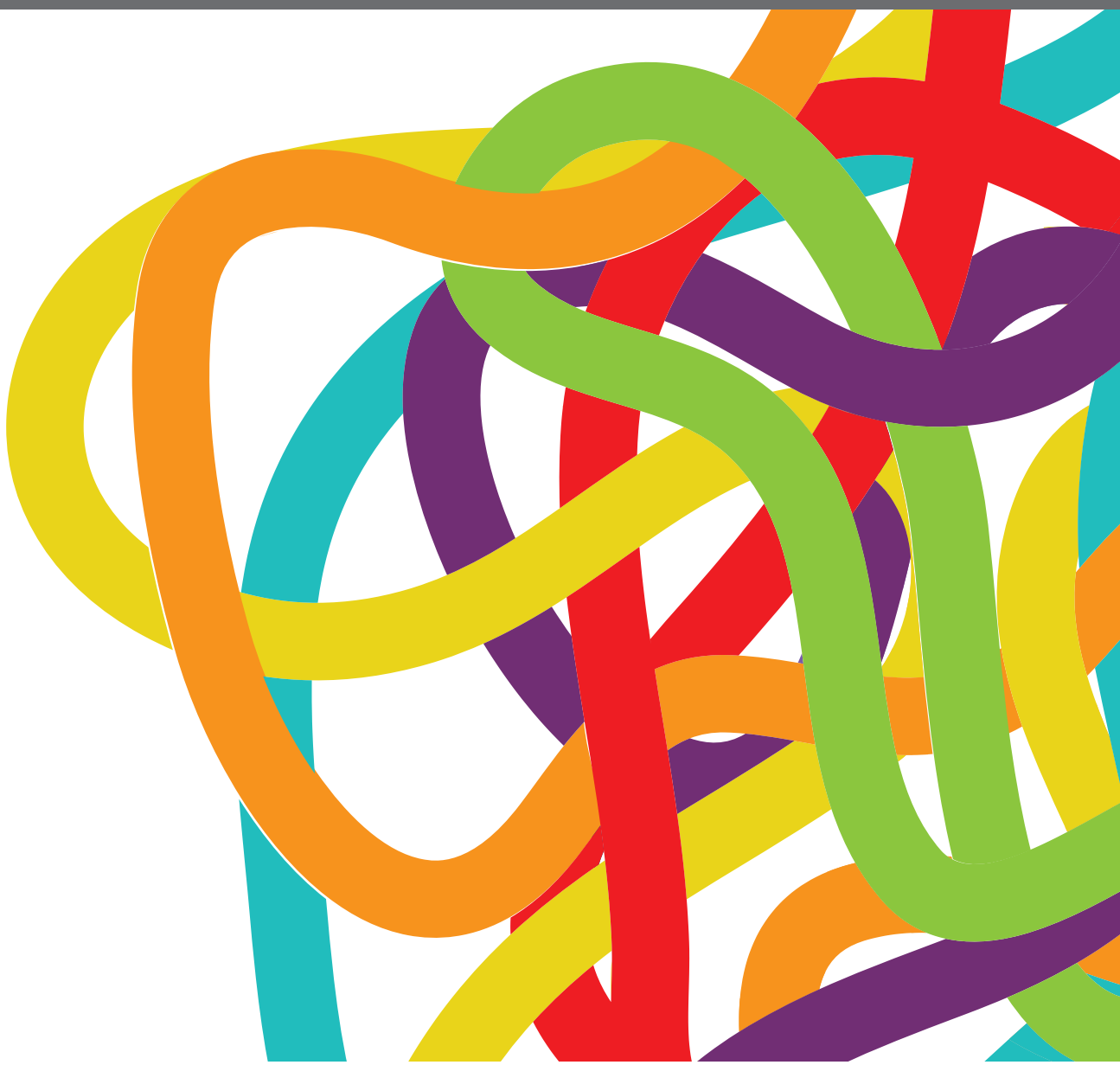
NOVEL METHODS FOR ONCOLOGIC IMAGING ANALYSIS: RADIOMICS, MACHINE LEARNING, AND ARTIFICIAL INTELLIGENCE

EDITED BY: Xuelei Ma, Lei Deng, Rong Tian and Chunxiao Guo

PUBLISHED IN: Frontiers in Oncology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-347-9
DOI 10.3389/978-2-88971-347-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

NOVEL METHODS FOR ONCOLOGIC IMAGING ANALYSIS: RADIOMICS, MACHINE LEARNING, AND ARTIFICIAL INTELLIGENCE

Topic Editors:
Xuelei Ma, Sichuan University, China
Lei Deng, Jacobi Medical Center, United States
Rong Tian, Sichuan University, China
Chunxiao Guo, University of Minnesota, United States

Citation: Ma, X., Deng, L., Tian, R., Guo, C., eds. (2021). Novel Methods for Oncologic Imaging Analysis: Radiomics, Machine Learning, and Artificial Intelligence. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-347-9

Table of Contents

frontiers in Oncology

EDITORIAL
published: 22 July 2021
doi: 10.3389/fonc.2021.628310

Editorial: Novel Methods for Oncologic Imaging Analysis: Radiomics, Machine Learning, and Artificial Intelligence

Hanyue Xu[1,2], *Lei Deng*[3], *Rong Tian*[4*] *and Xuelei Ma*[1*]

[1] Department of Biotherapy, West China Hospital, Sichuan University, Chengdu, China, [2] Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, China, [3] Department of Medicine, Jacobi Medical Center, Albert Einstein College of Medicine, Bronx, NY, United States, [4] Department of Radiology, West China Hospital, Sichuan University, Chengdu, China

OPEN ACCESS

Edited by:
Zaver Bhujwalla,
Johns Hopkins University,
United States

Reviewed by:
Meiyappan Solaiyappan,
Johns Hopkins University,
United States

***Correspondence:**
Rong Tian
zeyiqin@vip.126.com
Xuelei Ma
drmaxuelei@gmail.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 11 November 2020
Accepted: 08 July 2021
Published: 22 July 2021

Citation:
Xu H, Deng L, Tian R and Ma X (2021) Editorial: Novel Methods for Oncologic Imaging Analysis: Radiomics, Machine Learning, and Artificial Intelligence. Front. Oncol. 11:628310. doi: 10.3389/fonc.2021.628310

Keywords: machine learning, radiomics, tumor, artificial intelligence, diagnosis, prognosis

Editorial on the Research Topic

Novel Methods for Oncologic Imaging Analysis: Radiomics, Machine Learning, and Artificial Intelligence

INTRODUCTION

Radiomics is a quantitative and high-throughput radiological method that can aid in clinical decision-making, like treatment modality selection, and treatment plan optimization. By extracting plentiful of parameters from standard images, plenty of information that cannot be discovered by human naked eyes can be explored. Based on the hypothesis that these extra data provide additional information related to gene, protein and tumor phenotype, radiomics has gained increasing attention in cancer research. Meanwhile, because of the rich amount of data obtained in radiomics, sophisticated image analysis tools are required to analyze it. Many image-based signatures have been constructed by computer algorithms. Herein, this Research Topic recruited studies that exploring the usage of radiomics and artificial intelligence assisting clinical decision-making of tumors.

We are very glad to see that many excellent works were submitted to our Research Topic. In the end, a total of 36 papers were published, among which 34 were original studies and two were reviews. The researches were carried out in different countries, including China, USA, UK and France, and most of them were retrospective studies. They used various methods to explore the role of imaging in clinical decision-making. The methods used to select high-throughput imaging parameters can be divided into three levels, including the mathematical formulas level, Machine Learning level and Deep Learning level which belongs to Machine Learning but is more automatic. These kinds of analysis methods are constantly evolving to mimic the thinking patterns of the human brain, gaining the ability to analyze increasingly complicated data. However, for the lack of

open platform of images and non-uniform manual feature extraction, there is still a long way to go till a standard or a series of standardized radiomics signatures can be constructed.

PAPERS INCLUDED IN THIS RESEARCH TOPIC

Studies With Mathematical Formulas

As for the first level, using mathematical formulas, some studies generally extract parameters from images, then use statistical methods, like Mann–Whitney U-test, Spearman's rank correlation test, etc., to compare the internal and external differences of parameters, and then select the most heterogeneous data in different groups. After selecting the appropriate features, Machine Learning algorithms will be used to build models. This kind of studies included in the Research Topic used textures extracted from various images, like Positron Emission Tomography–Computed Tomography (PET/CT), CT, Magnetic Resonance Imaging (MRI), etc., to improve the accuracy of disease differentiation or prognosis prediction (Xu X. et al.; Xu H. et al.; Gao et al.; Hu et al.; Wang J. et al.; Zhou X. et al.; Zhang Y. et al.; Zhong et al.; Wu J. et al.; Zhang P. et al.; Chen W. et al.; Dong Y. et al.; Mai et al.; Li et al.). For instance, Zhang P. et al. differentiated seminomas and nonseminomas by MRI radiomics. Features were selected by comparing their heterogeneity among different groups and by assessing their relevance and redundancy. Then, Least Absolute Shrinkage and Selection Operator (LASSO), a regression analysis method, was used to select features to improve the mode prediction accuracy and interpretability (Zhang P. et al.). Mai et al. focused on the differentiation of phyllodes tumors and fibroadenoma with breast MRI texture analysis. They used a combination of a linear discriminant analysis and the K-Nearest Neighbor classifier to construct differentiative models (Mai et al.).

Studies With Machine Learning Algorithms

At the second analysis level, studies mainly used Machine Learning algorithms to select and classify radiomics features (Zhou H-F. et al.; Wang F. et al.; Yi et al.; Chen C. et al.; Huang et al.). Machine Learning algorithms build prediction models based on patterns in the training data and make predictions by comparing new instances to previous similar events, and they can be divided into supervised, unsupervised and semi-supervised learning algorithms, based on whether the data are labeled. Some studies selected one kind of algorithms to process data. Fei Wang et al. used LASSO to select features and Support Vector Machine (SVM) algorithm to constructed a predictive model and drew a nomogram to improve the preoperative T category accuracy (Wang F et al.). Huang et al. used LASSO regression model to select features and a multivariable logistic regression to develop predicting models. In addition, a nomogram was drawn by radiomics and clinical features to evaluate peritoneal metastasis status in gastric cancer (Huang et al.). Yi et al. predicted treatment response to neoadjuvant chemoradiotherapy in patients with locally advanced rectal cancer. Three aspects of the treatment response: not only partial clinical remission and good response, but also down-staging were evaluated. They used SVM rather than LASSO or Random Forest (RF) to regress features into a two-dimensional plane (Yi et al.).

Some other studies used multiple methods for feature selection and classification, as there are many kinds of Machine Learning methods with different advantages and drawbacks. They found the choice of classification methods accounted more than selection methods. Chen C. et al. used texture features to differentiate glioblastomas from metastatic brain tumors and differentiate glioblastoma from primary central nervous system lymphoma. In their studies, Linear Discriminant Analysis (LDA)-based models represented better performances than SVM-based models and Logistic Regression (LR)-based models (Chen C. et al.; Chen C. et al.). Similar results were found in other studies that compared different combinations (Tian et al.; Fan et al.; Zhang Y. et al.). For example, Yang Zhang et al. used five selection methods and nine classifiers. The combination of LASSO and LDA represented the best comprehensive performance (Zhang Y. et al.). LDA is a linear classifier whose decision boundary is a plane or a line, while SVM is a non-linear classifier with a decision boundary of a surface or a curved line. Although the above studies showed that LDA was superior to SVM, other studies uncovered the opposite results (Zhang Y. et al.; Payabvash et al.; Delzell et al.; Hong et al.). Zhang Y. et al. differentiated anaplastic oligodendroglioma from atypical low-grade oligodendroglioma. The best-performed combinations were various according to different image parameters. The combination of LASSO and RF classifier was the best for T1 images, while the combination of GBDT and RF classifier was the best for the fluid attenuated inversion recovery images (Zhang Y. et al.). In addition, Payabvash et al. differentiated posterior fossa tumors by using different Machine Learning classifiers, and also found RF models achieved greater accuracy. Delzell et al. used three types of classifying methods, including linear, nonlinear, and ensemble predictive classifying models, and found Elastic Net and SVM performed the best, while RF and Bagged Trees were the worst. It is impossible to draw firm conclusions about which method is the best, because there are too many influencing factors, such as sample size, parameter acquisition, extraction method, etc. However, at the very least, all the relevant articles show that Machine Learning methods are superior to manual methods, so more research on Machine Learning is necessary.

Studies With Deep Learning

Deep Learning is a subclass of Machine Learning that extends Deep Neural Networks to create complex neural architectures to solve difficult problems which would be impossible with traditional programming based on mathematical logic. Moawad et al. explored the feasibility of volumetric assessment of pre- and post- Transhepatic Arterial Chemotherapy And Embolization hepatocellular carcinoma using fully automated segmentation that based on a Convolutional Neural Network (CNN) approach (U-Net). For automated segmentation, attenuation of adjacent organs and the small size of lesions

were the main challenges. According to the assessment of response evaluation criteria in solid tumors, automated segmentation was a good substitute for manual segmentation (Moawad et al.). Sun et al. compared the deep CNN model based on breast ultrasound parameters with the radiomics model. Radiomics can be regarded as an accurate phenotypic analysis of medical images in which the imaging features are carefully defined in advance according to expert opinion. However, Deep Learning uses the raw data and analyzes the pixels and the voxel values by themselves. With convolution techniques, imaging features are automatically defined in the network. Thus, Deep Learning is the most artificial intelligent tool among these three analyzing levels. It is closest to human mode of thinking and can extract features and analyze them automatically.

Other Related Studies and Reviews

Moreover, there are some included studies focusing on optimizing the original features to promote the analysis results. Lacroix et al. optimized MR images before process with N4ITK bias field correction and normalizing voxel intensities with fat as a reference region. The results showed that correction of magnetic field heterogeneity and normalization of voxel values can promote the usage of radiomic features (Lacroix et al.). Wu W. et al. decomposed data by a non-linear kernelization method, Kernel Principal Component Analysis (KPCA), to find a new set of candidates and maximize the use of data. Lu et al. used concordance correlation coefficients to measure the fidelity in repeated experiments. A lot of features with good repeatability were found and their repeatability can be improved by using specific lesion-drawing methods. Zormpas-Petridis et al. proposed a novel multi-resolution hierarchical framework (SuperCRF) which can introduce the spatial context of a cell as additional information and improve the single cell classification algorithms. In other researches, topics related to radiomics and Machine Learning were discussed. Dong J. et al. and Ge et al. reviewed the usage of radiomics and Machine Learning in the management of cancers, and summarized computer-aided clinical decision-making as a promising solution.

CONCLUSION

In conclusion, the combination of radiomics and Machine Learning can provide clinical practice convenience, as long as some obstacles can be solved. The limitations of Machine Learning-based radiological decision-making mainly lie in the following aspects: Firstly, the data quality is uneven, and thus open data-platforms like http://www.predictcancer.org need to be built. Secondly, based on open image sources, algorithms of lesion delineation, feature extraction and signature construction require more standard reference to increase the generalization of the results. Thirdly, more studies that based on uniform data and algorithms and comparing the efficiency of computer-aid and conventional clinical decision-making, are required to better promote the usage of Artificial Intelligence in clinic.

This Research Topic involved many studies, which used the combination of radiomics and Machine Learning in tumor management. We appreciate all the reviewers and authors for their contributions to this Research Topic. We hope this Research Topic can arouse more attention in the related fields.

AUTHOR CONTRIBUTIONS

HX and LD wrote the first draft of the manuscript. RT and XM contributed to manuscript revision. All authors contributed to the article and approved the submitted version.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 26 June 2019
doi: 10.3389/fonc.2019.00552

MRI-Based Radiomics Predicts Tumor Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer

Xiaoping Yi[1,2†], Qian Pei[3†], Youming Zhang[1†], Hong Zhu[4], Zhongjie Wang[5], Chen Chen[6], Qingling Li[7], Xueying Long[1], Fengbo Tan[3], Zhongyi Zhou[3], Wenxue Liu[8], Chenglong Li[3], Yuan Zhou[3], Xiangping Song[3], Yuqiang Li[3], Weihua Liao[1], Xuejun Li[5], Lunquan Sun[2], Haiping Pei[3*], Chishing Zee[9] and Bihong T. Chen[10]*

[1] Department of Radiology, Xiangya Hospital, Central South University, Changsha, China, [2] Postdoctoral Research Workstation of Pathology and Pathophysiology, Basic Medical Sciences, Xiangya Hospital, Central South University, Changsha, China, [3] Department of General surgery, Xiangya Hospital, Central South University, Changsha, China, [4] Department of Radiation Oncology, Xiangya Hospital, Central South University, Changsha, China, [5] Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, China, [6] Department of Radiology, Zhuzhou 331 Hospital, Zhuzhou, China, [7] Department of Pathology, Xiangya Hospital, Central South University, Changsha, China, [8] Department of Cardiology, Xiangya Hospital, Central South University, Changsha, China, [9] Department of Radiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States, [10] Department of Diagnostic Radiology, City of Hope National Medical Center, Duarte, CA, United States

OPEN ACCESS

Edited by:
Xuelei Ma,
Sichuan University, China

Reviewed by:
Ana Paula Candiota,
Centre for Biomedical Network Research (CIBER), Spain
Seyedmehdi Payabvash,
Yale University, United States

***Correspondence:**
Haiping Pei
peihaipingxy@163.com
Weihua Liao
doctorliaoweihua@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *07 December 2018*
Accepted: *06 June 2019*
Published: *26 June 2019*

Citation:
Yi X, Pei Q, Zhang Y, Zhu H, Wang Z, Chen C, Li Q, Long X, Tan F, Zhou Z, Liu W, Li C, Zhou Y, Song X, Li Y, Liao W, Li X, Sun L, Pei H, Zee C and Chen BT (2019) MRI-Based Radiomics Predicts Tumor Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. Front. Oncol. 9:552. doi: 10.3389/fonc.2019.00552

Background: Conventional methods for predicting treatment response to neoadjuvant chemoradiotherapy (nCRT) in patients with locally advanced rectal cancer (LARC) are limited.

Methods: This study retrospectively recruited 134 LARC patients who underwent standard nCRT followed by total mesorectal excision surgery in our institution. Based on pre-operative axial T2-weighted images, machine learning radiomics was performed. A receiver operating characteristic (ROC) curve was performed to test the efficiencies of the predictive model.

Results: Among the 134 patients, 32 (23.9%) achieved pathological complete response (pCR), 69 (51.5%) achieved a good response, and 91 (67.9%) achieved down-staging. For prediction of pCR, good-response, and down-staging, the predictive model demonstrated high classification efficiencies, with an AUC value of 0.91 (95% CI: 0.83–0.98), 0.90 (95% CI: 0.83–0.97), and 0.93 (95% CI: 0.87–0.98), respectively.

Conclusion: Our machine learning radiomics model showed promise for predicting response to nCRT in patients with LARC. Our predictive model based on the commonly used T2-weighted images on pelvic Magnetic Resonance Imaging (MRI) scans has the potential to be adapted in clinical practice.

Novelty and Impact Statements: Methods for predicting the response of the locally advanced rectal cancer (LARC, T3-4, or N+) to neoadjuvant chemoradiotherapy (nCRT) is lacking. In the present study, we developed a new machine learning radiomics

method based on T2-weighted images. As a non-invasive tool, this method facilitates prediction performance effectively. It achieves a satisfactory overall diagnostic accuracy for predicting of pCR, good response, and down-staging show an AUC of 0.908, 0.902, and 0.930 in LARC patients, respectively.

Keywords: locally advanced rectal cancer (LARC), neoadjuvant chemoradiotherapy (nCRT), treatment response, magnetic resonance imaging (MRI), machine learning radiomics

INTRODUCTION

Rectal cancer is a common malignancy worldwide, accounting for ∼30–50% of colorectal cancer (1, 2). Moreover, in rectal cancer patients, lesions are usually located in middle-low rectum, which causes increased difficulty in treatment and worse prognosis, especially the locally advanced rectal cancer (LARC, T3-4 or N+) (3, 4). Currently, neoadjuvant chemoradiotherapy (nCRT) followed by total mesorectal excision is the recommended treatment for LARC patients, especially those with lesions located in the middle-low rectum (5). The advantages of nCRT are usually significant (6, 7). However, the response of LARC to nCRT varies widely, ranging from pathological complete response (pCR, ypT0N0M0) with no viable tumor cells left in the surgical specimen, to virtually no tumor regression at all (stable) or even tumor progression in a small group of patients (8, 9). Among these patients, pCR is not only associated with favorable disease-free and overall survival (7, 10), but also motivates the "watch-and-wait" treatment strategy, a non-operative option for patients achieving clinical complete response (11). Therefore, clinicians are motivated to identify ways to accurately predict patients' individual responses to nCRT.

Radiological examination has been considered to be one of the means most likely to accomplish this task (12). Among all modalities, Magnetic Resonance Imaging (MRI) is regarded as the most promising method because it uses no radiation, shows high soft tissue resolution, and has wide routine clinical application for evaluation of rectal cancer. Notably, some conventional and functional MRI methods have been reported to show some advantages in predicting tumor response to nCRT (13–15). Unfortunately, conventional MRI analysis remains limited when predict treatment response in individual patient using experience (16). There is a need to develop new methods.

Quantitative image data analysis, such as texture analysis and radiomics are procedures for converting clinical images into high-dimensional, exploitable, and quantitative imaging features by high-throughput extraction of data-characterization algorithms (17). In addition to clinical outcomes, the biomedical information contained in medical images, such as overall information about phenotype and microenvironment of the tumor, may be vitally important for evidence-based clinical decision support. In theory, all magnetic resonance images in different can be used as a source of analysis. In theory, for quantitative analysis, used features can be extracted from images of all modalities (12, 16, 18–22). However, T2 weighted image is almost the most widely used one, when considering the wide availability of images which can be stably acquired based on different machines. Quantitative image data analysis methods have the potential to reveal such biomedical information, providing an opportunity to improve decision-support in oncology and non-invasively (17, 23). The potential advantage of this kind of method has already been verified in colorectal cancer (24) and a variety of other cancers, including nasopharyngeal carcinoma (25), lung cancer (17), and breast cancer (26). Recently, some independent studies (12, 19–22, 24) reported that a multimodality MRI based radiomics model could predict RC tumor response to nCRT with an improved accuracy for pCR and good response prediction. However, due to the relatively small sample size, or the inclusion of multimodality images with other MRI sequences such as diffusion-weighted imaging, or the lack of integration of important relevant clinical pathological features, there is a need for improving accuracy of the prediction model.

In the present study, we retrospectively collected 134 consecutive surgically and pathologically confirmed LARC patients who received standard nCRT before surgery. We developed a machine learning radiomics model based on imaging data extracted from the T2-weighted images, and validated its prediction efficiency of treatment response to nCRT in patients with LARC.

Abbreviations: LARC, locally advanced rectal cancer; nCRT, neoadjuvant chemoradiotherapy; MRI, magnetic resonance imaging; RC, rectal cancer; CRC, colorectal cancer; pCR, pathological complete response; cCR, clinical complete response; AUC, area under curve; DWI, diffusion-weighted imaging; GR, good response; ROI, region of interest; ROC, receiver operating characteristic.

MATERIALS AND METHODS

Patients

This retrospective study was approved by our institutional review board (IRB No. 201610070). The written informed consents from patients were waived.

Medical data of consecutive biopsy-proven rectal adenocarcinoma patients with LARC treated with nCRT followed by total mesorectal excision between March 2009 and December 2017 in our institution were retrospectively analyzed. Complete clinical data, including MRI imaging of all patient's performed before radiotherapy, was analyzed. Details about the inclusion and exclusion criteria, clinical and pathological characteristics, and treatments information can be found in **Supplementary Files**. The patients recruiting process was shown in **Figure 1**.

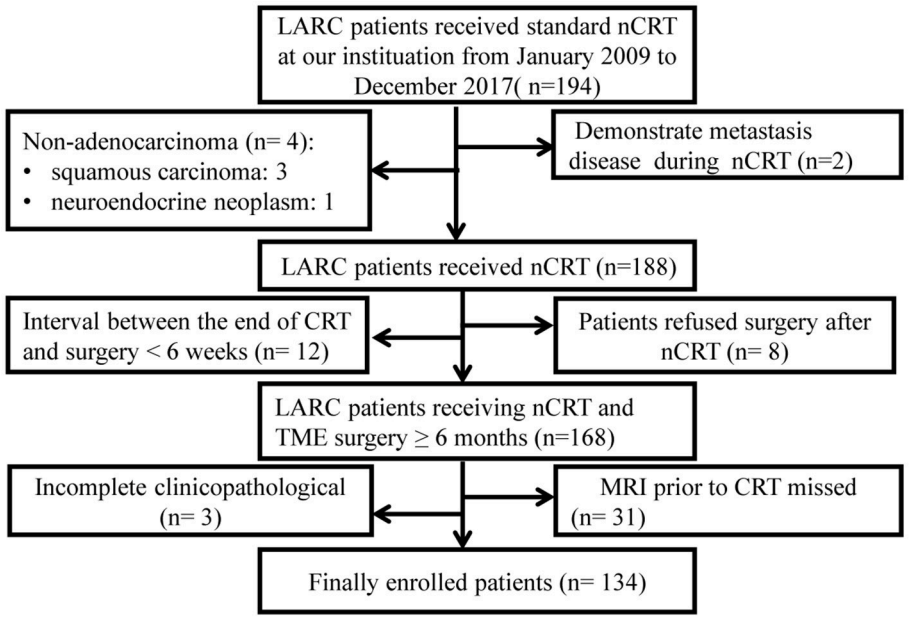


FIGURE 1 | Flow-chart. LARC, Locally advanced rectal cancer; nCRT, neoadjuvant chemoradiotherapy; TME, total mesorectal excision.

Pathological Assessments of Tumor Samples

Each specimen was sampled and evaluated by two experienced dedicated gastrointestinal pathologists. The two pathologists were both blind to the MRI data and clinical data. Criteria for pCRT and non-pCRT were defined as described in previous reports (8). We also classified TRG 3–4 into the good response (GR) group, and TRG 0–2 into the non-GR group according to Dowrak/Rödel's system (27). Changes in TNM staging were recorded by comparing to cTNM before the surgery, and responses were classed as either down-staging or non-down-staging (stability and progression). Details can be found in the **Supplementary Files**.

MRI Image Acquisition

All patients underwent an MRI scan in our hospital with either a 1.5 Tesla (Siemens, Erlangen, Germany) or a 3.0 Telsa scanner (GE, Milwaukee, US), using a phased-array body coil, 3–10 days before the start of chemoradiation. To ensure MRI image quality, a quality assurance check was performed biweekly by a hospital radiological physicist and executed bimonthly by the Siemens or GE engineer, as appropriate, according to the maintenance rules for the MRI scanners in our institute. Axial T2-weighted (T2w fast spin echo sequence) images (T2WI) and T1-weighted (T1w spin echo sequence) images (T1WI) were acquired regularly. Subsequently, multiphase T1w images were obtained before and after contrast injection, using a spoiled gradient echo sequence (LAVA/VIBE sequence). Contrast injection and data acquisition were triggered simultaneously. Briefly, a total of four repetitions were acquired, including one before the contrast injection and three after the injection (at 28, 65, and 120 s). For contrast, generally 90–100 ml of the gadolinium-based contrast media dimeglumine gadopentetate (Magnevist; Schering Diagnostics AG, Berlin, Germany) was administrated intravenously at a rate of 2.5 ml/s through a high pressure injector (Optistar LE, Liebel-Flarisheim Company, OH, USA).

Since all patients had at least three kinds of MRI images (T1WI, T2WI, and enhanced T1WI), MRI images from these three serials were included in the present study.

MRI Image Analysis

All MRI images of each patient were evaluated independently by two experienced abdominal radiologists (reader 1. C.C with 7 years of experience; reader 2, L.X.Y with 15 years of experience), who were totally blinded to all medical information. Final disagreement was resolved in a panel format including two additional radiologists (L.W.H and Y.X.P). The location and boundary of the tumor were confirmed, tumor size, the distance from the lower edge of the tumor to the anal canal, and the MRI-based TNM stage were recorded. The findings were recorded by consensus.

Texture Analysis Feature Extraction

For each patient, an anonymized representative axial T2WI image in which the lesion had the largest cross-sectional area was selected and retrieved from Picture Archiving and Communication System (PACS, Carestrem, Canada) using Digital Imaging and Communications in Medicine (DICOM) Works software (version 1.3.5). Subsequently, each image was transferred to a personal computer and inputted into the texture analysis software (MaZda Version 4.6, Institute of Electronics, Technical University of Lodz, Poland) (28).

Briefly, the process of texture analysis feature extraction was conducted by 2-steps as follows: (a) selecting and retrieving the suitable MRI images, and then (b) outlining the tumors as the regions of interest (ROI), and extracting quantitative texture analysis features by using the texture analysis software (MaZda Version 4.6, Institute of Electronics, Technical University of Lodz, Poland).

Tumors were outlined as a region of interest (ROI) by performing MaZda on T2WI, while using all other image sequences (especially gadolinium-enhanced images) as references in cases where the margin of the rectal lesion was difficult to define on unenhanced images. Briefly, a ROI was delineated initially by following the tumor outline, with notation that fat and air outside the mass are not included. Then, the ROI was saved for subsequent texture analysis. Contouring was performed carefully to cover the maximum extent of the tumor without exceeding the lesion border, to avoid contamination from adjacent normal rectal tissues or the intestinal lumen. For each ROI, a total of 340 quantitative features were automatically generated using MaZda software, including a gray level histogram, gradient, run-length matrix, co-occurrence matrix, autoregressive model, and wavelet transform analysis according to the software settings.

Evaluation of the Reproducibility of Radiomics Feature Extraction by the Two Radiologists

The reproducibility assessment of the features extracted by the two radiologists from the independent segmentations of T2WI images of all patients was performed. The inter-observer (reader 1 v reader 2) and intra-observer (reader 1 twice) correlation coefficient values were evaluated. The final consistency is evaluated by the following criteria regarding the correlation coefficient values: <0.20 indicates poor reproducibility, 0.21–0.40 fair reproducibility, 0.40–0.60 moderate reproducibility, 0.61–0.80 good reproducibility, and 0.81–1.00 excellent reproducibility. Generally, a correlation coefficient >0.75 is regarded as being in good agreement.

For the Kappa consistency test, excellent, good, and poor agreement were defined as kappa values of >0.81, in the range of 0.61–0.80, and <0.60, respectively.

The Mann-Whitney *U*-test was used to compare the values of each feature between the two groups. An independent samples *t*-test or Kruskal-Wallis *H* test, where appropriate, was used to assess the differences between the features generated by reader 1 (first time) and those generated by reader 2, as well as between the features generated twice by reader 1.

Inter-observer and intra-observer reproducibility of texture feature extraction was initially analyzed with 50 randomly chosen images from all T2WI images selected for evaluation by the two radiologists (reader 1, and reader 2). To assess the intra-observer reproducibility, reader 1 repeated the generation of texture features twice within a 2-week period following the same procedure. Reader 1 completed the workflow for the remaining images.

Statistical Analysis, Features Selection, Signature Generation, and Prediction Model Building

All statistical analyses were conducted using IBM SPSS version 20.0.0 (IBM Corporation, Armonk, NY, USA). To test the difference between groups, the Wilcoxon rank-sum test was performed for the quantitative features, and the chi-square test or fisher's exact test was performed for the qualitative features.

All data processing, data reduction and feature selection, and model built were performed using MATLAB 2017a (The Mathworks, Inc., Natick, MA, USA). The least absolute shrinkage and selection operator (LASSO) method, was used to select the most useful predictive features from the primary data set, and a radiomics score (Rad-score) was calculated for each patient at the mean time as a linear combination of selected features that were weighted by their respective coefficients. Based on these selected features, another classification model was also constructed by Random forest (RF), and the RF-score was generated. Subsequently, a combined classification model was finally built by the support vector machine (SVM) method (SVM-score), based on Rad-score and RF-score in the previous step. On the basis of two SVM-scores obtained, calculated from TE and TRC features, respectively, a final classification model was generated by using the SVM method again (SVM-score-final). Through the above steps, a total of seven models were generated representing each classification task, considering there are two kinds of data (Texture analysis [TA] features, and Traditional radiological-clinicopathological [TRC] features) that were used to build the model. The seven models include three models generated from TA features (model based on Rad-score, RF model, and first-step SVM model), three models generated from TRC features (model based on Rad-score, RF model, and first-step SVM model), and one combined SVM model.

The basic idea of this algorithm is to consider LASSO and RF as weak regressors and combine them using SVM. For each type of data, i.e., texture feature, we first use LASSO to obtain the Rad-score, and use its side product, i.e., important features, as the feature set of RF to obtain the RF-score. Since Rad-score and RF-score are independently acquired by two different weak regressors, using SVM to regress them in a two-dimensional plane achieves a better result than by them owns. Moreover, data sets Texture feature and Traditional radiological-clinicopathological data are also independent to each other. So for the same reason, we use SVM to regress the scores from Texture feature and Traditional radiological-clinicopathological data to get the final regression score. Finally, the regressed scores can be binarized for further prediction.

To evaluate the performance of the models, all patients were divided into two cohorts: a training cohort and a validation cohort. The models were developed in the training cohort, and tested in the validation cohort. The classification efficiencies of each kind of model mentioned above, including the receiver operating characteristic (ROC) curves, both in the training and validation cohort were calculated. A P-value < 0.05 was considered statistically significant. Details of the flow chart for building the classification model are shown in **Figure 2**.

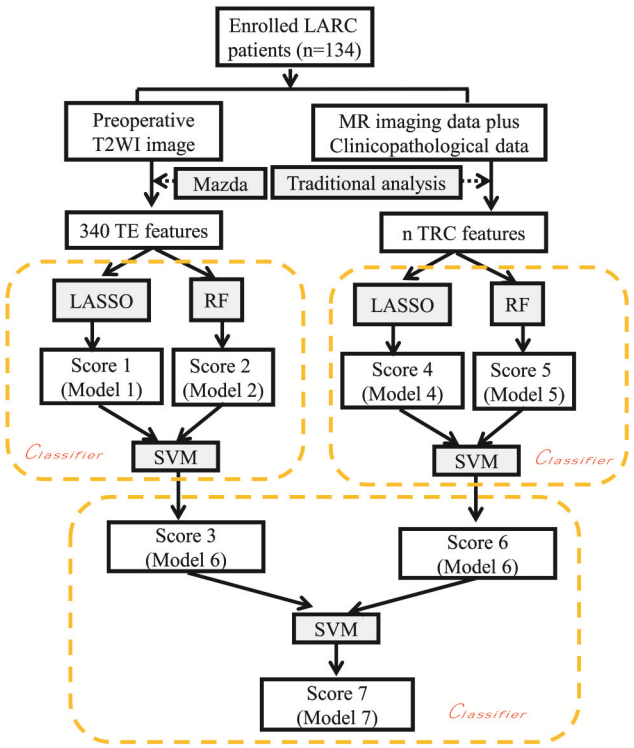


FIGURE 2 | Flow chart depicting construction of the classification models. LARC, locally advanced rectal cancer; TE, texture features; TRC Traditional radiological features and clinicopathological features; LASSO, least absolute shrinkage and selection operator; RF Random forest; SVM support vector machine. The n_x or n_y terms used here indicate the different numbers of selected features used in the LASSO method with three different reduction schemes based on 340 TE features and 31 TRC features, respectively. The texture analysis software used was MaZda Version 4.6 (Institute of Electronics, Technical University of Lodz, Poland).

RESULTS

Patients Characteristics

There were134 patients enrolled in the present study. Patients' characteristics in the training and validation cohorts were summarized in **Table 1**. Patients were randomly allocated into training cohort and validation cohort in a 3:1 ratio for building the pCR predictive model, the down-staging model, and the good response model. No differences were found between the training and validation cohorts in any of the three models. In addition, the patients' clinicopathological data mostly consisting of laboratory data were also used in building the predictive models.

The Classification Model Building and Predicting Efficiency

From a total of 340 features that were extracted from T2-wighted images for each patient, a set of features with corresponding numbers were selected by LASSO and used to calculate the Rad-scores for the pCR, Good Response, and Down-staging models.

Predicting Pathological Complete Response (pCR)

On the basis of the selected 10 texture and 8 clinicopathological features, a predictive model was finally constructed with SVM method for pCR prediction. The SVM model yielded an AUC of 90.78% in the training cohort, and 87.45% in the validation cohort (**Figure 3** and **Supplementary Figure 1**).

Predicting Good Response (GR)

The predictive model built based on the 10 texture features and 7 clinicopathological features achieved an AUC of 90.17% in the training cohort, and 89.72% in the validation cohort (**Figure 3** and **Supplementary Figure 1**).

Predicting Down-Staging

The predictive model with 10 texture features and 7 clinicopathological features showed an AUC of 92.97% in the training cohort, and 89.20% in the validation cohort (**Figure 3** and **Supplementary Figure 1**). Details about prediction efficiency of three kinds of models could be found in **Table 2**. The correlation matrix of the selected features used in the three kinds of models was showed as **Figure 4**.

DISCUSSION

To the best of our knowledge, this was the first cohort studied to date utilizing monosequence-MRI-based machine learning radiomics to predict tumor response to neoadjuvant chemoradiation therapy in patients with locally advanced rectal cancer. Our predictive model constructed with both radiomics features and clinicopathological data achieved higher accuracies than previously reported in the literature, with an AUC of more than 90%. Substantial evidence from prior studies has demonstrated that a number of clinicopathologcial and radiological features may help to predict treatment response (16, 18, 29). Nevertheless, no single factor has stood out to be the most reliable way for clinicians to use in decision making process (6, 16). It is important to distinguish the LARC patients who will likely to respond to nCRT from patients who would not. However, this has not been achieved yet. We introduced here a new imaging oriented strategy for a better prediction, which may have potential for clinical practice.

Our study is in general accord with prior research (19, 30, 31). Nie et al. (12) have reported a relatively satisfactory result by using a radiomics method, with an AUC of 0.84 for pCR and 0.89 for good response prediction. Most recently, Cui et al. (19) reported a further attempt on a bigger group LARC patients by similar methods, which show very high predictive efficiency with an AUC of 0.944. In addition, several LARC studies (20, 21) also perform similar radiomics-based studies with good experimental results, using features extracted from multimodality MR images including T2WI. However, there were obvious advantages in our study when compared to these studies. First, we fully evaluated three aspects of the treatment response: not only pCR and good response, but also down-staging. Our study has the potential to provide more information on the tumor and treatment response. Second, the number of enrolled patients in our study ($n = 134$), was larger than that in the Nie's ($n = 48$), and comparable

TABLE 1 | Clinicopathological characteristics in three tumor response predictive models.

	Down-staging predictive model			pCR predictive model			Good response predictive model		
	Training cohort	**Validation cohort**	***P*-value**	**Training cohort**	**Validation cohort**	***P*-value**	**Training cohort**	**Validation cohort**	***P*-value**
Gender									
Male	54 (57.4%)	26 (65.0%)	0.415	57 (61.3%)	23 (56.1%)	0.572	55 (58.5%)	25 (62.5%)	0.667
Female	40 (42.6%)	14 (35.0%)		36 (38.7%)	18 (43.9%)		39 (41.5%)	15 (37.5%)	
Age (years)	50.62 ± 10.29	54.70 ± 10.74	0.040	52.32 ± 10.68	50.73 ± 10.32	0.424	52.02 ± 10.82	51.4 ± 10.04	0.757
Distance from the anal verge (cm)	5.00 (3.00–6.00)	5.30 ± 2.15	0.264	5.0 (4.0–6.0)	5.00 (3.00–6.50)	0.185	5.00 (3.38–6.25)	5.00 (3.00–6.00)	0.233
Pathology type			0.463			0.320			0.087
Well/moderately differentiated adenocarcinoma	70 (74.5%)	33 (82.5%)		71 (76.3%)	32 (78.0%)		70 (74.5%)	33 (82.5%)	0.087
Poor differentiated adenocarcinoma	17 (18.1%)	6 (15.0%)		18 (19.4%)	5 (12.2%)		20 (21.3%)	3 (7.5%)	
Mucinous carcinomas	7 (7.4%)	1 (2.5%)		4 (4.3%)	4 (9.8%)		4 (4.3%)	4 (10.0%)	
Clinical T staging (cT)			1.000			1.000			0.364
cT2	3 (3.2%)	0		2 (2.2%)	1 (2.4%)		1 (1.1%)	2 (5.0%)	
Ct3	73 (77.7%)	29 (72.5%)		71 (76.3%)	31 (75.6%)		72 (76.6%)	30 (75.0%)	
cT4	18 (19.1%)	11 (27.5%)		20 (21.5%)	9 (22.0%)		21 (22.3%)	8 (20.0%)	
Clinical N staging (cN)			0.632			0.847			0.540
cN0	18 (19.1%)	10 (25.0%)		17 (18.3%)	11 (26.8%)		19 (20.2%)	9 (22.5%)	
cN1a	18 (19.1%)	10 (25.0%)		20 (21.5%)	8 (19.5%)		18 (19.1%)	10 (25.0%)	
cN1b	25 (26.6%)	8 (20.0%)		23 (24.7%)	10 (24.4%)		26 (27.7%)	7 (17.5%)	
cN1c	1 (1.1%)	0		1 (1.1%)	0		1 (1.1%)	0	
cN2a	20 (21.3%)	5 (12.5%)		19 (20.4%)	6 (14.6%)		15 (16.0%)	10 (25.0%)	
cN2b	12 (12.8%)	7 (17.5%)		13 (14.0%)	6 (14.6%)		15 (16.0%)	4 (10.0%)	

to Cui y's ($n = 186$), which can ensure the desired prediction results. Third, we also included conventional MRI findings and clinicopathological data which may further improve the prediction. Lastly, our radiomic features were extracted from only one sequence, i.e., the T2-weighted images, other than the multi-sequence MRI images used in previous studies. The T2 weighted images are commonly used in clinical practice, which is familiar to radiologists. In addition, it can be acquired easily and the images are quite stable in appearance, especially when compared with images obtained by special sequence, such as diffusion weighted images. Notably, diffusion weighted images are prone to distortion and susceptibility artifacts, which affect tumor segmentation and data extraction. Similarly, other sequences such as T1-weighted dynamic contrast enhanced images depend on the amount and distribution of the injected contrast-enhancing agent, which might be influenced by variable hemodynamic conditions in the patients.

The exact reason for why quantitative MRI-based texture data appear to be able to predict treatment response is still largely unknown. In theory, the biological phenotype of tumors, including treatment response, is largely determined by their underlying molecular subtypes, whose manifestations may vary. One of the phenotypes may be radiological heterogeneity, including inter- and intra-tumor heterogeneity. A large body of literature indicates that texture based radiomic modeling can evaluate tumor heterogeneity, and can correlate radiological findings with underlying genomic and biological characteristics, including prognosis and treatment response (17, 23). Our study may add into the literature in this regard as we have shown a predictive model for treatment response with high accuracy. From another perspective, the large amount of previous evidence (13, 15), supporting using advanced MRI-based radiomic features to predict different responses to nCRT in patients with rectal cancer.

In addition, we introduced clinicopathological features into the prediction model, which may contribute significantly to the improvement of prediction efficiencies. These features may represent, to some extent, some of the intrinsic properties of the tumor (32, 33). For example, the fecal occult blood test and red cell counts may indicate oxygen status of tumor. Neutrophil counts, Monocyte counts, globulin, or platelet counts, may actually reflect the immune status of LARC patients to some extent. The hypoxia and immune status of the tumor can influence tumor treatment response and mediate radiotherapy resistance (34, 35). Pathology type and distance from the anal verge also influence the tumor response, as has been shown in previous studies (36, 37). Our study results suggest that these clinicopathological data may play an important role in treatment response.

There were several limitations in our study. First, as a retrospective study, there may be a selection bias. Second, the sample size in our study was modest, which may affect the

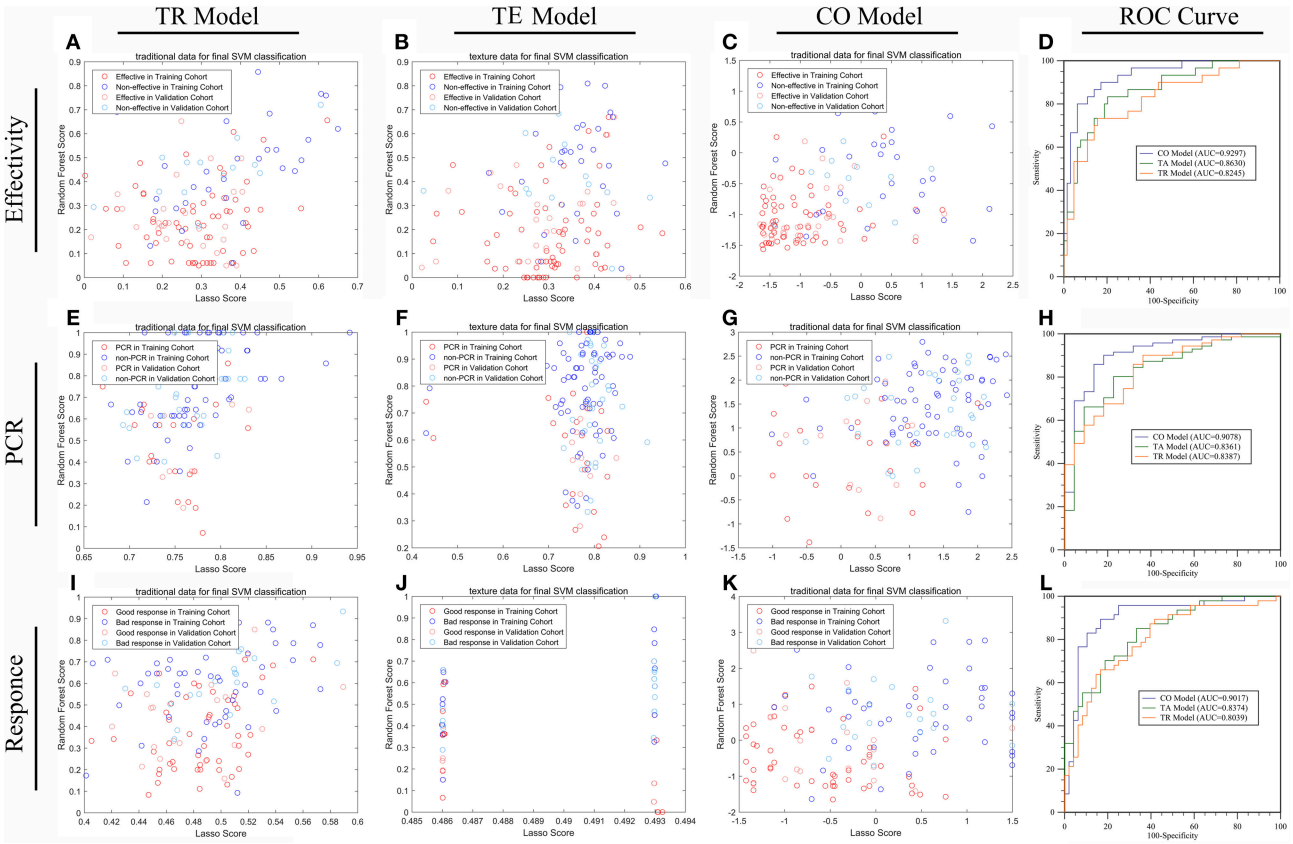


FIGURE 3 | The efficiencies of machine learning models predicting treatment response in LARC patients receiving nCRT. The distribution of patients with down-staging disease or not **(A–D)**, pCR or not **(E–H)**, and good response or not **(I–L)**, in TRC based model (model 6) **(A,E,I)**, TE based model (model 3) **(B,F,J)** and the combined TRC and TE based model (Model 7) **(C,G,K)** were demonstrated by scatter plots. The ROC test **(D,H,L)** shows that the efficiency of model 7 was significantly higher than that of either model 6 or model 3 in all three missions (all $P < 0.05$). There is no significant difference in prediction efficiency between model 6 and model 3 in any of the three missions (all $P > 0.05$).

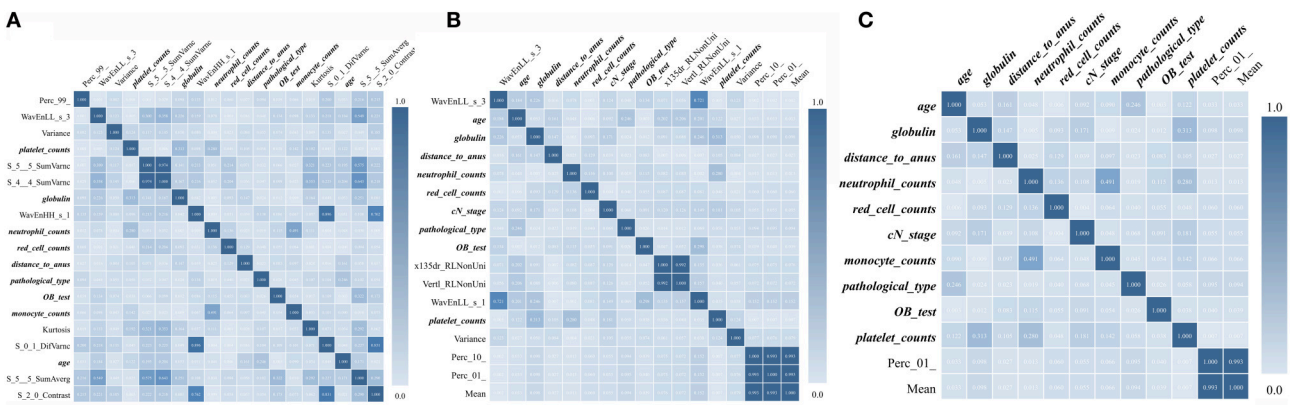


FIGURE 4 | Correlation matrix maps show the correlation among all TE and TRC features used in predictive models. **(A)** Down-staging model. **(B)** PCR model. **(C)** Good-response model. TRC features are expressed in bold fonts.

accuracy and stability of the predictive models. Third, both the building and validation of the models were conducted in our institution with a single dataset. A multicenter prospective study might be helpful to further validate and optimize our prediction models. The texture features were extracted from the largest cross-sectional area of the tumor rather than from the

TABLE 2 | The efficience of models to predict the treatment response in LARC petients.

	Training cohort			Validation cohort		
	Score 3	**Score 6**	**Score 7**	**SVM1**	**SVM2**	**SVM3**
DOWN-STAGING						
AUC	0.8630 (95% CI: 78.36–94.25%)	0.8245 (95% CI: 73.07–91.83%)	0.9297 (95% CI: 87.62–98.31%)	0.8006 (95% CI: 65.28–94.84%)	0.8462 (95% CI: 72.46–96.77%)	0.8920 (95% CI: 79.40–99.01%)
Specificity	0.82812 (95% CI: 68.750–95.312%)	0.85938 (95% CI: 59.375–96.875%)	0.90625 (95% CI: 73.438–98.438%)	0.7778 (95% CI: 51.85–96.30%)	0.7778 (95% CI: 51.85–96.30%)	0.7778 (95% CI: 62.96–92.59%)
Sensitivity	0.83333 (95% CI: 63.333–96.667%)	0.73333 (95% CI: 56.667–93.333%)	0.9000 (95% CI: 73.333–100.000%)	0.8462 (95% CI: 61.54–100.00%)	0.9231 (95% CI: 69.23–100.00%)	0.9231 (95% CI: 76.92–100.00%)
Accuracy	0.82979 (95% CI: 74.468–89.362%)	0.81915 (95% CI: 69.149–89.362%)	0.89362 (95% CI: 79.787–95.745%)	0.8000 (95% CI: 65.00–92.50%)	0.8250 (95% CI: 67.50–92.56%)	0.8500 (95% CI: 70.00–92.50%)
PCR						
AUC	0.8361 (95% CI: 74.13–93.09%)	0.8387 (95% CI: 74.83–92.91%)	0.9078 (95% CI: 83.15–98.41%)	0.8194 (95% CI: 69.08–94.79%)	0.7581 (95% CI: 58.56–93.05%)	0.8745 (95% CI: 74.82–99.49%)
Specificity	0.86364 (95% CI: 63.64–100.00%)	0.77273 (95% CI: 54.55–100.00%)	0.86364 (95% CI: 72.73–100.00%)	1.00 (95% CI: 80.00–100.00%)	0.9000 (95% CI: 40.00–100.00%)	0.9000 (95% CI: 50.00–100.00%)
Sensitivity	0.77465 (95% CI: 54.93–91.55%)	0.85915 (95% CI: 49.30–95.78%)	0.88732 (95% CI: 69.01–97.18%)	0.67742 (95% CI: 48.39–90.32%)	0.67742 (95% CI: 32.26–100.00%)	0.80645 (95% CI: 51.61–100.00%)
Accuracy	0.78495 (95% CI: 64.52–89.25%)	0.82796 (95% CI: 61.29–90.32%)	0.88172 (95% CI: 75.27–94.62%)	0.7561 (95% CI: 60.98–90.24%)	0.73171 (95% CI: 48.78–92.68%)	0.85366 (95% CI: 63.35–95.18%)
GOOD-RESPONSE						
AUC	0.8374 (95% CI: 75.95–91.53%)	0.8039 (95% CI: 71.42–89.36%)	0.9017 (95% CI: 83.30–97.05%)	0.7920 (95% CI: 65.24–93.16%)	0.7744 (95% CI: 62.65–92.23%)	0.8972 (95% CI: 80.19–99.25%)
Specificity	0.77083 (95% CI: 54.17–97.92%)	0.8125 (95% CI: 52.08–93.75%)	0.875 (95% CI: 70.83–97.92%)	0.7143 (95% CI: 38.10–100.00%)	0.7143 (95% CI: 38.10–95.24%)	0.8571 (95% CI: 66.67–100.00%)
Sensitivity	0.80435 (95% CI: 50.00–95.65%)	0.73913 (95% CI: 56.52–95.65%)	0.9130 (95% CI: 76.09–100.00%)	0.8947 (95% CI: 47.37–100.00%)	0.8421 (95% CI: 52.63–100.00%)	0.8947 (95% CI: 68.42–100.00%)
Accuracy	0.7766 (95% CI: 70.21–85.11%)	0.7766 (95% CI: 69.15–85.11%)	0.88298 (95% CI: 81.92–93.62%)	0.7500 (95% CI: 65.00–87.50%)	0.7750 (95% CI: 65.00–87.50%)	0.8750 (95% CI: 75.00–95.00%)

entire tumor, which may raise questions as to whether these features were optimally representative of the characteristics of the entire tumor. Lastly, the MRI images used in the texture feature extraction were obtained from three different MRI scanners (Siemens and GE) in our hospital, and differences among the scanners may potentially influence the texture features and the subsequent model building. Future research is needed to standardize the signal intensity among different MRI scanners.

CONCLUSION

Our study showed a predictive model built with radiomic features and clinicopathological data was promising to predict tumor response to neoadjuvant chemoradiation in patients with locally advanced rectal cancer. In addition, our method developed with information from the clinically obtained T2-weighted sequence may be used as a complimentary tool to assist clinical decision making. Nevertheless, future prospective multicenter studies with larger samples will be needed to validate our study result and to optimize the prediction models for clinical practice.

AUTHOR CONTRIBUTIONS

XY, QP, HP, and WeiL conceived and designed the experiments. XY, QP, and BC wrote the paper. All authors analyzed the data and contributed to reagents, materials, analysis tools, and also read and approved the final manuscript.

FUNDING

This study was supported by the following: the Nature Scientific Foundations of China (81702956, 81602399), the Postdoctoral Science Foundation of Central South University (No.185705), the Strategy-Oriented Special Project of Central South University in China (ZLXD2017003), and the Beijing CSCO Fund (Y-MX2014-002).

ACKNOWLEDGMENTS

We thank Doctors Hongling Yin and Guanghui Gong, from the Department of Pathology (Xiangya Hospital, Central South University) for the detailed pathological analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00552/full#supplementary-material

REFERENCES

1. Bailey CE, Hu CY, You YN, Bednarski BK, Rodriguez-Bigas MA, Skibber JM, et al. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. *JAMA Surg.* (2015) 150:17–22. doi: 10.1001/jamasurg.2014.1756
2. Gu J, Chen N. Current status of rectal cancer treatment in China. *Colorectal Dis.* (2013) 15:1345–50. doi: 10.1111/codi.12269
3. Liu Z, Xiong L, Ouyang G, Ma L, Sahi S, Wang K, et al. Investigation of copper cysteamine nanoparticles as a new type of radiosensitiers for colorectal carcinoma treatment. *Sci Rep.* (2017) 7:9290. doi: 10.1038/s41598-017-09375-y
4. Sun PL, Li B, Ye QF. Effect of neoadjuvant cetuximab, capecitabine, and radiotherapy for locally advanced rectal cancer: results of a phase II study. *Int J Colorectal Dis.* (2012) 27:1325–32. doi: 10.1007/s00384-012-1446-2
5. Schmoll HJ, Van Cutsem E, Stein A, Valentini V, Glimelius B, Haustermans K, et al. ESMO Consensus Guidelines for management of patients with colon and rectal cancer. a personalized approach to clinical decision making. *Ann Oncol.* (2012) 23:2479–516. doi: 10.1093/annonc/mds236
6. Dayde D, Tanaka I, Jain R, Tai MC, Taguchi A. Predictive and prognostic molecular biomarkers for response to neoadjuvant chemoradiation in rectal cancer. *Int J Mol Sci.* (2017) 18: E573. doi: 10.3390/ijms18030573
7. Zorcolo L, Rosman AS, Restivo A, Pisano M, Nigri GR, Fancellu A, et al. Complete pathologic response after combined modality treatment for rectal cancer and long-term survival: a meta-analysis. *Ann Surg Oncol.* (2012) 19:2822–32. doi: 10.1245/s10434-011-2209-y
8. Park IJ, You YN, Agarwal A, Skibber JM, Rodriguez-Bigas MA, Eng C, et al. Neoadjuvant treatment response as an early response indicator for patients with rectal cancer. *J Clin Oncol.* (2012) 30:1770–6. doi: 10.1200/JCO.2011.39.7901
9. Ryan JE, Warrier SK, Lynch AC, Heriot AG. Assessing pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a systematic review. *Colorectal Dis.* (2015) 17:849–61. doi: 10.1111/codi.13081
10. Maas M, Nelemans PJ, Valentini V, Das P, Rödel C, Kuo LJ, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol.* (2010) 11:835–44. doi: 10.1016/S1470-2045(10)70172-8
11. Renehan AG, Malcomson L, Emsley R, Gollins S, Maw A, Myint AS, et al. Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. *Lancet Oncol.* (2016) 17:174–83. doi: 10.1016/S1470-2045(15)00467-2
12. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin Cancer Res.* (2016) 22:5256–64. doi: 10.1158/1078-0432.CCR-15-2997
13. Barbaro B, Vitale R, Valentini V, Illuminati S, Vecchio FM, Rizzo G, et al. Diffusion-weighted magnetic resonance imaging in monitoring rectal cancer response to neoadjuvant chemoradiotherapy. *Int J Radiat Oncol Biol Phys.* (2012) 83:594–9. doi: 10.1016/j.ijrobp.2011.07.017
14. Lu W, Jing H, Ju-Mei Z, Shao-Lin N, Fang C, Xiao-Ping Y, et al. Intravoxel incoherent motion diffusion-weighted imaging for discriminating the pathological response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Sci Rep.* (2017) 7:8496. doi: 10.1038/s41598-017-09227-9
15. Martens MH, Subhani S, Heijnen LA, Lambregts DM, Buijsen J, Maas M, et al. Can perfusion MRI predict response to preoperative treatment in rectal cancer. *Radiother Oncol.* (2015) 114:218–23. doi: 10.1016/j.radonc.2014.11.044
16. Ryan JE, Warrier SK, Lynch AC, Ramsay RG, Phillips WA, Heriot AG. Predicting pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a systematic review. *Colorectal Dis.* (2016) 18:234–46. doi: 10.1111/codi.13207
17. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2014) 5:4006. doi: 10.1038/ncomms5644
18. Al-Sukhni E, Attwood K, Mattson DM, Gabriel E, Nurkin SJ. Predictors of pathologic complete response following neoadjuvant chemoradiotherapy for rectal cancer. *Ann Surg Oncol.* (2016) 23:1177–86. doi: 10.1245/s10434-015-5017-y
19. Cui Y, Yang X, Shi Z, Yang Z, Du X, Zhao Z, et al. Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol.* (2019) 29:1211–20. doi: 10.1007/s00330-018-5683-9
20. Giannini V, Mazzetti S, Bertotto I, Chiarenza C, Cauda S, Delmastro E, et al. Predicting locally advanced rectal cancer response to neoadjuvant therapy with 18F-FDG PET and MRI radiomics features. *Eur J Nucl Med Mol Imaging.* (2019) 46:878–88. doi: 10.1007/s00259-018-4250-6
21. Bibault JE, Giraud P, Housset M, Durdux C, Taieb J, Berger A, et al. Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. *Sci Rep.* (2018) 8:12611. doi: 10.1038/s41598-018-35359-7
22. Wang G, He L, Yuan C, Huang Y, Liu Z, Liang C. Pretreatment MR imaging radiomics signatures for response prediction to induction chemotherapy in patients with nasopharyngeal carcinoma. *Eur J Radiol.* (2018) 98:100–6. doi: 10.1016/j.ejrad.2017.11.007
23. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
24. Horvat N, Veeraraghavan H, Khan M, Blazic I, Zheng J, Capanu M, et al. MR Imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy. *Radiology.* (2018) 287:833–43. doi: 10.1148/radiol.2018172300
25. Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett.* (2017) 403:21–7. doi: 10.1016/j.canlet.2017.06.004
26. Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: radiomics and breast cancer. *Breast Cancer Res Treat.* (2018) 169:217–29. doi: 10.1007/s10549-018-4675-4
27. Rödel C, Martus P, Papadoupolos T, Füzesi L, Klimpfinger M, Fietkau R, et al. Prognostic significance of tumor regression after preoperative chemoradiotherapy for rectal cancer. *J Clin Oncol.* (2005) 23:8688–96. doi: 10.1200/JCO.2005.02.1329
28. Szczypiński PM, Strzelecki M, Materka A, Klepaczko A. MaZda–a software package for image texture analysis. *Comput Methods Programs Biomed.* (2009) 94:66–76. doi: 10.1016/j.cmpb.2008.08.005
29. Das P, Skibber JM, Rodriguez-Bigas MA, Feig BW, Chang GJ, Wolff RA, et al. Predictors of tumor response and downstaging in patients who receive preoperative chemoradiation for rectal cancer. *Cancer.* (2007) 109:1750–5. doi: 10.1002/cncr.22625
30. Zhang S, Shi L, Mao D, Peng W, Sheng C, Ding C, et al. Use of jianpi jiedu herbs in patients with advanced colorectal cancer: a systematic review and meta-analysis. *Evid Based Complement Alternat Med.* (2018) 2018:6180810. doi: 10.1155/2018/6180810
31. Rossi L, Bijman R, Schillemans W, Aluwini S, Cavedon C, Witte M, et al. Texture analysis of 3D dose distributions for predictive modelling of toxicity rates in radiotherapy. *Radiother Oncol.* (2018) 129:548–53. doi: 10.1016/j.radonc.2018.07.027
32. Song X, Zhu H, Pei Q, Tan F, Li C, Zhou Z, et al. Significance of inflammation-based indices in the prognosis of patients with non-metastatic colorectal cancer. *Oncotarget.* (2017) 8:45178–89. doi: 10.18632/oncotarget.16774
33. Tan X, Wen Q, Wang R, Chen Z. Chemotherapy-induced neutropenia and the prognosis of colorectal cancer: a meta-analysis of cohort studies. *Expert Rev Anticancer Ther.* (2017) 17:1077–85. doi: 10.1080/14737140.2017.1380521
34. Formenti SC, Demaria S. Effects of chemoradiation on tumor-host interactions: the immunologic side. *J Clin Oncol.* (2008) 26:1562–3. doi: 10.1200/JCO.2007.15.5499
35. Ostheimer C, Gunther S, Bache M, Vordermark D, Multhoff G. Dynamics of heat shock protein 70 serum levels as a predictor of clinical response

in non-small-cell lung cancer and correlation with the hypoxia-related marker osteopontin. *Front Immunol.* (2017) 8:1305. doi: 10.3389/fimmu.2017.01305

36. McCawley N, Clancy C, O'Neill BD, Deasy J, McNamara DA, Burke JP. Mucinous rectal adenocarcinoma is associated with a poor response to neoadjuvant chemoradiotherapy: a systematic review and meta-analysis. *Dis Colon Rectum*. (2016) 59:1200–8. doi: 10.1097/DCR.0000000000000635
37. Patel SV, Roxburgh CS, Vakiani E, Shia J, Smith JJ, Temple LK, et al. Distance to the anal verge is associated with pathologic complete response to neoadjuvant therapy in locally advanced rectal cancer. *J Surg Oncol.* (2016) 114:637–41. doi: 10.1002/jso.24358

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 22 August 2019
doi: 10.3389/fonc.2019.00806

Radiomics-Based Machine Learning in Differentiation Between Glioblastoma and Metastatic Brain Tumors

Chaoyue Chen[1,2,3†], *Xuejin Ou*[1,2,4], *Jian Wang*[5], *Wen Guo*[1,2,3,4] *and Xuelei Ma*[1,2*†]

[1] Department of Biotherapy, Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [2] State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Collaborative Innovation Center for Biotherapy, Sichuan University, Chengdu, China, [3] Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu, China, [4] West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China, [5] School of Computer Science, Nanjing University of Science and Technology, Nanjing, China

Purpose: To investigative the diagnostic performance of radiomics-based machine learning in differentiating glioblastomas (GBM) from metastatic brain tumors (MBTs).

Method: The current study involved 134 patients diagnosed and treated in our institution between April 2014 and December 2018. Radiomics features were extracted from contrast-enhanced T1 weighted imaging (T1C). Thirty diagnostic models were built based on five selection methods and six classification algorithms. The sensitivity, specificity, accuracy, and area under curve (AUC) of each model were calculated, and based on these the optimal model was chosen.

Result : Two models represented promising diagnostic performance with AUC of 0.80. The first model was a combination of Distance Correlation as the selection method and Linear Discriminant Analysis (LDA) as the classification algorithm. In the training group, the sensitivity, specificity, accuracy, and AUC were 0.75, 0.85, 0.80, and 0.80, respectively; and in the testing group, the sensitivity, specificity, accuracy, and AUC of the model were 0.69, 0.86, 0.78, and 0.80, respectively. The second model was the Distance Correlation as the selection method and logistic regression (LR) as the classification algorithm, with sensitivity, specificity, accuracy, and AUC of 0.75, 0.85, 0.80, 0.80 in the training group and 0.69, 0.86, 0.78, 0.80 in the testing group.

Conclusion: Radiomic-based machine learning has potential to be utilized in differentiating GBM from MBTs.

Keywords: radiomics, machine learning, glioblastomas, metastatic brain tumors, texture analysis

OPEN ACCESS

Edited by:
Roger M. Bourne, University of Sydney, Australia

Reviewed by:
William Ian Duncombe Rae, University of Sydney, Australia
Zhongxiang Ding, Hangzhou First People's Hospital, China

***Correspondence:**
Xuelei Ma
drmaxuelei@gmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *13 June 2019*
Accepted: *07 August 2019*
Published: *22 August 2019*

Citation:
Chen C, Ou X, Wang J, Guo W and Ma X (2019) Radiomics-Based Machine Learning in Differentiation Between Glioblastoma and Metastatic Brain Tumors. Front. Oncol. 9:806. doi: 10.3389/fonc.2019.00806

INTRODUCTION

Glioblastomas (GBM) and metastatic brain tumors (MBTs) are commonly identified brain tumors in the adult population. Pre-surgery diagnosis between these lesions is critical to assist in efficient treatment planning, especially for MBTs with brain metastases detected before the primary tumor (1). Magnetic resonance imaging (MRI) is highly recommended for radiological examination as a non-invasive tool due to the advantage of identifying the location and size of lesions (2, 3).

However, conventional MR imaging is limited in differentiating GBM from solitary MBTs due to lacking characteristics on their imaging, and their contrast-enhancement patterns may mimic each other. Moreover, advanced MR techniques, like Dynamic Susceptibility Contrast Enhanced (DSC) MR imaging and proton magnetic resonance spectroscopy (HMRS), are not significant in the diagnosis of these lesions either given the similarities and the increased vascularity between these tumors or the metabolite ratios (4–8). Evidently, even with the quantitative information that individual MR techniques provided on specific properties of the tumor, the single radiological technique is not enough to provide a tumor characterization.

Considering MR data was able to reflect the pathophysiology of tumors visually, the quantitative radiomics-based analysis may provide a feasible solution to assist in the demanding process. Texture analysis (TA) is the mathematical method to calculate the voxel-intensity heterogeneity of images, including computed tomography (CT) and magnetic resonance imaging (MRI), and showed promising diagnostic ability in various lesions (9, 10). Previous studies have investigated the diagnostic ability of pattern recognition techniques combined with TA in order to aid physicians in making clinical decisions (3, 11, 12). However, the optimal diagnostic model is still controversial because the performance of models could be significantly different with various combinations of classification algorithms and the selection method on radiomics features. In the present study, we performed a radiomic-based machine learning method in discriminating GBM from MBTs with five selection methods and six classification algorithms to bring about the intuitional selection of an optimal model. Therefore, the purpose of our study was to assess the contribution of pattern recognition techniques using radiomics features in the different models to distinguish GBM from MBTs and to select the optimum one in terms of diagnostic value.

METHODS

Patient and MR Imaging Sequence Selection

This retrospective study was performed in our institution. The patients were selected from the neurosurgery department treated between April 2014 and December 2018. The initial selection enrolled potentially qualified patients who had records of intraoperative frozen-section confirmation on GBM or MBTs. Then we viewed the electronic medical records to collect the information we needed for analysis, including name, gender, age, and pathology report. Patients were excluded if the history of other types of intracranial diseases were documented or observed in MRI. The preoperative MR images were also collected from the radiological department through Picture Archiving and Communication Systems (PACS) (**Figure 1**).

In this study, we focused on conventional MR sequences, including T1-weighted imaging (T1WI), contrast-enhanced T1-weighted imaging (T1C), T2-weighted imaging (T2WI), and fluid attenuated inversion recovery (FLAIR), as they are the routine examination for patients with intracranial tumor. After the initial evaluation on images, contrast-enhanced T1 weighted-imaging (T1C) was chosen among all the sequences for further analysis due to the rather precise separation of tumor tissue from brain tissue.

Conventional MR Imaging Examination Protocols

The MR scans were performed using the 3.0T Siemens Trio Scanners in the MR Research Center. High-resolution 3-dimensional T1-weighted images were collected using MPRAGE sequence. The parameters were as follows: TR/TE/TI = 1,900/2.26/900 ms, 176 axial slices with thickness = 1 mm, axial FOV = 25.6 × 25.6 cm^2, Flip angle = 9°, and data matrix = 256 × 256. Dimeglumine (0.1 mmol/Kg) was the contrast agent for contrast-enhanced imaging, and multi-directional data of contrast-enhanced MRI were collected during the continuous interval time of 90–250 s.

Texture Feature Extraction

Two neurosurgeons participated in the statistic extraction of texture features using LifeX software (http://www.lifexsoft.org) with the assistance of senior radiologists. Following the software protocol, they drew along the whole lesion in each slice to obtain the 3D-texture features. In each layer of the image, the regions of interest (ROI) were carefully drawn along the boundary of tumor tissue (including the necrosis and vessels within tissue). The peritumoral edema band and adjacent structure invasion were separated from the primary tumor with the difference in contrast enhancement. After segmentation on the whole tumor, the software automatically calculated and extracted texture features with default protocols (**Figure 2**). To ensure the validity and reproducibility of the procedure, the surgeons conducted data extraction twice, and the difference between two sets was examined with Manny-Whitney *U*-test. We adjusted the $q < 0.01$ as significant (before was $p < 0.05$) to avoid the interference of false-positive errors rising from a large number of texture features. The results suggested that none of the features were significantly different, implying that the results could be reliable and reproducible (**Supplement Material 1**).

Texture features were calculated from two orders. In the first order, features on shape- and histogram-based matrixes were extracted; and in the second order, features on the gray-level co-occurrence matrix (GLCM), neighborhood gray-level dependence matrix (NGLDM), gray-level zone length matrix (GLZLM), and gray-level run length matrix (GLRLM) were extracted. Finally, we built a statistical dataset of the radiomic statistics consisting of 43 features for machine-learning analysis.

Classification Procedure

The establishment on the diagnostic model involved two parts: feature selection and classification algorithm (or known as classifier) deployment. The feature selection serviced the purpose that the numbers of features were so many that overfitting was inevitable for classification of algorithms. Considering the

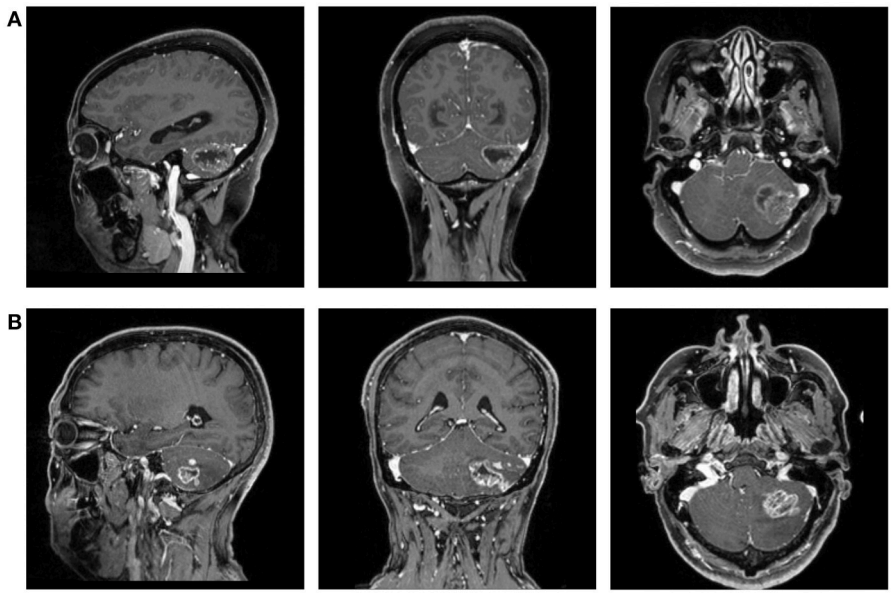


FIGURE 1 | The magnetic resonance images (T1C) of a patient with **(A)** GBM and **(B)** MBTs.

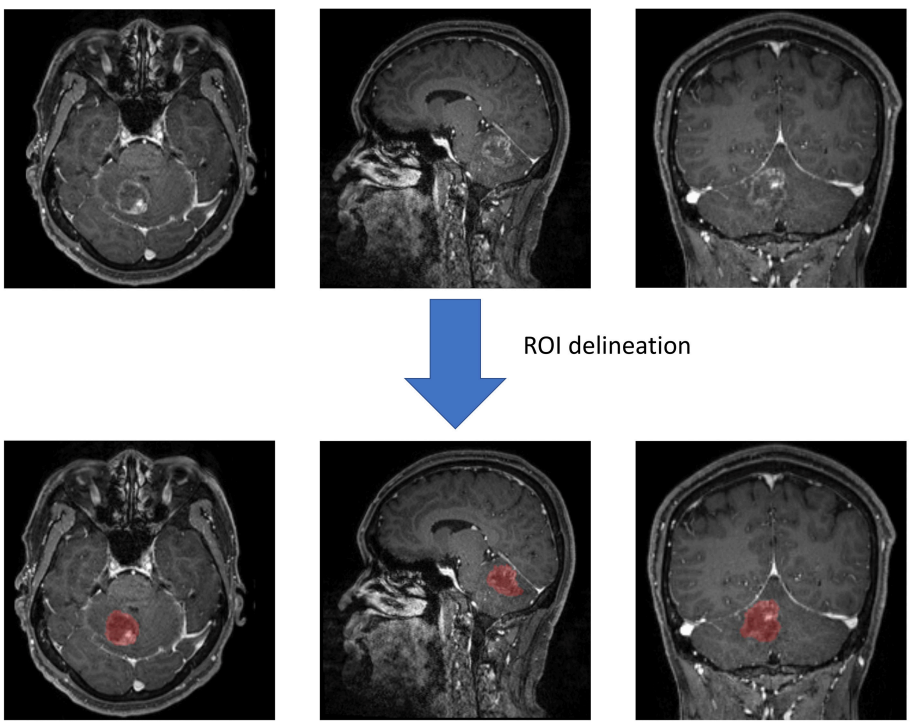


FIGURE 2 | Screen capture of regions of interest (ROI) delineation.

optimal selection method could be different for algorithms, five selection methods were evaluated in our study, including distance correlation, random forest (RF), least absolute shrinkage and selection operator (LASSO), eXtreme gradient boosting (Xgboost), and Gradient Boosting Decision Tree (GBDT). The selected features were adopted into classification algorithms to establish models.

Six classification algorithms were evaluated in our study, including Linear Discriminant Analysis (LDA, also known as Fisher Linear Discriminant), Support Vector Machine (SVM),

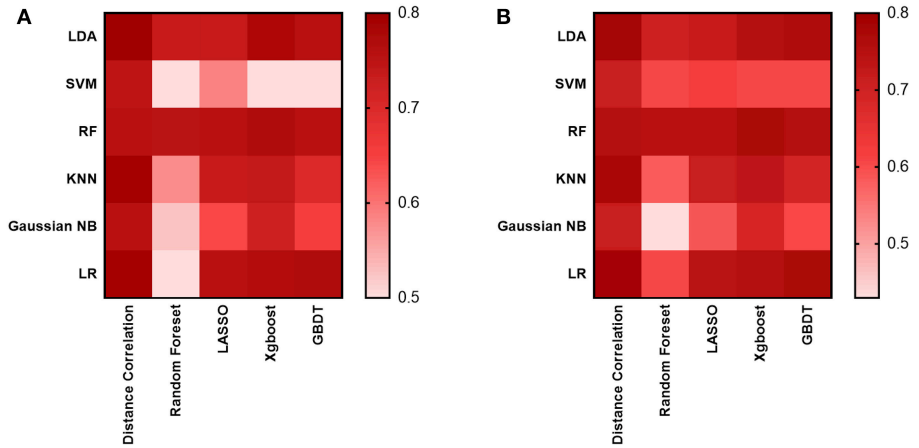


FIGURE 3 | Heat map of the classifiers for differentiating between GBM and MBTs. **(A)** The AUC of the training group. **(B)** The AUC of the testing group.

TABLE 1 | Results of the optimal discriminative model in distinguishing GBM from MBTs in the training and the testing groups.

Model	Training group				Testing group			
	AUC	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity
Distance correlation + LDA	0.80	0.80	0.75	0.85	0.80	0.78	0.69	0.86
Distance correlation + LR	0.83	0.83	0.79	0.87	0.80	0.79	0.71	0.85

AUC, area under curve; LDA, linear discriminant analysis; LR, Logistic Regression.

random forest (RF), k-nearest neighbor (KNN), GaussianNB, and logistic regression (LR). Patients were divided as the training group and the testing group on a proportion of 4:1. Area under the receiver operating characteristic curve (AUC) of each model was calculated to assess their diagnostic performance. For each model, the progress of machine learning was repeated over 100 times to obtain the realistic distribution of classification accuracies.

All procedures involving human participants were in accordance with the ethical standards of the institutional and/or national research committee. The Ethics Committee of Sichuan University approved this retrospective study. The written informed consent was necessary before radiological examination (written informed consent for patients <16 years old was signed by parents or guardians) for all patients. They agreed to undertake examination if needed and were informed that the statistics (including MR image) might be used for academic purposes in the future.

RESULT

Patients Selection

A total number of 134 patients were enrolled in this study. Seventy-six of the patients were diagnosed with GBM, and 58 of them were diagnosed with MBTs. The average ages of patients were 46.9 and 57.6, respectively. The gender ratio for each type of tumor (Male: Female) was 10:9 and 9:5, respectively. The pathology reports represented that the majority of MBTs were originated from lung cancer and breast cancer ($N = 54$).

Diagnostic Performance of Models

As for the diagnostic models we evaluated, 30 models were established to select the suitable one, which was defined as the one with the highest AUC in the testing group. The results suggested the AUC of models mostly hovered around between 0.70 and 0.76 (**Figure 3**), and the highest value was 0.80 observed in two models: the Distance Correlation + LDA and the Distance Correlation + LR (**Table 1**). The details of each model performance are summarized in **Supplement Material 2**.

For the first model (the Distance Correlation + LDA), in the training group, the sensitivity, specificity, accuracy, and AUC of the model were 0.75, 0.85, 0.80, and 0.80, respectively. And in the testing group, the sensitivity, specificity, accuracy, and AUC of the model were 0.69, 0.86, 0.78, and 0.80. For the second model (the Distance Correlation + LR) in the training group, the sensitivity, specificity, accuracy, and AUC of the model were 0.79, 0.87, 0.83, and 0.83, respectively. And in the testing group, the sensitivity, specificity, accuracy, and AUC of the model were 0.71, 0.85, 0.79, and 0.80, respectively. The LDA distribution suggested these two models represented similar diagnostic performance (**Figure 4**). **Figure 5** shows one example of 100 independent validation cycles of the model, representing the distribution of the first and second direct LDA canonical functions.

DISCUSSION

In the present study, we investigated the diagnostic ability of pattern recognition techniques combined with texture features extracted from conventional MRI in discriminating GBM from MBTs. MRI could provide excellent information on soft tissue differentiation to enable the exact localization of the tumors and to assist in the prediction of tumor response to therapy evaluation (13). However, pathological identification is the weakness of conventional MRI bringing additional advanced imaging techniques, which required additional fees and equipment, into tumor characterization and treatment. Our study made the evaluation on six classification algorithms consisting of five selection methods and six classification algorithms to identify the optimal model.

The diagnosis between MBTs and GBM on conventional MRI is rather straightforward because of the clinical history or observation of multiple lesions. The differences in tumor growth could lead to characteristic descriptions that GBM usually extends by infiltration, while MBTs usually arise within the brain parenchyma and grow by expansion, leading to comprising surrounding brain tissue (14). However, the emergence of lesions with a solitary enhancing appearance lacking information on primary tumors brings difficulty on differential diagnosis because high-grade GBM can present similar contrast enhancement patterns (15). The accurate and early diagnosis of these lesions is clinically important because the surgical planning, medical staging, and therapeutic approach can significantly vary from each other. Given that MR scan is the conventional radiological examination for patients, TA on T1C has the potential to serve as a feasible solution in clinical application without requiring additional fees. Previous studies have illustrated that TA combined with machine learning could assist in the diagnosis of various brain tumors, such as GBM from primary central nerve system lymphoma and meningioma from GBM (16, 17). Moreover, it has also been applied in tumor grade system and gene mutation prediction (18–22). The researchers illustrated the potential of artificial intelligence in lightening the clinical workload and improving early diagnostic accuracy.

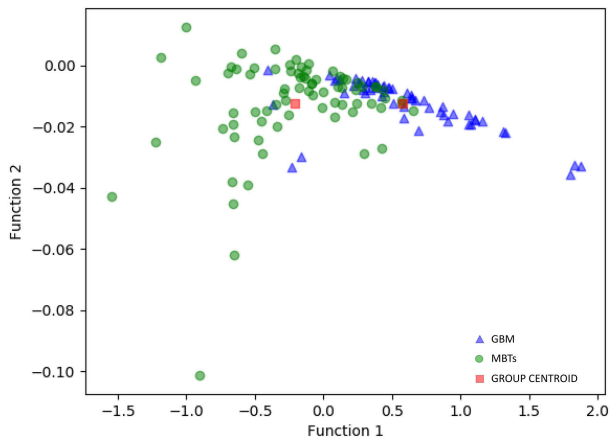


FIGURE 4 | Distribution of the discriminant functions of LDA in discriminating GBM from MBTs.

Compared with the previous studies, our study involved various selection methods and classification algorithms to choose the optimal model with the best performance. Thirty models were evaluated, and two of them represented feasible diagnostic ability with AUC of 0.80 (the Distance Correlation + LDA and the Distance Correlation + LR). In the previous study, the SVM classifier was usually proven to be the suitable classifier compared to the others, which made sense considering that SVM is the suitable algorithm for small sample size. Our study illustrated that the feasible optimal classifiers were LDA and LR, while overfittings were almost observed in all SVM-based models (**Supplement Material 2**). LDA and LR are considered as the state-of-the-art on pattern recognition classifiers, with much better performance in some cases. LDA is also taken as the ground truth number of parameters in terms of performance. The mechanisms of classifiers provide a possible explanation of the differences in results. Both LDA and LR are the linear classifiers, while SVM is the non-linear classifier. The main difference of two types of classifiers consists in the shape of the decision boundary: plane or straight line in the first case, and surface or curved line in the second case. The choice of classification algorithm should be a tradeoff between computational burden

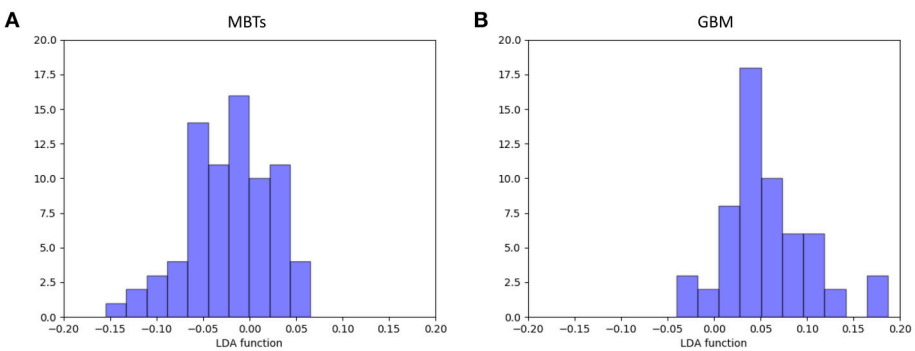


FIGURE 5 | Example of distributions of the LDA function of **(A)** MBTs and **(B)** GBM for one cycle.

and performance (23). This theory also demonstrated why SVM could be the suitable algorithm for a small sample size (50∼60) while LDA/LR was suitable for a relatively large sample size (>100). However, it is worth noting that the diagnostic performances of classifiers did not improve much in the current research, even with the change in classification algorithm. All studies applying machine learning in discrimination of MBTs from GBM represented similar diagnostic performance with AUC in the testing group of ∼0.80, even when radiomics features were selected with various selection methods and extracted from various sequences (11, 12, 24). More research is required to verify our results and to investigate the algorithm with better diagnostic performance.

There were some limitations in the current study. First and foremost, this study was a single central, retrospective study, bringing inevitable selection bias (**Supplement Material 3**). Second, the inhomogeneous histological subcategories of MBTs could reduce the accuracy in the differentiation. Future investigations with a larger sample size are required to assess the ability of classification algorithms and texture parameters in characterizing the lesion subtype. Third, only texture features retrieved from T1C images were adapted into classifiers, while features from other sequences (like T2WI and FLAIR) and advanced MR techniques were not explored. Fourth, the models were not validated in the other dataset, and we cannot guarantee the diagnostic ability of our models for external datasets due to the various protocols of imaging acquisition and MR scanners. However, the analysis protocol and image processing procedure were open-source packages and they should be validated and reproduced.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

XM participated in the conceptualization and revised intellectual content in the manuscript. CC collected MR image, participated in MRI features extraction, and drafted this manuscript. XO collected MR image, participated in MRI features extraction. JW deployed the machine-learning algorism and responsible for statistical analysis. WG assisted in MRI features extraction.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00806/full#supplementary-material

Supplement Material 1 | The repeatability of texture feature extraction examined by Mann-Whitney *U*-test.

Supplement Material 2 | The diagnostic performance of each model.

Supplement Material 3 | The radiomics quality score of research.

REFERENCES

1. Chiang IC, Kuo YT, Lu CY, Yeung KW, Lin WC, Sheu FO., et al. Distinction between high-grade gliomas and solitary metastases using peritumoral 3-T magnetic resonance spectroscopy, diffusion, and perfusion imagings. *Neuroradiology*. (2004) 46:619–27. doi: 10.1007/s00234-004-1246-7
2. Earnest F, Kelly PJ, Scheithauer BW, Kall BA, Cascino TL, Ehman RL. et al. Cerebral astrocytomas: histopathologic correlation of MR and CT contrast enhancement with stereotactic biopsy. *Radiology*. (1988) 166:823–7. doi: 10.1148/radiology.166.3.2829270
3. Devos A, Simonetti AW, van der Graaf M, Lukas L, Suykens JA, Vanhamme L, et al. The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumor classification. *J Magn Reson*. (2005) 173:218–28. doi: 10.1016/j.jmr.2004.12.007
4. Korfiatis P, Erickson B. Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas. *Clin Radiol*. (2019) 74, 367–73. doi: 10.1016/j.crad.2019.01.028
5. Lohmann P, Werner JM, Shah NJ, Fink GR, Langen KJ, Galldiks N. Combined amino acid positron emission tomography and advanced magnetic resonance imaging in glioma patients. *Cancers*. (2019) 11:153. doi: 10.3390/cancers11020153
6. Law M, Cha S, Knopp EA, Johnson G, Arnett J, Litt AW. High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. *Radiology*. (2002) 222:715–21. doi: 10.1148/radiol.2223010558
7. Liu X, Tian W, Kolar B, Yeaney GA, Qiu X, Johnson MD, et al. MR diffusion tensor and perfusion-weighted imaging in preoperative grading of supratentorial nonenhancing gliomas. *Neuro Oncol*. (2011) 13:447–55. doi: 10.1093/neuonc/noq197
8. Fan G, Sun B, Wu Z, Guo Q, Guo Y. *In vivo* single-voxel proton MR spectroscopy in the differentiation of high-grade gliomas and solitary metastases. *Clin Radiol*. (2004) 59:77–85. doi: 10.1016/j.crad.2003.08.006
9. Kassner A, Thornhill RE. Texture analysis: a review of neurologic MR imaging applications. *AJNR Am J Neuroradiol*. (2010) 31:809–16. doi: 10.3174/ajnr.A2061
10. Dennie C, Thornhill R, Sethi-Virmani V, Souza CA, Bayanati H, Gupta A, et al. Role of quantitative computed tomography texture analysis in the differentiation of primary lung cancer and granulomatous nodules. *Quant Imaging Med Surg*. (2016) 6:6–15. doi: 10.3978/j.issn.2223-4292.2016.02.01
11. Svolos P, Tsolaki E, Kapsalaki E, Theodorou K, Fountas K, Fezoulidis I, et al. Investigating brain tumor differentiation with diffusion and perfusion metrics at 3T MRI using pattern recognition techniques. *Magn Reson Imaging*. (2013) 31:1567–77. doi: 10.1016/j.mri.2013.06.010
12. Tsolaki E, Svolos P, Kousi E, Kapsalaki E, Fountas K, Theodorou K, et al. Automated differentiation of glioblastomas from intracranial metastases using 3T MR spectroscopic and perfusion data. *Int J Comput Assist Radiol Surg*. (2013) 8:751–61. doi: 10.1007/s11548-012-0808-0
13. Provenzale JM, Mukundan S, Barboriak DP. Diffusion-weighted and perfusion MR imaging for brain tumor characterization and assessment of treatment response. *Radiology*. (2006) 239:632–49. doi: 10.1148/radiol.2393042031
14. Cha S. Update on brain tumor imaging: from anatomy to physiology. *AJNR Am J Neuroradiol*. (2006) 27:475–87.
15. Oh J, Cha S, Aiken AH, Han ET, Crane JC, Stainsby JA, et al. Quantitative apparent diffusion coefficients and T2 relaxation times in characterizing contrast enhancing brain tumors and regions of peritumoral edema. *J Magn Reson Imaging*. (2005) 21:701–8. doi: 10.1002/jmri.20335
16. Nakagawa M, Nakaura T, Namimoto T, Kitajima M, Uetani H, Tateishi M, et al. Machine learning based on multi-parametric magnetic resonance imaging to differentiate glioblastoma multiforme from primary

cerebral nervous system lymphoma. *Eur J Radiol*. (2018) 108:147–54. doi: 10.1016/j.ejrad.2018.09.017
17. Nguyen AV, Blears EE, Ross E, Lall RR, Ortega-Barnett J. Machine learning applications for the differentiation of primary central nervous system lymphoma from glioblastoma on imaging: a systematic review and meta-analysis. *Neurosurg Focus*. (2018) 45:E5. doi: 10.3171/2018.8.FOCUS18325
18. Zhang X, Yan LF, Hu YC, Li G, Yang Y, Han Y, et al. Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features. *Oncotarget*. (2017) 8:47816–30. doi: 10.18632/oncotarget.18001
19. Li Y, Qian Z, Xu K, Wang K, Fan X, Li S, et al. MRI features predict p53 status in lower-grade gliomas via a machine-learning approach. *Neuroimage Clin*. (2018) 17:306–11. doi: 10.1016/j.nicl.2017.10.030
20. Zacharaki EI, Wang S, Chawla S, Soo Yoo D, Wolf R, Melhem ER, et al. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn Reson Med*. (2009) 62:1609–18. doi: 10.1002/mrm.22147
21. Zarinabad N, Wilson M, Gill SK, Manias KA, Davies NP, Peet AC. Multiclass imbalance learning: Improving classification of pediatric brain tumors from magnetic resonance spectroscopy. *Magn Reson Med*. (2017) 77:2114–24. doi: 10.1002/mrm.26318
22. Takada M, Sugimoto M, Masuda N, Iwata H, Kuroi K, Yamashiro H, et al. Prediction of postoperative disease-free survival and brain metastasis for HER2-positive breast cancer patients treated with neoadjuvant chemotherapy plus trastuzumab using a machine learning algorithm. *Breast Cancer Res Treat*. (2018) 172:611–8. doi: 10.1007/s10549-018-4958-9
23. Dellacasa Bellingegni A, Gruppioni E, Colazzo G, Davalli A, Sacchetti R, Guglielmelli E, et al. NLR, MLP, SVM, and LDA: a comparative analysis on EMG data from people with trans-radial amputation. *J Neuroeng Rehabil*. (2017) 14:82. doi: 10.1186/s12984-017-0290-6
24. García-Gómez JM, Luts J, Julià-Sapé M, Krooshof P, Tortajada S, Robledo JV, et al. Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. *Magma*. (2009) 22:5–18. doi: 10.1007/s10334-008-0146-y

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 30 August 2019
doi: 10.3389/fonc.2019.00821

CT-Based Radiomics Signature for the Preoperative Discrimination Between Head and Neck Squamous Cell Carcinoma Grades

Wenli Wu[1], *Junyong Ye*[2], *Qi Wang*[3], *Jin Luo*[2] *and Shengsheng Xu*[1]*

[1] *Department of Radiology, The First Affiliated Hospital, Chongqing Medical University, Chongqing, China,* [2] *Key Laboratory of Optoelectronic Technology and Systems of the Ministry of Education, Chongqing University, Chongqing, China,* [3] *Department of Information, The First Affiliated Hospital, Chongqing Medical University, Chongqing, China*

Background: Radiomics has been widely used to non-invasively mine quantitative information from medical images and could potentially predict tumor phenotypes. Pathologic grade is considered a predictive prognostic factor for head and neck squamous cell carcinoma (HNSCC) patients. A preoperative histological assessment can be important in the clinical management of patients. We applied radiomics analysis to devise non-invasive biomarkers and accurately differentiate between well-differentiated (WD) and moderately differentiated (MD) and poorly differentiated (PD) HNSCC.

Methods: This study involved 206 consecutive HNSCC patients (training cohort: $n = 137$; testing cohort: $n = 69$). In total, we extracted 670 radiomics features from contrast-enhanced computed tomography (CT) images. Radiomics signatures were constructed with a kernel principal component analysis (KPCA), random forest classifier and a variance-threshold (VT) selection. The associations between the radiomics signatures and HNSCC histological grades were investigated. A clinical model and combined model were also constructed. Areas under the receiver operating characteristic curves (AUCs) were applied to evaluate the performances of the three models.

Results: In total, 670 features were selected by the KPCA and random forest methods from the CT images. The radiomics signatures had a good performance in discriminating between the two cohorts of HNSCC grades, with an AUC of 0.96 and an accuracy of 0.92. The specificity, accuracy, sensitivity, positive predictive value (PPV), and negative predictive value (NPV) of the abovementioned method with a VT selection for determining HNSCC grades were 0.83, 0.92, 0.96, 0.94, and 0.91, respectively; without VT, the corresponding results were 0.70, 0.83, 0.88, 0.80, and 0.84. The differences in accuracy, sensitivity and NPV were significant between these approaches ($p < 0.05$). The AUCs with VT and without VT were 0.96 and 0.89, respectively ($p < 0.05$). Compared to the combined model and the radiomics signatures, The clinical model had a worse performance, and the differences were significant ($p < 0.05$). The combined model had the best performance, but the difference between the combined model and the radiomics signature weren't significant ($p > 0.05$).

OPEN ACCESS

Edited by:
Chunxiao Guo,
University of Texas MD Anderson Cancer Center, United States

Reviewed by:
Zhongxiang Ding,
Hangzhou First People's Hospital, China
Seyedmehdi Payabvash,
Yale University School of Medicine, United States

***Correspondence:**
Shengsheng Xu
xuss@cqmu.edu.cn

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *29 May 2019*
Accepted: *09 August 2019*
Published: *30 August 2019*

Citation:
Wu W, Ye J, Wang Q, Luo J and Xu S (2019) CT-Based Radiomics Signature for the Preoperative Discrimination Between Head and Neck Squamous Cell Carcinoma Grades. Front. Oncol. 9:821. doi: 10.3389/fonc.2019.00821

Conclusions: The CT-based radiomics signature could discriminate between WD and MD and PD HNSCC and might serve as a biomarker for preoperative grading.

Keywords: head and neck cancer, grade, computed tomography, radiomics signature, biomarker

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common malignant tumor worldwide. Many factors affect the prognosis of patients with HNSCC; among these factors, the histological differentiation grade was reported to correlate with lymph node status, distant metastases, survival and prognosis (1–4). A pretreatment histopathologic grade evaluation for HNSCC provides information for clinical decision making. Although the histological differentiation grade is routinely confirmed by biopsy and surgical resection in many head and neck cancer centers, invasive biopsy is sometimes of little predictive value in early-stage oral SCC (5). In addition, intra-tumor heterogeneity is an issue. Biopsies do not exactly reflect the overall pathophysiology of the lesion.

Some non-invasive functional imaging modalities have been developed in the clinic, such as diffusion-weighted imaging (DWI), dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), perfusion-weighted imaging (PWI), and positron emission tomography (PET), all of which have been applied in the grading of HNSCC (6–8). These imaging modalities play important roles in the evaluation of disease grade to some extent, but combining clinical visual assessments is necessary to increase the overall accuracy.

Radiomics, which refers to an enhanced deep analysis of the molecular aspects of tumors and accounts for intrinsic susceptibility in the long-term follow-up, is a qualitative and quantitative analysis of a large amount of radiologic data extracted in a high-throughput manner to obtain predictive or prognostic information from cancer patients (9, 10). Radiomics is suitable for providing some predictive, classifying, and prognostic information for HNSCC patients (11–13). A few radiomics studies have been conducted based on MRI regarding the staging and grading of HNSCC (14–16). Although the vast majority of radiomics analyses were conducted on CT images, no studies exist about radiomics models based on CT signatures to differentiate HNSCC grades.

A large number of machine-learning methods were used to evaluate their applying values in HNSCC patients (17, 18). In this study, we will use another analysis method based on CT radiomics signatures to evaluate its predictive value in differentiating between HNSCC grades (WD vs. MD/PD).

MATERIALS AND METHODS

Study Population

We collected patients with head and neck tumors confirmed to be SCC by surgical pathology in our hospital from January 2012 to February 2018. This study was approved by the institutional review board of our hospital (approval number 2019-178), and informed consent was waived. All patients underwent both precontrast and multiple-phase pretreatment contrast enhanced multi-slice spiral computed tomography (MSCT) scans. In this study, the patients were chosen and excluded according to the criteria presented in **Figure 1**. A total of 206 consecutive patients were identified met the criteria. These patients were randomly divided into a training cohort and a testing cohort at a ratio of 9:1 by a computer. We retrospectively analyzed the clinical information of all patients, including race, age, sex, tumor sites, tumor differentiation, tumor node metastasis (TNM) classification, and stage.

CT Image Acquisition

All CT scans were performed using a GE Discovery 750 HD (GE Healthcare, Milwaukee, WI, USA) multidetector CT scanner. The CT scanning area was from the skull base down to the thorax inlet. The scanning parameters were as follows: 120 kV; 80 mA; pitch 0.984; detector collimation, 64 × 0.625 mm; rotation time, 0.6 s; matrix, 512 × 512; section thickness, 5 mm; and field of view, 220–250 × 220–250 mm. First, a non-contrast enhanced

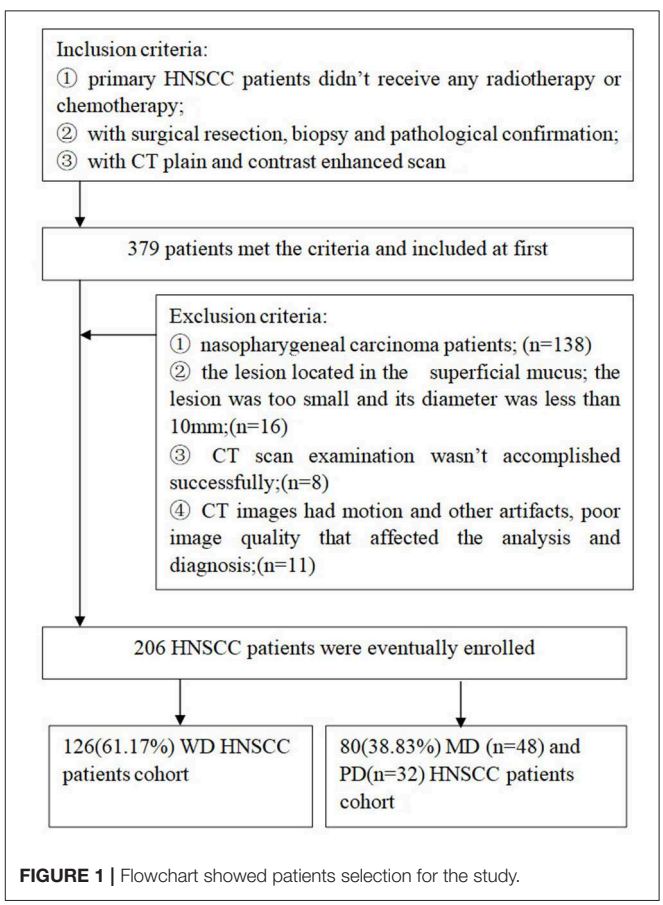


FIGURE 1 | Flowchart showed patients selection for the study.

CT scan was performed, and then a contrast-enhanced CT scan was performed in the arterial phase (25–30 s), portal venous phase (60–65 s), and delayed phase (120 s), after an intravenous injection of non-ionic iodinated contrast medium (Ultravist 370, Bayer Schering Pharma, Berlin, Germany) (dose 1.5 mL/kg, injection rate 3.5 mL/s).

Image Analysis

Preprocessing

Lesion segmentation and labeling

We segmented and labeled the lesions on picture archiving and communication systems (PACS) (Carestream Health Inc., Rochester, NY, USA). First, the doctor's terminal was opened to view the enrolled patients' CT images, especially the portal venous phase contrast-enhanced images, and then the slice on which the lesion was the most obviously displayed was determined. Second, the image window width was adjusted to 350 Hounsfield units (HU), and the window level was adjusted to 40 HU. Third, the curve measurement button on the toolbar was clicked to set the line color to red. Then, the largest solid part of the tumor was encircled to include the markedly enhanced area and excluded the necrotic or cystic areas. The enclosed area was recognized as a region of interest (ROI) and could be round, oval or another irregular shape. The area of the ROI was more than 1 cm^2, which guaranteed a large enough area for analysis. Finally, the image with the ROI was exported and saved in a JPG format for subsequent processing. The segmentation and labeling processes were performed by two head and neck radiologists (J.F. and Y.T. with 5 and 10 years of diagnostic experience, respectively). Any discrepancies that occurred were resolved by a consensus between the two radiologists.

The goal of preprocessing was to delineate tumor regions, but first, the coordinates of the tumor area needed to be detected. Because the CT image was almost gray, the red line could easily be detected by a sliding a 64 × 64 rectangle to scan the whole image from the left top with step size of 1. This sliding rectangle recorded the coordinates of the vertex as soon as the rectangle came into contact with the closed red line.

We used a 64 × 64 window to scan the whole image with a step size of 1. Once the closed red line was found, the scanning process was stopped. Since we used the red line to contour the tumor, the window had a 100% overlap with the tumor at this time. The segmentation process was performed by two head and neck radiologists (W.X. and C.Y. with 8 and 11 years of diagnostic experience, respectively). We used an original non-annotated image in case the annotated red line interfered with the prediction. The coordinates recorded by the sliding rectangle could help delineate a 64 × 64 tumor region on the original image. These delineated images are called patches. Only the tumor region was considered when discriminating WD HNSCC from MD/PD HNSCC so that we could focus on the tumor and reduce the amount of noise interference. In addition, compared to a complete tumor region, a 64 × 64 patch contained some tissues around the tumor, which could also contribute to the tumor grade.

To extract the shape features, we need an additional mask to describe the shape of the tumor. We extracted the edge of the

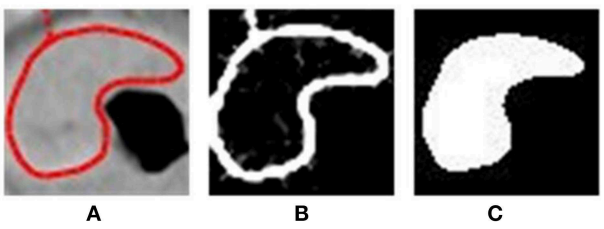


FIGURE 2 | Steps of preprocessing: **(A)** cutting off the patches of ROI; **(B)** detecting the edge; **(C)** fulfilling the edge and generating mask.

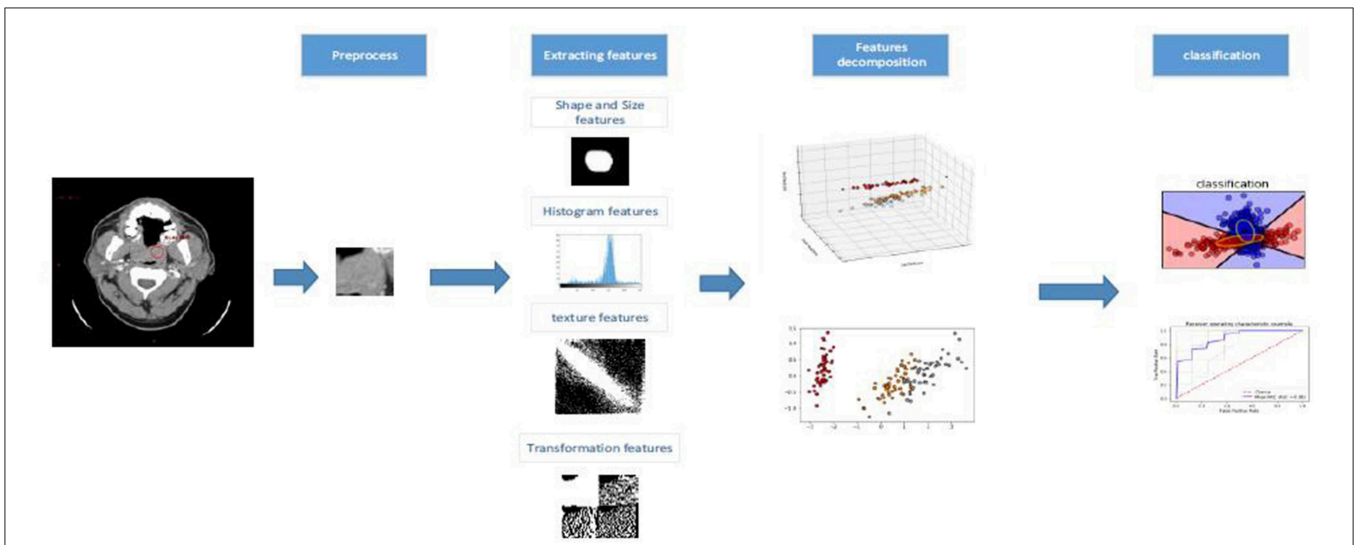


FIGURE 3 | The workflow of proposed kernelized radiomics model in HNSCC.

patch, primarily by keeping only the red parts of the image and then filling in the edge and erasing the small annotated area to generate a mask. We used both the segmented patches and masks to extract all features (**Figure 2**).

Extracting radiomics features

We extracted 670 radiomics features from the portal venous phase contrast-enhanced CT images. These features quantified the phenotypic HNSCC characteristics and were divided into four feature groups: shape and size features; histogram features; texture features; and transformation features. All features are shown in **Tables S1–S4**, and we used all features to construct the random forest model. The workflow of the radiomics analysis is shown in **Figure 3**. The preprocessing and feature extraction methods were coded in MATLAB and python using scikit-image (19).

Feature decomposition and classification

Although 670 features were extracted for each patient, these features did not contribute equally to discriminating between WD HNSCC and MD/PD HNSCC. The features with low discrimination capabilities or those highly correlated with each other would overfit the classifiers and lead to a poor outcome. Therefore, feature decomposition was performed to find a set of candidate features with excellent discrimination capabilities and significant differences before grade prediction. In addition to feature selection, feature decomposition could also generate new features that are more capable of discrimination and have less correlation with each other than the original features. We used a non-linear kernelization method in the analysis. KPCA, which could be seen as a non-linear version of PCA, is a perfect answer to non-linear requests. In this paper, the following radial basis function (RBF) kernel was used:

$$k_{RBF}\left(x_i, x_j\right) = exp(-\frac{\left\| x_i - x_j \right\|^2}{2\sigma^2}). \quad (1)$$

Then, the features extracted from the CT image could be processed by the KPCA algorithm with a RBF kernel. The decomposition and classification methods were implemented using scikit-learn (20), followed by a random forest classifier, and we finally obtained our proposed kernelized radiomics model. All experiments were performed under a Windows OS on a machine with CPU Intel Xeon E5 2687W V3, GPU NVIDIA GeForce 1080ti, and 16*8GB of RAM.

Kernelized radiomics model building

To build our kernelized radiomics model, we first decided on the dimensions of the kernelized features. When using the RBF kernel, we tuned the dimension value from 30 to 200 with steps of 10.

Because a little imbalance existed between the positive and negative samples in our dataset, AUC, instead of accuracy, was used to select the dimension value.

Since the dimension of the kernelized features had been decided, we still needed to select the classifier parameters. For an ensemble learning method using random feature selection, the main factors that could affect the performance of the random forest model are the number of basic learners (decision tree), maximum depth of each decision tree and number of randomly selected features. We used a gidsearch to search for the best values of these parameters, which tuned one parameter while freezing the others.

We removed features with a training set variance lower than 0.8. We used the python and sklearn library to implement this method, which first calculated the variance of each feature and then removed features with a low variance.

Clinical and combined model building

According to previous studies (1, 4, 21–25), some clinical and radiological characteristics are related to the differentiation grades of HNSCC. The TN classification, stage and enhancement types were selected as the clinical parameters for clinical model building (**Supplementary Data Sheet 2**). These clinical

TABLE 1 | HNSCC patients information and tumor characteristics in the study.

Information/ characteristic	Testing cohort	Training cohort	*p*-value
Age	63.57 ± 12.01 (31–87)	61.18 ± 11.87 (27–86)	0.18
Sex			0.74
Male	53 (76.8%)	108 (78.8%)	
Female	16 (23.2%)	29 (21.2%)	
Tumor primary location			0.45
Oral cavity	35 (50.7%)	71 (51.8%)	
Oropharynx	12 (17.4%)	13 (9.5%)	
Hypoharynx	12 (17.4%)	28 (20.4%)	
Larynx	10 (14.5%)	22 (16.1%)	
Others	0	3 (2.2%)	
Tumor differentiation			0.95
WD	42 (60.9%)	84 (61.3%)	
MD/PD	27 (39.1%)	53 (38.7%)	
T classification			0.64
T1–2	19 (27.5%)	42 (30.7%)	
T3–4	50 (72.5%)	95 (69.3%)	
N classification			0.52
N0	38 (55.1%)	69 (51.1%)	
N+	31 (44.9%)	68 (48.9%)	
Stage			0.79
I–II	14 (20.3%)	30 (21.9%)	
III–IV	55 (79.7%)	107 (78.1%)	
Enhancement types			
Observer 1			0.70
Homogeneous 1	23 (33.3%)	42 (30.7%)	
Heterogeneous 1	46 (66.7%)	95 (69.3%)	
Observer 2			0.23
Homogeneous 2	22 (31.9%)	33 (24.1%)	
Heterogeneous 2	47 (68.1%)	104 (75.9%)	

Age data are mean ± standard deviation, age range in parentheses, other data are number (percentage). P > 0.05.

and radiological characteristics and radiomics signatures were integrated to build the combined model.

Statistical Analysis

The discriminating performance of this model was evaluated with ROC curves and AUCs, and these values were compared using DeLong tests. The differences in clinical characteristics between the training and validation sets were evaluated using Student's *t*-tests and chi-square tests, and a $p < 0.05$ was considered statistically significant. IBM SPSS software ver. 24 (IBM Corp., Armonk, NY, USA) and open-source machine learning studio were used for statistical analysis. The inter-observer agreement in evaluating the enhancement types (homogeneous/heterogeneous) was assessed with kappa statistics: a kappa value between 0.00 and 0.20 indicates a slight agreement; a value between 0.21 and 0.40 indicates a fair agreement; a value between 0.41 and 0.60 indicates a moderate agreement; a value between 0.61 and 0.80 indicates a substantial agreement; and a value between 0.81 and 1.00 indicates an almost perfect agreement.

RESULTS

Patient Population Information and Tumor Characteristics

The clinical information of the patients and HNSCC characteristics in this study are summarized in **Table 1**. The testing cohort included 69 patients (53 males and 16 females). The training cohort included 137 patients (108 males and 29 females). All patients were Chinese, with no patients who were white, black or of other races. Regarding the tumor TNM classifications, only two patients were classified as M1, and the others were classified as M0; therefore, we did not conduct statistical assessments on the M stage. There were no differences between the training and testing cohorts in terms of age, sex, tumor primary location, histological differentiation, TN classification, stage or enhancement types ($p > 0.05$).

The *p*-value of the kappa statistics analysis was 0.000 ($p < 0.05$), indicating that inter-observer agreement existed. The kappa value was 0.510 [95% CI (confidence interval, CI) 0.379–0.642]. The degree of inter-observer agreement regarding enhancement types was moderate.

After the parameters were finished tuning, a dimension of 130 corresponded to the biggest AUC (AUC = 0.97). Therefore, we obtained a 130-dimensional vector after kernelizing the features of the sample (**Figure 4**). We built our kernelized model, which used KPCA with a kernelized dimension of 130 as a feature decomposer and random forest classifier, because these parameter values led to the best model performance in terms of AUC.

We obtained the top two features: smoothness and GLCM_t_45_d_1_Con_2. There were significant differences between the WD and MD/PD HNSCC cohorts ($p < 0.05$).

Performance of the Models

On the basis of VT selection, which eliminated the features with a variance <0.8, the kernelized radiomics model from the CT images achieved the best classification performance. The accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of using the kernelized radiomics models both with and without VT to differentiate WD HNSCC from MD/PD HNSCC are shown in **Table 2**. ACC, SEN, and NPV in the cohort with VT selection were significantly higher than those without VT selection. The AUCs of the models with VT and without VT are 0.96 and 0.89, respectively (**Figure 5**). There was significant difference between them ($p < 0.05$).

We used 3-fold validation to split our entire dataset into three parts and recursively used two parts as the training set and one as the testing set. The model was trained on the

TABLE 2 | The performances of kernelized models with and without VT selection.

	ACC	SEN	SPE	PPV	NPV	AUC
With VT selection	0.92	0.96	0.83	0.94	0.91	0.96
Without VT selection	0.83	0.88	0.70	0.80	0.84	0.89
p-value	0.002▲	0.002▲	0.131	0.113	0.000▲	0.000▲

ACC, Accuracy; SEN, Sensitivity; SPE, Specificity. ▲$p < 0.05$.

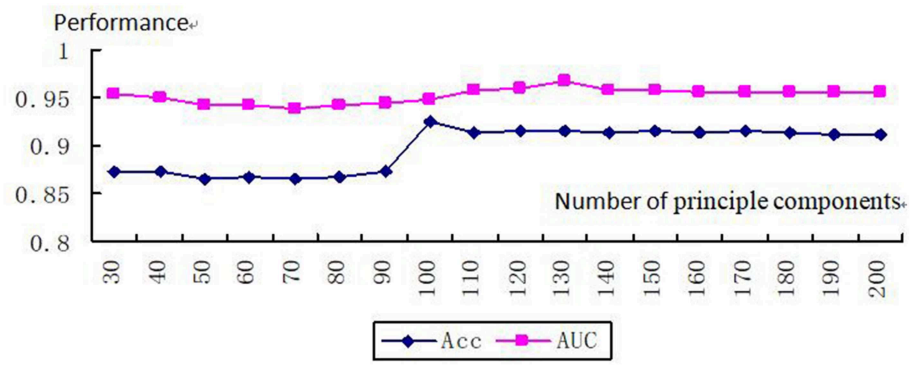


FIGURE 4 | Tuning number of principle components.

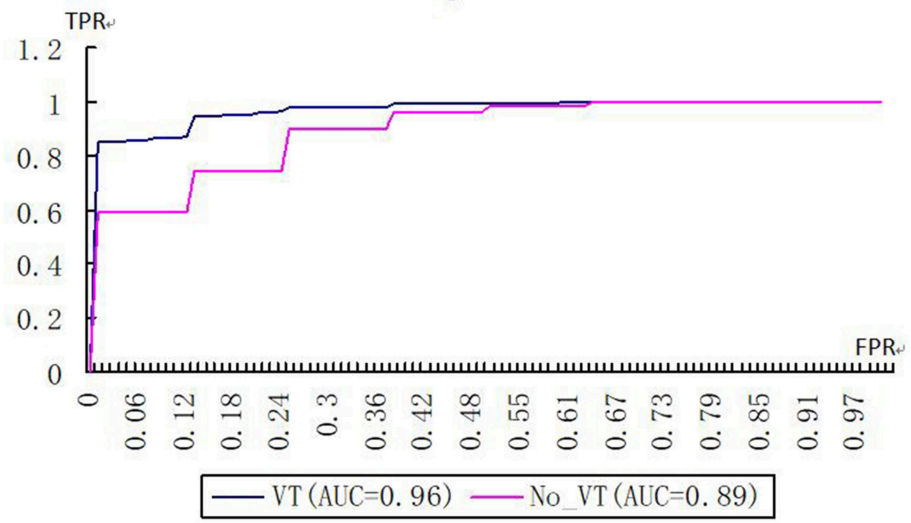


FIGURE 5 | Receiver operating characteristic curves of kernelized models with and without VT selection (FPR false positive rate, TPR true positive rate).

training set and validated on the testing set, which the model could not learn from. We used ACC, SEN, SPE, PPV, NPV, and AUC to describe the performance of the model, which has the ability to ignore an unbalance between samples with different classes.

The performances of each model in discriminating tumor grades are summarized in **Table 3** and **Supplementary Data Sheet 3**. The clinical model had lower performance parameters than the radiomics signature and the combined model, and there were significant differences among these models ($p < 0.05$). The combined model had a relatively higher ACC, SEN, and NPV than the radiomics signature, but there were no significant differences between these two models ($p > 0.05$). The AUCs of the three models are shown in **Figure 6**. The AUC of the clinical model was much lower than that of the radiomics signature and that of the combined model, which were significant differences ($p < 0.05$). The AUC of the radiomics signature was slightly lower than that of the combined model, but the difference was not significant ($p > 0.05$).

TABLE 3 | Discrimination performances of clinical model, radiomics signature features, and the combined model.

Models	ACC	SEN	SPE	PPV	NPV	AUC
Clinical▲♀	0.68	0.87	0.38	0.69	0.68	0.63
Radiomics*▲	0.92	0.96	0.83	0.94	0.91	0.96
Combined*♀	0.93	0.97	0.83	0.90	0.92	0.97
*	0.72	0.52	1.00	0.97	0.54	0.94
p value▲	0.00	0.016	0.00	0.00	0.00	0.00
♀	0.00	0.003	0.00	0.00	0.00	0.00

*$p > 0.05$, ▲$p < 0.05$, ♀$p < 0.05$.

DISCUSSION

In this study, we combined a RBF KPCA with a random forest classifier for the prediction of HNSCC tumor grade, especially for differentiating WD tumors from MD/PD tumors. A total of 670 features were extracted from each tumor lesion. In total, 130 dimensions were from the PCA based on the highest AUCs at different dimension levels (30–200). These 130 dimensions were used as the inputs for the random forest model. Notably, the application of VT selection to eliminate features with variance <0.8 improved the AUC. We also constructed a clinical and a combined model, and evaluated their performances; the combined model achieved the best performance.

As the solid cancer is spatially and temporally heterogeneous, radiomics is advantageous for non-invasively capturing intra-tumoral heterogeneity from medical imaging (10). Radiomics has been reported for grading brain gliomas and can discriminate high- vs. low-grade gliomas (26–28). Although other modalities such as PET, DWI, histogram analysis of apparent diffusion coefficient (ADC) maps, PWI and DCE-MRI have been used to differentiate the histologic grades of HNSCC (6–8), these multi-parameter imaging methods provide information regarding the composition of HNSCC to reflect metabolism, cellularity, and perfusion. There might exist complex associations among those parameters depending on tumor grade (6). Additionally, intra- and inter-observer variability are important factors in whether these radiology diagnostic tools are independently reliable. In this study, the AUC and ACC of our constructed model were higher than those of PWI (8).

Radiomics is a promising tool for the non-invasive characterization of tumor phenotypes. In our study, we extracted a large number of quantitative features from contrast-enhanced CT images: ROIs were characterized regarding their shape and size features, histogram features, texture features

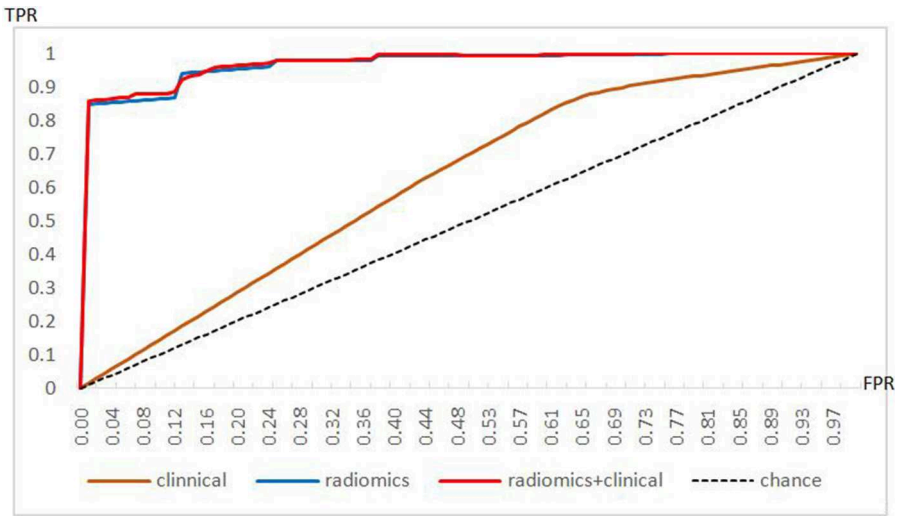


FIGURE 6 | Receiver operating characteristic curves of the performances of three models.

and transformation features. A few radiomics studies have been performed based on MRI to stage and grade HNSCC, and these studies used various methods and obtained some quantitative parameters. Ren et al. (14) also constructed radiomics signatures with the method of least absolute shrinkage and selection operator (LASSO) logistic regression and explored the associations between radiomics signatures and HNSCC stage. The researchers used MRI with contrast-enhanced T1-weighted imaging (CET1WI) and T2-weighted imaging (T2WI) and found that there were three radiomics signatures that were significantly different between stage III-IV and stage I-II in both the testing and training cohorts. Fujima et al. (15) used MRI histograms and a texture analysis of fat-suppressed T2WI to predict the histological grade of HNSCC and found that the relative mean signal and contrast were significantly lower in poorly differentiated SCC than in the well/moderately differentiated SCC. The homogeneity was higher in poorly differentiated SCC than in the well/moderately differentiated SCC. Ahn et al. (16) studied different b values to determine whether histogram analyses of ADC maps can differentiate histologic grades of HNSCC; the researchers found that at a high b value (2,000 s/mm^2), the mean ADC and kurtosis ratio were significantly different among cohorts of different grades, and the diagnostic accuracies varied among various cohorts.

On the basis of VT selection, which eliminated features with a variance <0.8, the kernelized radiomics model from CT images achieved a good performance. The ACC, SEN, and NPV of the kernelized radiomics models with VT were significantly higher than those of the model without VT. The variance threshold could clearly help improve the performance of the model in grading HNSCC.

Of all 670 features extracted from the portal venous phase contrast-enhanced images, the top two features were smoothness and GLCM_t_45_d_1_Con_2. The smoothness feature concerns the texture of the image, which is either smooth or rough. When the image contains constant gray level intensity values, the texture is smooth. When the intensity levels rapidly vary, the texture is considered rough. In this study, the images of the WD cohort were smoother than those of the MD/PD cohort. We speculated that a WD tumor would resemble normal squamous epithelium, be slightly more keratinized, have slight atypia nuclei, and show less necrosis than a MD/PD tumor; these observations reflect the pathological characteristics of the WD tumor and may relate to smoothness. Regarding the feature GLCM_t_45_d_1_Con_2, GLCM describes the spatial relationship of the pixels and characterizes the image texture by calculating how often pairs of pixels with specific values and spatial relationships occur in an image. HNSCC tumors of different grades have various pathological characteristics, including cellularity, necrosis, vessels, desmoplasia, and inflammatory infiltration, all of which have various pixel values and spatial relationships. Fujima et al. (15) also reported that the contrast and homogeneity parameters of the GLCM texture features based on MRI were significantly different between WD/MD and PD SCC patients. GLCM features may be useful for determining HNSCC grade. Surov et al. (29) reported that ADC histogram parameters represent the proliferation potential and cellularity of HNSCC. In G1/2 and G3 tumors, various ADC parameters correlated with Ki67 expression, cellularity, cell count, and total nucleic area, all of which depend on the tumor grade.

To assess the performance of the radiomics signature for discriminating among HNSCC grades, we additionally constructed two models, a clinical model and a combined model, and compared the performances of these models. Among these three models, the combined model achieved the best performance, although there were no significant differences between the radiomics signature and the combined model. When we incorporated clinical and radiological information into the radiomics signature, the performance of the combined

model was not significantly different from that of the radiomics signature, which explains why the radiomics signature also played a predominant role in discriminating between HNSCC grades. A computerized algorithm analysis can make quantitative and qualitative improvements in grading HNSCC tumors with CT images. In prospective radiomics, a signature analysis may serve as a useful, non-invasive tool that is extensively applied in clinical practice.

There were several limitations in our study. First, this was a retrospective and single center study. The study data are limited; multi-center datasets, larger sample data and prospective studies will be needed to validate the performance of our model. Second, the ROIs were subjectively identified by observers according to the most significantly enhanced area inside the tumor on one slice of a CT image. Only 2-dimensional (2D) analysis, rather than 3-dimensional (3D) analysis, was conducted for the radiomics analysis. 3D analyses tend to be more representative of tumor tissue heterogeneity, but a 3D analysis may be more complex and time-consuming. In the future, we will use the automatic segmentation method to define the ROIs. Finally, the methodology used in this study needs to be improved. As machine learning techniques develop, deep learning method has emerged. Convolutional neural network (CNN) is a representative, more advanced method in deep learning. In the future, if the study sample size is enough for deep learning, we will try CNN method for image segmentation and feature extraction. Then the model can be worthy of explaining more.

CONCLUSIONS

In conclusion, in this study, we constructed a radiomics model that could non-invasively discriminate WD HNSCC from MD/PD HNSCC. This radiomics model could be used in precision medicine and improve therapeutic strategies in the clinic. The radiomics model had a better performance with the use of a KPCA, random forest classifier and VT selection and may serve as a potential method for assessing imaging biomarkers for HNSCC patients.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

WW designed and wrote the manuscript. JY designed the imaging analysis method. QW collected and analyzed the data. JL analyzed the data and performed the statistics analysis. SX designed and reviewed the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00821/full#supplementary-material

REFERENCES

1. Janot F, Klijanienko J, Russo A, Mamet JP, de Braud F, El-Naggar AK, et al. Prognostic value of clinicopathological parameters in head and neck squamous cell carcinoma: a prospective analysis. *Br J Cancer*. (1996) 73:531–8. doi: 10.1038/bjc.1996.92
2. Qian X, Sinikovic B, Schreiber F, Ochsenreither S, Klinghammer K, Wollenberg B, et al. pN status predicts outcomes in surgically treated pT1-pT2 patients of various disease stages with squamous cell carcinoma of the head and neck: a 17-year retrospective single center cohort study. *Eur Arch Otorhinolaryngol*. (2018) 275:2787–95. doi: 10.1007/s00405-018-5108-z
3. Subramaniam N, Murthy S, Balasubramanian D, Low TH, Vidhyadharan S, Clark JR, et al. Adverse pathologic features in T1/2 oral squamous cell carcinoma classified by the American Joint Committee on Cancer eighth edition and implications for treatment. *Head Neck*. (2018) 40:2123–8. doi: 10.1002/hed.25168
4. Xu QS, Wang C, Li B, Li JZ, Mao MH, Qin LZ, et al. Prognostic value of pathologic grade for patients with oral squamous cell carcinoma. *Oral Dis*. (2018) 24:335–46. doi: 10.1111/odi.12727
5. Dik EA, Ipenburg NA, Kessler PA, van Es RJJ, Willems SM. The value of histological grading of biopsy and resection specimens in early stage oral squamous cell carcinomas. *J Craniomaxillofac Surg*. (2018) 46:1001–6. doi: 10.1016/j.jcms.2018.03.019
6. Leifels L, Purz S, Stumpp P, Schob S, Meyer HJ, Kahn T, et al. Associations between ^{18}F-FDG-PET, DWI, and DCE parameters in patients with head and neck squamous cell carcinoma depend on tumor grading. *Contrast Media Mol Imaging*. (2017) 2017:5369625. doi: 10.1155/2017/5369625
7. Dong Ji X, Yan S, Xia S, Guo Y, Shen W. Quantitative parameters correlated well with differentiation of squamous cell carcinoma at head and neck: a study of dynamic contrast-enhanced MRI. *Acta Radiol*. (2019) 60:962–8. doi: 10.1177/0284185118809543
8. Abdel Razek AAK, Nada N. Arterial spin labeling perfusion-weighted MR imaging: correlation of tumor blood flow with pathological degree of tumor differentiation, clinical stage and nodal metastasis of head and neck squamous cell carcinoma. *Eur Arch Otorhinolaryngol*. (2018) 275:1301–7. doi: 10.1007/s00405-018-4950-3
9. Bourgier C, Colinge J, Aillères N, Fenoglietto P, Brengues M, Pèlegrin A, et al. Radiomics: definition and clinical development. *Cancer Radiother*. (2015) 19:532–7. doi: 10.1016/j.canrad.2015.06.008
10. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
11. Fujita A, Buch K, Li B, Kawashima Y, Qureshi MM, Sakai O. Difference between HPV-Positive and HPV-Negative non-oropharyngeal head and neck cancer: texture analysis features on CT. *J Comput Assist Tomogr*. (2016) 40:43–7. doi: 10.1097/RCT.0000000000000320
12. Zhang H, Graham CM, Elci O, Griswold ME, Zhang X, Khan MA, et al. Locally advanced squamous cell carcinoma of the head and neck: CT texture and histogram analysis allow independent prediction of overall survival in patients treated with induction chemotherapy. *Radiology*. (2013) 269:801–9. doi: 10.1148/radiol.13130110
13. Ou D, Blanchard P, Rosellini S, Levy A, Nguyen F, Leijenaar RTH, et al. Predictive and prognostic value of CT based radiomics signature in locally advanced head and neck cancers patients treated with concurrent chemoradiotherapy or bioradiotherapy and its added value to Human Papillomavirus status. *Oral Oncol*. (2017) 71:150–5. doi: 10.1016/j.oraloncology.2017.06.015
14. Ren J, Tian J, Yuan Y, Dong D, Li X, Shi Y, et al. Magnetic resonance imaging based radiomics signature for the preoperative discrimination of stage I-II and III-IV head and neck squamous cell carcinoma. *Eur J Radiol*. (2018) 106:1–6. doi: 10.1016/j.ejrad.2018.07.002

15. Fujima N, Homma A, Harada T, Shimizu Y, Tha KK, Kano S, et al. The utility of MRI histogram and texture analysis for the prediction of histological diagnosis in head and neck malignancies. *Cancer Imaging*. (2019) 19:5. doi: 10.1186/s40644-019-0193-9
16. Ahn SJ, Choi SH, Kim YJ, Kim KG, Sohn CH, Han MH, et al. Histogram analysis of apparent diffusion coefficient map of standard and high B-value diffusion MR imaging in head and neck squamous cell carcinoma: a correlation study with histological grade. *Acad Radiol*. (2012) 19:1233–40. doi: 10.1016/j.acra.2012.04.019
17. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJ. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *Front Oncol*. (2015) 5:272. doi: 10.3389/fonc.2015.00272
18. Jethanandani A, Lin TA, Volpe S, Elhalawani H, Mohamed ASR, Yang P, et al. Exploring applications of radiomics in magnetic resonance imaging of head and neck cancer: a systematic review. *Front Oncol*. (2018) 8:131. doi: 10.3389/fonc.2018.00131
19. Van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, et al. Scikit-image: image processing in Python. *PeerJ*. (2014) 2:e453. doi: 10.7717/peerj.453
20. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. (2011) 12:2825–30.
21. Kolokythas A, Park S, Schlieve T, Pytynia K, Cox D. Squamous cell carcinoma of the oral tongue: histopathological parameters associated with outcome. *Int J Oral Maxillofac Surg*. (2015) 44:1069–74. doi: 10.1016/j.ijom.2015.01.027
22. Thomas B, Stedman M, Davies L. Grade as a prognostic factor in oral squamous cell carcinoma: a population-based analysis of the data. *Laryngoscope*. (2014) 124:688–94. doi: 10.1002/lary.24357
23. González-García R, Naval-Gías L, Rodríguez-Campo FJ, Sastre-Pérez J, Muñoz-Guerra MF, Gil-Díez Usandizaga JL. Contralateral lymph neck node metastasis of squamous cell carcinoma of the oral cavity: a retrospective analytic study in 315 patients. *J Oral Maxillofac Surg*. (2008) 66:1390–8. doi: 10.1016/j.joms.2008.01.012
24. Kendi AT, Corey A, Magliocca KR, Nickleach DC, Galt J, Switchenko JM, et al. 18F-FDG-PET/CT parameters as imaging biomarkers in oral cavity squamous cell carcinoma, is visual analysis of PET and contrast enhanced CT better than the numbers? *Eur J Radiol*. (2015) 84:1171–6. doi: 10.1016/j.ejrad.2015.02.030
25. Koyasu S, Nakamoto Y, Kikuchi M, Suzuki K, Hayashida K, Itoh K, et al. Prognostic value of pretreatment 18F-FDG PET/CT parameters including visual evaluation in patients with head and neck squamous cell carcinoma. *Am J Roentgenol*. (2014) 202:851–8. doi: 10.2214/AJR.13.11013
26. Zacharaki EI, Wang S, Chawla S, Soo Yoo D, Wolf R, Melhem ER, et al. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn Reson Med*. (2009) 62:1609–18. doi: 10.1002/mrm.22147
27. Ditmer A, Zhang B, Shujaat T, Pavlina A, Luibrand N, Gaskill-Shipley M, et al. Diagnostic accuracy of MRI texture analysis for grading gliomas. *J Neurooncol*. (2018) 140:583–9. doi: 10.1007/s11060-018-2984-4
28. Cho HH, Lee SH, Kim J, Park H. Classification of the glioma grading using radiomics analysis. *PeerJ*. (2018) 6:e5982. doi: 10.7717/peerj.5982
29. Surov A, Meyer HJ, Winter K, Richter C, Hoehn AK. Histogram analysis parameters of apparent diffusion coefficient reflect tumor cellularity and proliferation activity in head and neck squamous cell carcinoma. *Oncotarget*. (2018) 9:23599–607. doi: 10.18632/oncotarget.25284

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 03 September 2019
doi: 10.3389/fonc.2019.00844

Three-Dimensional Texture Analysis Based on PET/CT Images to Distinguish Hepatocellular Carcinoma and Hepatic Lymphoma

Hanyue Xu[1,2], *Wen Guo*[2], *Xiwei Cui*[2], *Hongyu Zhuo*[3], *Yinan Xiao*[2], *Xuejin Ou*[2], *Yunuo Zhao*[2], *Tao Zhang*[2] *and Xuelei Ma*[1,3]*

[1] State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University and Collaborative Innovation Center for Biotherapy, Chengdu, China, [2] West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China, [3] West China Hospital, Sichuan University, Chengdu, China

Objectives: This study compared the diagnostic ability of image-based parameters with texture parameters in the differentiation of hepatocellular carcinoma (HCC) and hepatic lymphoma (HL) by positron emission tomography–computed tomography (PET/CT).

Methods: Patients with pathological diagnosis of HCC and HL were included in this study. Image-based and texture parameters were obtained by manual drawing of region of interest. Receiver operating characteristic (ROC) was used to test the diagnostic capacity of each parameter. Binary logistic regression was used to transform the most discriminative image-based parameters, texture parameters, and the combination of these parameters into three regression models. ROC was used to test the diagnostic capacity of these models.

Result: Ninety-nine patients diagnosed with HCC ($n = 76$) and HL ($n = 23$, 10 primary HL, 13 secondary HL) by histological examination were included in this study (From 2011 to 2018, West China hospital). According to the AUC and p-value, 2 image-based parameters and five texture parameters were selected. The diagnostic ability of texture-based model was better than that of image-based model, and after combination of those two groups of parameters the diagnostic capacity improved.

Conclusion: Texture parameters can differentiate HCC from HL quantitatively and improve the diagnostic ability of image-based parameters.

Keywords: hepatocellular carcinoma, hepatic lymphoma, positron emission tomography–computed tomography, texture, differentiation

OPEN ACCESS

Edited by:
Georgios S. Limouris, School of Health Sciences, National and Kapodistrian University of Athens, Greece

Reviewed by:
Yong-An Chung, Incheon St. Mary's Hospital, South Korea
Athanasios Zafeirakis, Army Share Fund Hospital (NIMTS), Greece

***Correspondence:**
Xuelei Ma
drmaxuelei@gmail.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *02 June 2019*
Accepted: *19 August 2019*
Published: *03 September 2019*

Citation:
Xu H, Guo W, Cui X, Zhuo H, Xiao Y, Ou X, Zhao Y, Zhang T and Ma X (2019) Three-Dimensional Texture Analysis Based on PET/CT Images to Distinguish Hepatocellular Carcinoma and Hepatic Lymphoma. Front. Oncol. 9:844. doi: 10.3389/fonc.2019.00844

INTRODUCTION

Malignant hepatic nodules include primary malignant hepatic neoplasms, such as hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC), and metastatic diseases from bile duct, lymphoid cells, endothelial cells. Hepatocellular carcinoma (HCC) accounts for almost 80% of all primary malignant hepatic neoplasms (1). Primary hepatic lymphoma (PHL) is rare, while secondary hepatic lymphoma (SHL), widespread lymphoma with liver involved, appears in 50% patients with non-Hodgkin lymphoma and 20% patients with Hodgkin lymphoma (2).

Differentiation between HCC and hepatic lymphoma, both in its primary and secondary form, is difficult. The images of HL is variable and can be similar to that of HCC, and several studies have reported cases about patients with PHL mimicking HCC (3–5). SHL is characterized by hepatosplenomegaly and systemic involvement which contributes to clinic diagnosis, but radiological diagnosis of SHL still requires further improvement since it has no specific features in sonography, and can appear as a solitary lesion (2, 6). Although the Biopsy has been extensively used as a clinic tool to distinguish HCC from HL, cancer cells extracted only represent part of the lesion and therefore may cause selection bias. Positron emission tomography-computed tomography (PET/CT), a non-invasive clinic examination, reveals the anatomical structure as well as the glucose metabolism of tissues and therefore has been widely used in the diagnosis and prognosis of hepatic lesions (7, 8). However, the accuracy of PET/CT in the diagnosis of different hepatic lesions remains to be improved (9).

Texture analysis is a newly-developed high throughput way to extract digital information from images that naked eyes cannot detect, and can thus explore more characteristics of images (10). Some studies have adopted texture analysis combined with PET/CT to differentiate benign from malignant mediastinal lymph nodes and distinguish malignant from benign bone and soft-tissue lesions (11, 12). Aiming at distinguishing HCC from HL in a non-invasive way, we explored the ability of texture and image-based parameters of PET/CT in differentiating HCC and HL.

METHODS

Patients

This study was approved by the West China Hospital Ethics Committee and had a waiver of patients written informed consent. From Jan. 2011 to Dec. 2018, all patients diagnosed with liver lesions by PET/CT were recruited, and only patients with pathological diagnosis of HCC and HL were included. Their information was obtained from the clinical and radiological databases of our hospital. The inclusion criteria were having: (1) images obtained before treatment for hepatic lesions, (2) images obtained from a same system, (3) complete clinical and radiological information, (4) pathological diagnosis of HCC or HL; exclusion criteria were having: (1) incomplete image or clinical information, (2) FDG uptake of liver lesions below or comparable to background activity, (3) liver transplantation.

PET/CT Examination

Patients were fasted 4–6 h and had serum glucose concentration <200 mg/dl before the intravenous injection of 185–370 MBq of 18F-FDG (4 MBq/kg of body weight). After injection, patients rested in a quiet room for 1 h. Then, a whole-body PET/CT scanner (Gemini GXL; Philips Medical Systems, The Netherlands) was used for imaging. During imaging process, patients were in supine position with both arms extended in the cranial direction and breathing quietly. PET images were obtained at 2 min/bed.

The CT images were acquired simultaneously with parameters as follows: 40 mAs, 120 kVp, a slice thickness of 2 mm, and a pitch of 4 mm. After acquisition completed, the transverse, sagittal, and coronal plane images of CT and PET were reconstructed automatically by the computer. The PET images were reconstructed by the line of response (LOR) method after a CT attenuation correction.

Radiomics Extraction

All scans were analyzed by two senior residents independently (HYX, 3 year of training; WG, 4 year of training) and were supervised by a senior radiologist (XLM, 13 years of experience) in order to handle the non-consensus. All of them were blinded to the histological outcomes. Each region of interest (ROI) was manually drawn along the liver lesion, slice by slice on axial images, by using a dedicated software for image analysis (LIFEx software, version 3.74, French Alternative Energies and Atomic Energy Commission). **Figure 1** shows PET/CT images of two case examples, HCC and HL, respectively. Intra-luminal water, cavity, and necrotic components that can be distinguished from tumor solid portion by naked eye were excluded via a fixed 40% threshold of SUV max.

Radiomics Features Analysis

A total of 45 radiomics parameters were extracted from images and divided into seven categories, including conventional PET/CT parameters (SUV and TLG), Histogram (HISTO), Shape value, Gray Level Co-occurrence Matrix (GLCM), Gray-Level Run Length Matrix (GLRLM), Neighborhood Gray-Level Different Matrix (NGLDM), Gray-Level Zone Length Matrix (GLZLM). The conventional PET/CT parameters included the minimum, average, maximum Standardized Uptake Value (SUV), and the Total Lesion Glycolysis (CONV_TLG) in the Volume of Interest. HISTO reflects the gray level of ROIs, regarding the "number of gray level" as "the size of bin". Shape value is the sum total of volume of ROIs in mL and in voxels. GLCM describes the gray-level value distribution of voxel pairs along from 13 different directions at different distances in the ROI. GLRLM corresponds to the amount of homogeneous in 13 directions of the ROIs. NGLDM reflects the difference of gray-level between one voxel and its 26 neighbors in three dimensions. GLZLM describes the size of homogeneous zones for each gray-level in three dimensions.

The Mann-Whitney U test (U test) and $\chi 2$ test were used for comparing the baseline characteristics of those two groups. Since not all the parameters contributed to differentiating HCC and HL, we resorted to the results of operation characteristic curve (ROC curve) to select the most discriminative parameters in each category (**Table 2**). The binary logistic regression was used to transform the group of related parameters into a set of corresponding variables by three models. The discrimination ability of these models based on image-based parameters, texture parameters, and the combination of the two parameters were measured by the Area Under Curve (AUC) of ROC curves. *P*-values <0.05 were considered to be statistically significant. All statistical analyses were performed by SPSS (version 25, IBM, USA).

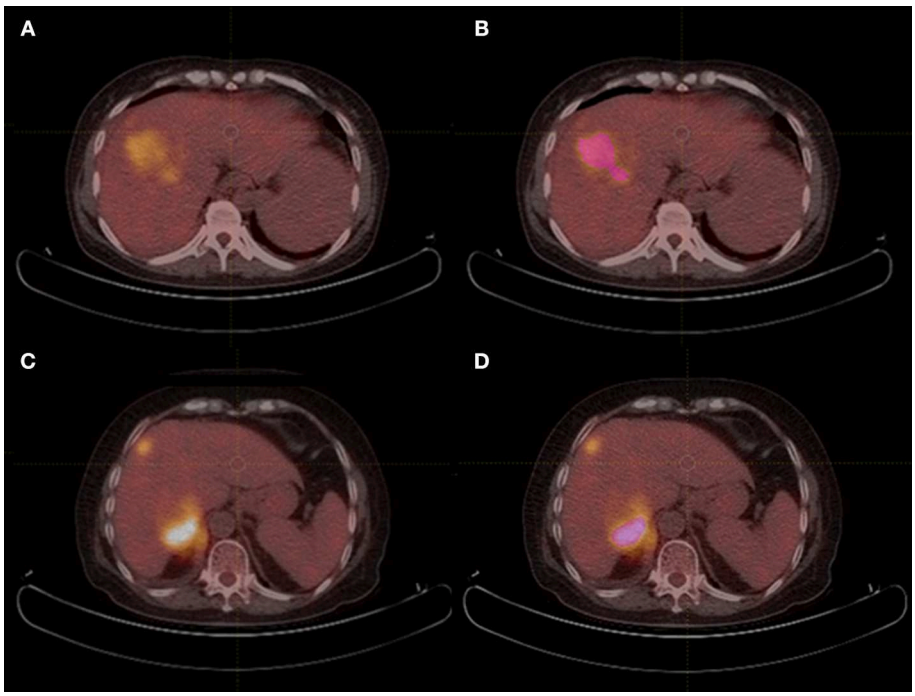


FIGURE 1 | PET/CT images of HCC and HL case examples. **(A,B)** An example of hepatocellular carcinoma mimicking hepatic lymphoma, and region of interest was drawn in **(B)**. (CONVENTIONAL_SUVmin = 2.1, CONVENTIONAL_TLG = 293.5, SHAPE_Compacity = 2.41, GLCM_Correlation = 0.637, GLRLM_GLNU = 105.7, NGLDM_Contrast=0.055, and GLZLM_GLNU=14.4); **(C,D)** An example of secondary hepatic lymphoma mimicking hepatocellular carcinoma, and region of interest was drawn in **(D)**. (CONVENTIONAL_SUVmin = 6.6, CONVENTIONAL_TLG = 425.9, SHAPE_Compacity = 1.98, GLCM_Correlation = 0.567, GLRLM_GLNU = 43.4, NGLDM_Contrast=0.238, and GLZLM_GLNU = 18.9).

RESULT

Patients

The characteristics of patients and lesions were summarized in **Table 1**. There were 74 men (mean age, 54 years ± 14.5; age range 22–86 years) and 25 women (mean age, 52 years ± 16.3; age range 19–76 years). Based on histopathological proof, 23 patients had HL (10 PHL, 13 SHL) and 76 patients had HCC were prospectively included. The median ages of patients with HL was 51 and that of HCC was 54. Based on the TNM classification of malignant tumors, patients with HCC were divided into four groups, one patient of I stage, 16 of II stage, 14 of III stage, and 45 of IV stage. Among patients with HL, there were 12 Diffuse large B cell lymphoma (DLBCL), four B cell lymphoma (except DLBCL), six Hodgkin lymphoma, and one NK/T-cell lymphoma.

Imaging Features

A total of six image-based parameters and 39 texture parameters were extracted and compared. The top two image-based parameters (CONVENTIONAL_SUVmin, CONVENTIONAL_TLG, AUC: 0.642, 0.686, $p < 0.05$) and the top five discriminative texture parameters (SHAPE_Compacity, GLCM_Correlation, GLRLM_GLNU, NGLDM_Contrast, and GLZLM_GLNU, AUC: 0.784, 0.726, 0.774, 0.721, 0.704, $p < 0.05$) were selected by ROC analysis (**Supplementary Table 1**). Binary logistic regression was used to transform groups of parameters into correspondent predictive models, including models transformed from image-based parameters, texture features, and the combination of those two kinds of parameters. Three predictive models were shown in **Table 3**: MODimage, MODtexture, and MODcombination. **Table 4** showed the ROC results of these three models. AUC of the model transformed from image-based parameters was 0.822, with sensitivity of 69.6%, specificity of 73.7%. AUC of the model related with texture parameters was 0.870, with increased sensitivity of 91.3% and specificity of 77.6%. AUC of model transformed from the combination of image-based parameters and texture parameters was 0.898, with the same sensitivity and specificity as that of texture-based model (**Figure 2**).

DISCUSSION

In this study, we used image-based parameters and texture parameters from 18F-FDG PET/CT to differentiate HCC from HL, and found many parameters significantly different between those two diseases. By comparing AUC of diagnostic models of image-based parameters, texture parameters, and the combination of the two parameters, we found that texture parameters presented better diagnostic ability than image-based parameters and that combination of the two parameters possessed a more effective diagnostic capacity than the other two groups.

The differentiation between HCC and HL is necessary, as their managements are different. Early stage HCC can be treated by excision of lesions, while the main treatment choice for secondary HL is multiagent chemotherapy (2). As an invasive method, the liver biopsy is prone to selection bias and may cause metastasis though it can offer reliable proofs for cancer diagnosis (13). The traditional radiological diagnosis could provide images with summarized features of the lesion. For instance, on contrast-enhanced images, HCC is characterized by "wash-out" pattern and fibrous tumor capsule (13). However, the images of HL are less specific and too variable to provide solid evidence for clinic diagnosis (4). Consequently, non-invasive and precise methods are required to differentiate HCC between HL.

PET/CT can reveal the metabolic characters of organs, and as tumor cells have enhanced glycolysis, they have higher 18F-FDG uptake compared with normal tissue. Previous studies have indicated that the SUV metrics of lymphoma was higher than that of carcinoma in PET/CT images when differentiating renal carcinoma and lymphoma with renal involvement (14, 15). However, a study claimed that the SUV max was not

TABLE 1 | Patient characteristics.

	HL (*N* = 23) Median (Range or %)	HCC (*N* = 76) Median (Range or %)	*P*-value
Gender			
Male	10	64	<0.05
Female	13	12	
Age	51 (19–85)	54 (23–86)	0.878
Histopathologic diagnosis of HL			
Diffuse large B cell lymphoma (DLBCL)	12 (52%)	NA	NA
B cell lymphoma (except DLBCL)	4 (17%)		
Hodgkin lymphoma	6 (26%)		
NK/T-cell lymphoma	1 (5%)		
TNM Stage			**SUV mean (sd)**
I	NA	1	2.67
II		16	3.41 (0.96)
III		14	3.63 (1.33)
IV		45	4.82 (1.90)
all		76	4.27 (1.75)
Ann Arbor Stage			**SUV mean (sd)**
II	2	NA	4.79 (0.50)
IV	21		6.17 (5.05)
all	23		6.05 (4.83)

HL, hepatic lymphoma; HCC, hepatocellular carcinoma; NK, natural killer; NA, not applicable; sd, standard deviation.

TABLE 3 | Regression models composed of image-based parameters, texture features, and the combination of those two kinds of parameters.

Model	Formula
MODimage	−2.154 CONVENTIONAL_SUVmin + 2.349 CONVENTIONAL_TLG - 1.065
MODtexture	20.405 SHAPE_Compacity-0.031 GLCM_Correlation+0.888 GLRLM_GLNU-2.498 NGLDM_Contrast-18.289 GLZLM_GLNU-0.758
MODcombination	36.534 SHAPE_Compacity+0.122 GLCM_Correlation+0.926 GLRLM_GLNU-1.783 NGLDM_Contrast-16.767 GLZLM_GLNU-0.975 CONVENTIONAL_SUVmin-17.756 CONVENTIONAL_TLG-0.76

TABLE 4 | Comparison of differential diagnostic ability of the three predictive models.

Test result variable(s)	Sensitivity	Specificity	AUC (95% CI)	Asymptotic Sig.b
Image based	0.696	0.737	0.822(0.740–0.904)	<0.001
Texture	0.913	0.776	0.870(0.788–0.953)	<0.001
Combination	0.913	0.776	0.898(0.838–0.959)	<0.001

TABLE 2 | The results of ROC analysis of optimal image-based and texture parameters in PET and CT images for hepatocellular carcinoma vs. hepatic lymphoma.

	HCC		HL		AUC	*P*-value
	Median	Range	Median	Range		
Image-based parameters						
CONVENTIONAL_SUVmin (SUV)	2.28	1.06–4.64	3.73	0.89–9.87	0.642	0.039
CONVENTIONAL_TLG (mL)	751.67	8.22–4403.85	552.96	11.50–6299.10	0.686	0.007
Texture parameters						
SHAPE_Compacity	2.53	0.77–5.78	1.52	0.00–6.13	0.784	<0.001
GLCM_Correlation	0.63	0.20–0.86	0.52	0.20–0.78	0.726	0.001
GLRLM_GLNU	238.79	8.58–2777.08	104.75	3.22–1622.28	0.774	<0.001
NGLDM_Contrast	0.08	0.01–0.46	0.22	0.03–1.42	0.721	0.001
GLZLM_GLNU	21.53	1.00–148.42	13.83	1.25–121.32	0.704	<0.001

GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run-length matrix; GLNU, Gray-Level Non-Uniformity; NGLDM, Neighborhood Gray-Level Different Matrix; GLZLM, gray-level zone-length matrix.

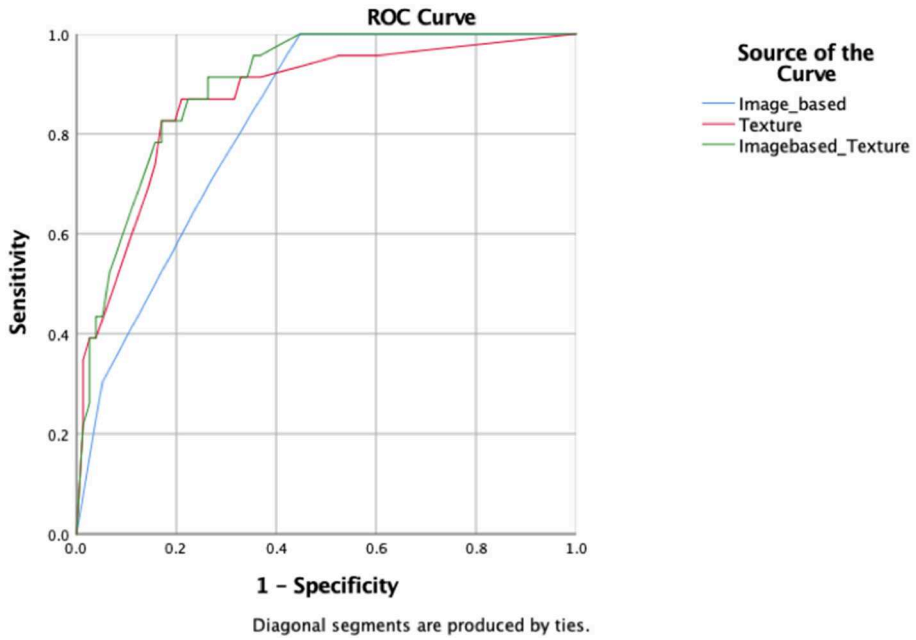


FIGURE 2 | ROC curves of the three radiomic predictive models.

conclusive when distinguishing between primary nasopharyngeal lymphoma and nasopharyngeal carcinoma (16). Though many kinds of lymphoma are FDG avid, the diagnostic capacity of PET/CT remains unconfirmed (8). In our study, the SUV min of HL was higher than that of HCC while the TLG of HL was lower than that of HCC. The SUV min, a measurement of metabolic activity per body weight, could reflect the lowest point of metabolic activity within the tumor, but TLG takes into account the volume of the tumor lesion additionally. Therefore, this result may be affected by the volume of HCC.

Texture analysis can quantify image features by extracting the distribution and relation of pixel or voxel grayscale in images. Some studies have applied texture analysis to classifying benign and malignant liver lesions or stratifying different stages of liver fibrosis (17–19). However, no study has used texture analysis to distinguish HCC from HL, and it may because the morbidity of HL is relatively low, and the similarity of malignant lesions is more than that of benign and malignant liver lesions. In our study, texture parameters are more effective than image-based parameters in differentiation HCC from HL (AUC: 0.822 VS. 0.870). Previous studies compared carcinoma with lymphoma via texture analysis and proposed that the pixel gray-level value has a tight correlation with diagnosis, which is further confirmed by our results (20, 21). Early texture analysis in CT to differentiate malignant and benign liver lesion found that the First Order Statistics (FOS) performed best (22). However, in our study, FOS is less effective, while the secondary features, GLCM_Correlation, GLRLM_GLNU, NGLDM_Contrast, and GLZLM_GLNU, presented more significant differences between these two diseases. The result demonstrated that the gray levels of ROIs of HCC and HL were not distinguishing enough, and thus the second-order features such as gray-level value comparison were necessary.

Based on the better diagnostic capacity of texture parameters, we hypothesized that the combination of image-based and texture parameters contributes more to the clinic diagnosis of cancers. The result of AUC indicated that the combination group improved diagnostic capacity (AUC: 0.898), though the sensitivity of the specificity of the combination model remained the same as that of the texture model. Consistent with our previous studies, the same results were found in the diagnosis of breast carcinoma and breast involved lymphoma (23). Besides, another study found the same result in the differentiation of benign and malignant breast tumors (24). The combination of texture and image-based parameters could quantify and enhance the accuracy of the imaged-based PET/CT diagnosis.

However, the limited number of lymphoma group which did not include all kinds of lymphoma may lead to selected bias and therefore impact the accuracy of our result. Moreover, characterized as labor-intensive, ROIs are subject to manual measurement errors when compared with rectangular ROIs used drawn by computers. Finally, the relationship between texture parameters and histopathological structures requires further study.

In conclusion, our study confirmed the role of texture analysis in diagnosing different pathological cancer types and therefore proposed a new method for differentiating HCC and HL. Although both image-based and texture parameters can distinguish HCC from HL, the latter one is more efficient and the combination of the two parameters contribute to the diagnosis of HCC and HL more effectively.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

XM and HX contributed to the conception and design this study. WG and XC developed the methodology. XO and HZ analyzed and interpreted the data. YZ, YX, and TZ wrote the manuscript and approved the final submission of the study. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00844/full#supplementary-material

REFERENCES

1. Fernandez MP, Redvanly RD. Primary hepatic malignant neoplasms. *Radiol Clin North Am.* (1998) 36:333–48. doi: 10.1016/S0033-8389(05)70026-9
2. Rajesh S, Bansal K, Sureka B, Patidar Y, Bihari C, Arora A. The imaging conundrum of hepatic lymphoma revisited. *Insights Imaging.* (2015) 6:679–92. doi: 10.1007/s13244-015-0437-6
3. Bohlok, De Grez T, Bouazza F, De Wind R, El-Khoury M, Repullo D, et al. Primary hepatic lymphoma mimicking a hepatocellular carcinoma in a cirrhotic patient: case report and systematic review of the literature. *Case Rep Surg.* (2018) 2018:9183717. doi: 10.1155/2018/9183717
4. Lee J, Park KS, Kang MH, Kim Y, Son SM, Choi H, et al. Primary hepatic peripheral T-cell lymphoma mimicking hepatocellular carcinoma: a case report. *Ann Surg Treat Res.* (2017) 93:110–114. doi: 10.4174/astr.2017.93.2.110
5. Kaneko K, Nishie A, Arima F, Yoshida T, Ono K, Omagari J, et al. A case of diffuse-type primary hepatic lymphoma mimicking diffuse hepatocellular carcinoma. *Ann Nucl Med.* (2011) 25:303–7. doi: 10.1007/s12149-010-0460-0
6. Abe H, Kamimura K, Kawai H, Kamimura H, Domori K, Kobayashi Y, et al. Diagnostic imaging of hepatic lymphoma. *Clin Res Hepatol Gastroenterol.* (2015) 39:435–42. doi: 10.1016/j.clinre.2014.11.001
7. Parsai A, Miquel ME, Jan H, Kastler A, Szyszko T, Zerizer I. Improving liver lesion characterisation using retrospective fusion of FDG PET/CT and MRI. *Clin Imaging.* (2019) 55:23–28. doi: 10.1016/j.clinimag.2019.01.018
8. Haug AR. Imaging of primary liver tumors with positron-emission tomography. *Q J Nucl Med Mol Imaging.* (2017) 61:292–300. doi: 10.23736/S1824-4785.17.02994-6
9. Tan GJ, Berlangieri SU, Lee ST, Scott AM. FDG PET/CT in the liver: lesions mimicking malignancies. *Abdom Imaging.* (2014) 39:187–95. doi: 10.1007/s00261-013-0043-3
10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
11. Bayanati H, Thornhill RE, Souza CA, Sethi-Virmani V, Gupta A, Maziak D, et al. Quantitative CT texture and shape analysis: can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer? *Eur Radiol.* (2015) 25:480–7. doi: 10.1007/s00330-014-3420-6
12. Xu R, Kido S, Suga K, Hirano Y, Tachibana R, Muramatsu K, et al. Texture analysis on (18)F-FDG PET/CT images to differentiate malignant and benign bone and soft-tissue lesions. *Ann Nucl Med.* (2014) 28:926–35. doi: 10.1007/s12149-014-0895-9
13. Ronot M, Vilgrain V. Hepatocellular carcinoma: diagnostic criteria by imaging techniques. *Best Pract Res Clin Gastroenterol.* (2014) 28:795–812. doi: 10.1016/j.bpg.2014.08.005
14. Nicolau C, Sala E, Kumar A, Goldman DA, Schoder H, Hricak H, et al. Renal masses detected on FDG PET/CT in patients with lymphoma: imaging features differentiating primary renal cell carcinomas from renal lymphomatous involvement. *Am J Roentgenol.* (2017) 208:849–53. doi: 10.2214/AJR.16.17133
15. Ye XH, Chen LH, Wu HB, Feng J, Zhou WL, Yang RM, et al. 18F-FDG PET/CT evaluation of lymphoma with renal involvement: comparison with renal carcinoma. *South Med J.* (2010) 103:642–9. doi: 10.1097/SMJ.0b013e3181e23cb0
16. Cho KS, Kang DW, Kim HJ, Lee JK, Roh HJ. Differential diagnosis of primary nasopharyngeal lymphoma and nasopharyngeal carcinoma focusing on CT, MRI, and PET/CT. *Otolaryngol Head Neck Surg.* (2012) 146:574–8. doi: 10.1177/0194599811434712
17. Wu CM, Chen YC, Hsieh KS. Texture features for classification of ultrasonic liver images. *IEEE Trans Med Imaging.* (1992) 11:141–52. doi: 10.1109/42.141636
18. Stocker D, Marquez HP, Wagner MW, Raptis DA, Clavien PA, Boss A, et al. MRI texture analysis for differentiation of malignant and benign hepatocellular tumors in the non-cirrhotic liver. *Heliyon.* (2018) 4:e00987. doi: 10.1016/j.heliyon.2018.e00987
19. Vicas C, Lupsor M, Badea R, Nedevschi S. Usefulness of textural analysis as a tool for noninvasive liver fibrosis staging. *J Med Ultrasonics.* (2011) 38:105–17. doi: 10.1007/s10396-011-0307-x
20. Huang Z, Li M, He D, Wei Y, Yu H, Wang Y, et al. Two-dimensional texture analysis based on CT images to differentiate pancreatic lymphoma and pancreatic adenocarcinoma: a preliminary study. *Acad Radiol.* (2018) 26:e189–95. doi: 10.1016/j.acra.2018.07.021
21. Kunimatsu A, Kunimatsu N, Kamiya K, Watadani T, Mori H, Abe O. Comparison between glioblastoma and primary central nervous system lymphoma using MR image-based texture analysis. *Magn Reson Med Sci.* (2018) 17:50–7. doi: 10.2463/mrms.mp.2017-0044
22. Mougiakakou SG, Valavanis IK, Nikita A, Nikita KS. Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers. *Artif Intell Med.* (2007) 41:25–37. doi: 10.1016/j.artmed.2007.05.002
23. Ou X, Wang J, Zhou R, Zhu S, Pang F, Zhou Y, et al. Ability of (18)F-FDG PET/CT radiomic features to distinguish breast carcinoma from breast lymphoma. *Contrast Media Mol Imaging.* (2019) 2019:4507694. doi: 10.1155/2019/4507694
24. Li Z, Yu L, Wang X, Yu H, Gao Y, Ren Y, et al. Diagnostic performance of mammographic texture analysis in the differential diagnosis of benign and malignant breast tumors. *Clin Breast Cancer.* (2018) 18:621–7. doi: 10.1016/j.clbc.2017.11.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 06 September 2019
doi: 10.3389/fonc.2019.00876

Glioblastoma and Anaplastic Astrocytoma: Differentiation Using MRI Texture Analysis

Zerong Tian[1†], *Chaoyue Chen*[1†], *Yimeng Fan*[2,3], *Xuejin Ou*[4], *Jian Wang*[5], *Xuelei Ma*[6,7*] *and Jianguo Xu*[1*]

[1] *Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu, China,* [2] *Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, China,* [3] *State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, Chengdu, China,* [4] *West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China,* [5] *School of Computer Science, Nanjing University of Science and Technology, Nanjing, China,* [6] *Department of Biotherapy, Cancer Center, West China Hospital, Sichuan University, Chengdu, China,* [7] *State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University Collaborative Innovation Center for Biotherapy, Chengdu, China*

OPEN ACCESS

Edited by:
Sebastian Cerdan, Spanish National Research Council (CSIC), Spain

Reviewed by:
Ana Paula Candiota, Centre for Biomedical Network Research (CIBER), Spain
Xudong Shen, Guizhou Medical University (GMU), China

***Correspondence:**
Xuelei Ma
drmaxuelei@gmail.com
Jianguo Xu
drjianguoxu@gmail.com

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *19 June 2019*
Accepted: *23 August 2019*
Published: *06 September 2019*

Citation:
Tian Z, Chen C, Fan Y, Ou X, Wang J, Ma X and Xu J (2019) Glioblastoma and Anaplastic Astrocytoma: Differentiation Using MRI Texture Analysis. Front. Oncol. 9:876. doi: 10.3389/fonc.2019.00876

Introduction: Glioblastoma and anaplastic astrocytoma (ANA) are two of the most common primary brain tumors in adults. The differential diagnosis is important for treatment recommendations and prognosis assessment. This study aimed to assess the discriminative ability of texture analysis using machine learning to distinguish glioblastoma from ANA.

Methods: A total of 123 patients with glioblastoma ($n = 76$) or ANA ($n = 47$) were enrolled in this study. Texture features were extracted from contrast-enhanced Magnetic Resonance (MR) images using LifeX package. Three independent feature-selection methods were performed to select the most discriminating parameters:Distance Correlation, least absolute shrinkage and selection operator (LASSO), and gradient correlation decision tree (GBDT). These selected features (datasets) were then randomly split into the training and the validation group at the ratio of 4:1 and were fed into linear discriminant analysis (LDA), respectively, and independently. The standard sensitivity, specificity, the areas under receiver operating characteristic curve (AUC) and accuracy were calculated for both training and validation group.

Results: All three models (Distance Correlation + LDA, LASSO + LDA and GBDT + LDA) showed promising ability to discriminate glioblastoma from ANA, with AUCs $\geq$0.95 for both the training and the validation group using LDA algorithm and no overfitting was observed. LASSO + LDA showed the best discriminative ability in horizontal comparison among three models.

Conclusion: Our study shows that MRI texture analysis using LDA algorithm had promising ability to discriminate glioblastoma from ANA. Multi-center studies with greater number of patients are warranted in future studies to confirm the preliminary result.

Keywords: texture features, machine learning, linear discriminant analysis, differential diagnosis, glioblastoma, anaplastic astrocytoma

INTRODUCTION

Glioblastoma and anaplastic astrocytoma (ANA) are two of the most common primary brain tumors in adults (1). There is a true increase in incidence rates, especially in the elderly (1–3). In clinical practice, it is difficult to differentiate patients with glioblastoma from those with ANA before surgery or biopsy, because the symptoms and signs of the two tumors are relatively uniform and non-specific (4, 5). However, the management for them are different, such as the chemotherapy protocol, dosage, and mode of administration (6). For example, for patients with ANA (WHO grade III), it is recommended to receive radiotherapy or TMZ after resection or biopsy; while for patients newly diagnosed with glioblastoma (WHO grade IV), it is radiotherapy plus concurrent TMZ, followed by adjuvant TMZ. According to previous studies, glioblastoma and ANA grow by invasion into normal brain tissue, spread through the cerebrospinal fluid (CSF), and extend beyond a single carotid or vertebral artery distribution, thus they both have a poor response to medical management and become leading causes of cancer-related death in adults (7, 8). Besides, the prognosis of glioblastoma and ANA are different. In the elderly population, there is no significant difference in prognosis between glioblastoma and ANA, but the difference may exist in younger population (9). Therefore, it is hard but crucial to distinguish glioblastoma from ANA.

Magnetic Resonance Imaging (MRI) is the optimal neuroimaging in the preoperative diagnosis of glioblastoma and ANA for its multiplanar capability and superior soft tissue contrast. Although some studies demonstrated that the presence of ring-like enhancement and necrosis was suggestive of glioblastoma, in most cases, both glioblastoma and ANA appear as irregular shapes on MR images (hyperdense on T2-weighted sequence and hypodense on T1-weighted sequence) with various degree of Gd-based contrast enhancement and edema, of which the differences were usually imperceptible to the human eye (10–12).

Recently, texture analysis (TA), also known as radiomics, has been widely applied in different fields. Researchers found that TA was a feasible and promising method to facilitate differential diagnosis, since it enabled acquisition of additional quantitative information from MR images which was invisible to human assessment (13–15). TA describes the frequency distribution and the spatial organization of voxel value to reveal the possible differences in tumor tissue (16). Previous studies have explored the feasibility of applying TA in differential diagnosis, subtype classification of tumors and detection of heterogeneity of tumor tissue (17–19). To our acknowledgment, the application of TA in differential diagnosis between glioblastoma and ANA has not been reported yet. The purpose of this study was to evaluate the discriminative ability of MRI texture analysis using machine learning algorithms to differentiate glioblastoma and ANA.

Abbreviations: ANA, Anaplastic astrocytoma; LDA, linear discriminant analysis; MRI, Magnetic Resonance Imaging; MR, Magnetic Resonance; LASSO, Least absolute shrinkage and selection operator; GBDT, Gradient correlation decision tree; CSF, Cerebrospinal fluid; TA, Texture analysis; ROI, Regions of interest; HISTO, Histogram-based matrix; GLCM, Grey-level co-occurrence matrix; GLRLM, Grey-level run length matrix; GLZLM, Grey-level zone length matrix; NGLDM, Neighborhood grey-level dependence matrix; AUC, Area under the receiver operating characteristic curve; PCNSL, Primary central nervous system lymphoma; MLP, Multilayer perceptron; IDH, Isocitrate dehydrogenase.

MATERIALS AND METHODS

Patient Selection

We retrospectively searched our institution database and screened all patients histopathologically diagnosed as glioblastoma or ANA, from January 2015 to December 2018. Eligibility criteria for qualified patients were: (1) conclusive histopathological diagnosis of glioblastoma or ANA; (2) elaborate electronic medical records, especially pathologic material; (3) diagnostic MR scan at our institution before surgical resection. Exclusion criteria were: (1) history of intracranial disease (e.g., brain trauma, intracranial infection or other types of brain tumor), considering the interference of scar tissue on the intensity of the images; (2) presence of motion artifact on MRI; (3) history of treatments before MR scan (e.g., surgery, chemotherapy or radiotherapy); (4) patients who did not reach the criteria for diagnosis of glioblastoma or ANA according to the 2016 WHO classification system. A senior neuropathologist with 10-year experience judged whether the patient met the criteria (the 2016 WHO classification system) for glioblastoma or ANA. The institutional review board approved this retrospective study. The written informed consent was obtained from participants enrolled in this study. The Ethics Committee of Sichuan University and radiology department of our institution have approved of the utilization of the statistics for this study.

MR Image Acquisition

For all patients included in this study, contrast-enhanced T1-weighted sequences were available and were obtained on 3.0T Siemens Trio Scanner with the following parameters: TR/TE/TI = 1900/2.26/900 ms, Flip angle = 9 °, 20 axial slices with thickness = 5 mm, axial FOV = 25.6 × 25.6 cm^2 and data matrix = 256 × 256. Contrast-enhanced T1-weighted imaging used gadopentetate dimeglumine (0.1 mmol/Kg) was the contrast agent for contrast-enhanced image, and multi-directional data of contrast-enhanced MRI were collected during the continuous interval time of 90–250 s.

Texture Extraction

In our study, LifeX package (http://www.lifexsoft.org) was used to extract texture features. Post-contrast T1-weighted (T1C) images were selected for further analysis due to the clear depiction of tumor location and border (20). Region of interest (ROI) was manually drawn slice-by-slice in the axial plane along the lesions on contrast-enhanced images to obtain texture features. Two experienced neurosurgeons, blind to patients' medical records and histopathological diagnosis, drew the ROI followed by editing by a senior radiologist and a senior neurosurgeon. The disagreements were addressed by discussing and consulting with the senior radiologist and the senior neurosurgeon. A total of 40 texture features were extracted from the MRI images, including minValue, meanValue, maxValue,

stdValue, and parameters derived from six matrixes: Histogram-based matrix (HISTO), Shape, Gray-level co-occurrence matrix (GLCM), Gray-level run length matrix (GLRLM), Gray-level zone length matrix (GLZLM), and Neighborhood gray-level dependence matrix (NGLDM).

Features Selection

There were 40 texture features in total derived from six selected matrixes. The explanation of the 40 texture features were shown in **Supplementary Table 1**. The statistics of these texture features were shown in **Supplementary Table 2**. Feature selection was performed to determine relevant features and thereby avoid overfitting. Besides, the machine learning algorithm applied in this study could not take all 40 texture features into analysis. Three independent feature-selection methods were used to select optimal texture features, including Distance Correlation, least absolute shrinkage, and selection operator (LASSO), and gradient correlation decision tree (GBDT). Three subsets of texture features were thereby formed and constituted three different datasets.

Classification

Linear discriminant analysis (LDA) is a robust classification method to separate two classes by searching for the linear combination of predictors that maximizes the separation between groups. In this study, three classification models were established based on LDA algorithm: Distance Correlation + LDA, LASSO + LDA, and GBDT + LDA. Datasets were fed into LDA algorithm, respectively, and independently. Each dataset was randomly split into training and validation group at the ratio of 4:1. The model trained by training group was then applied to the independent validation group to evaluate its performance. To appraise the robustness of LDA algorithm, the procedure was repeated for 100 cycles with different, random and independent case assignment. A confusion matrix was determined using the true assignment from histopathology and predictions of LDA algorithm. The standard sensitivity, specificity, the areas under receiver operating characteristic curve (AUC) and accuracy were calculated for both the training and validation group to reveal the discriminative ability of the models. The comparison of three models (Distance Correlation + LDA, LASSO + LDA, and GBDT + LDA) was carried out to determine the optimal discriminative model for glioblastoma and ANA. The flowchart of MRI classification by texture features is shown in **Figure 1**.

RESULTS

Patients Characteristics

A total of 133 patients with glioblastoma ($n = 76$) or ANA ($n = 57$) fulfilled inclusion criteria. All patients with glioblastoma were enrolled in this study, while 10 patients with ANA were excluded according to the exclusion criteria. Finally, 76 patients with glioblastoma and 46 patients with ANA were included in this study. The mean ages of patients were 46.9 (15–67) and 40.0 (7–69), respectively. All patients underwent surgically tumor resection in our neurosurgery department from 2015 to 2018. **Figure 2** shows two cases of the axial plane of contrast-enhanced images in patients with glioblastoma and ANA.

Glioblastoma vs. ANA

There were three models analyzed in this study, including Distance Correlation + LDA, LASSO + LDA, GBDT + LDA. The texture features used for classification in these models were shown in **Supplementary Table 3**. The performance of each model was presented in **Table 1** (including sensitivity, specificity, accuracy, and AUC of the training and the validation group). LASSO + LDA achieved the best performance with the highest AUCs in both training and validation group. The sensitivity, specificity, accuracy and AUC for its training group were 0.989, 0.993, 0.996, and 0.997, respectively; and for validation group, they were 0.927, 0.989, 0.968, and 0.974, respectively. In addition, Distance Correlation + LDA and GBDT + LDA also showed promising ability to discriminate glioblastoma from ANA, with AUC $\geq$0.95 for both training groups and validation groups.

Figure 3 shows the relationship between the canonical discriminative functions from LASSO + LDA models for the glioblastoma and ANA groups (triangles and circles) and for the group centroids (squares). Minimal overlapping was observed in this figure. Qualitatively, analysis of the data selected by LASSO could separate glioblastoma from ANA. **Figure 4** shows the distribution of the direct LDA function determined for the glioblastoma and ANA for one of the 100 independent training cycles in the data analysis to illustrate the performance of the LASSO + LDA model. There were clear shifts of LDA function values, with left shift for ANA and right shift for glioblastoma.

DISCUSSION

The pre-treatment differential diagnosis between glioblastoma and ANA is important considering the significant difference in treatment strategy and patient prognosis. MR scan, the main radiological preoperative examination for brain tumors, is highly recommended as the good sensitivity in lesion detection. However, the accurate diagnosis before operation is still challenging due to the reason that both tumors present similar characteristics on conventional MR images which are beyond human naked eye assessment (4, 5). In this study, we extracted texture features making quantitative description of images to maximize the utilization of MR examination, with which three LDA-based models were established. The results demonstrated that MRI-based texture analysis combining with LDA algorithm could enable the feasible differentiation between glioblastoma and ANA.

TA is a mathematical approach to characterize the heterogeneity of voxel value on images. It could visualize spatial histologic heterogeneity which is invisible to human eye assessment (21). Theoretically, the characteristics of lesions images could be quantitively analyzed as texture features due to their different enhanced patterns on MR images (22). Moreover, previous studies suggested the textures features could reflect a series of abnormal pathology process of tumor such as edema, effusion, and necrosis, providing a potential mechanism for texture features in discriminating glioblastoma from ANA (23, 24).

Artificial intelligence has been widely explored in recent researches. Combined with texture features extracted from

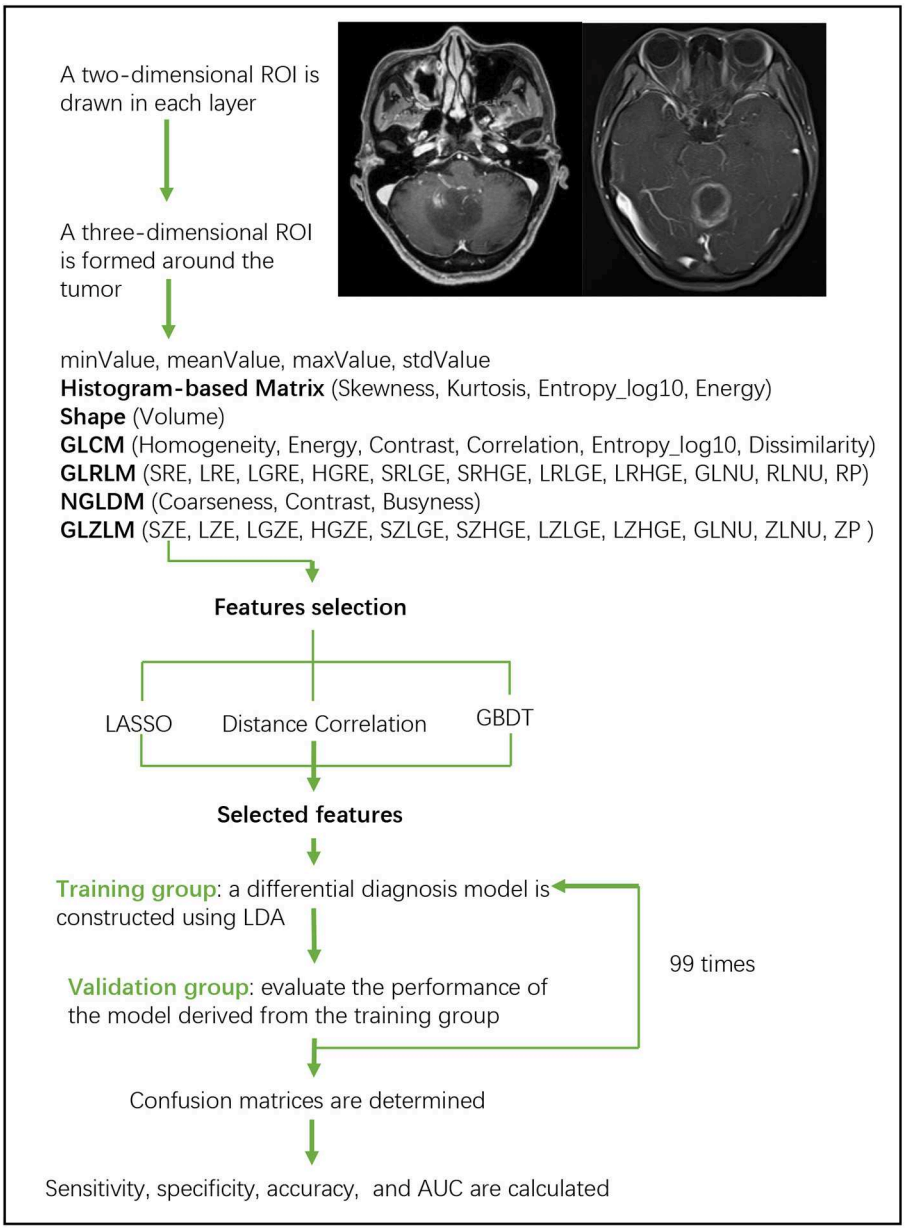


FIGURE 1 | Flowchart of MRI classification by texture features. ANA, anaplastic astrocytoma; LDA, linear discriminant analysis; MRI, Magnetic Resonance Imaging; LASSO, least absolute shrinkage and selection operator; GBDT, gradient correlation decision tree; HISTO, histogram-based matrix; GLCM, Gray-level co-occurrence matrix; GLRLM, Gray-level run length matrix; GLZLM, Gray-level zone length matrix; NGLDM, Neighborhood gray-level dependence matrix; AUC, area under the receiver operating characteristic curve.

images, it was reported to assist in tumor grading, clinical diagnosis, and outcome prediction. A study aimed to evaluate the diagnostic performance of TA-based machine-learning algorithms in differentiating PCNSL from glioblastoma presented optimal performance with the mean AUC of 0.921, while the AUC of three readers were all < 0.8. Thus, the researchers concluded that the diagnostic performance of TA-based machine-learning algorithms was superior to that of human readers (25). Other studies with similar purpose also demonstrated similar results with AUCs higher than 0.85 (18, 26). Moreover, researchers aiming to apply machine learning in astrocytoma grading also reported promising ability in discrimination (27). In our study, the classification models were established based on LDA algorithms. LDA is the statistic classifier combining inputted parameters into a discriminant function to classify cases in different groups (28). Our results demonstrated that LDA-based model represented promising

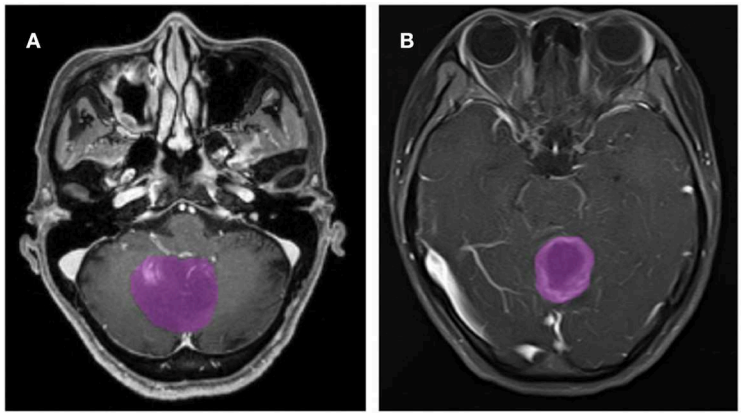


FIGURE 2 | Examples of two cases from the contrast-enhanced MR images in patients with glioblastoma and ANA. **(A)** Contrast-enhanced images with ANA, **(B)** contrast-enhanced images with glioblastoma. ANA, Anaplastic astrocytoma.

TABLE 1 | Discrimination between glioblastoma and ANA.

	Training				Validation			
	Sensitivity	**Specificity**	**Accuracy**	**AUC**	**Sensitivity**	**Specificity**	**Accuracy**	**AUC**
Distance Correlation	0.995	0.979	0.987	0.982	0.996	0.955	0.972	0.966
LASSO	**0.989**	**0.993**	**0.996**	**0.997**	**0.927**	**0.989**	**0.968**	**0.974**
GBDT	0.909	0.991	0.963	0.970	0.918	0.994	0.964	0.972

Entries in bold were most significant. ANA, anaplastic astrocytoma; AUC, area under the receiver operating characteristic curve; LASSO, least absolute shrinkage and selection operator; GBDT, gradient correlation decision tree.

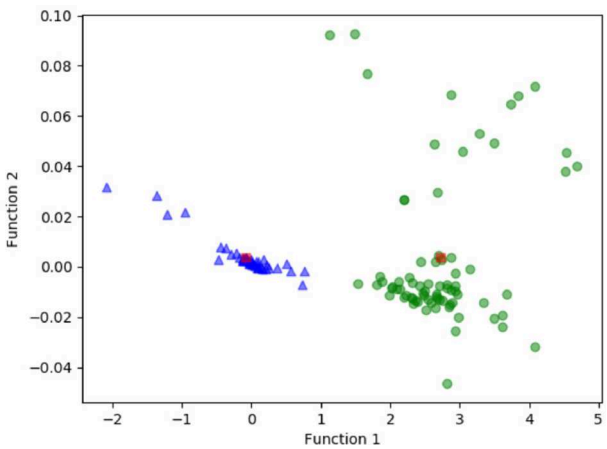


FIGURE 3 | Relationship between the canonical discriminative functions from LASSO + LDA models for the glioblastoma and ANA groups (triangles and circles) and for the group centroids (squares). Minimal overlapping was observed in this figure. Qualitatively, analysis of the data selected by LASSO could separate glioblastoma from ANA. LASSO, least absolute shrinkage and selection operator; LDA, linear discriminant analysis; ANA, Anaplastic astrocytoma.

performance in accurate diagnosis between glioblastoma and ANA.

The adoption on optimal features for machine learning algorithms was challenging but was necessary relative to diagnostic performance. Previous studies perform feature selection with varied methods: Mann-Whitney *U* test with AUC of ROC, Student's *t*-test with recursive feature elimination, random forest, and entropy-based discretization, respectively (18, 25, 29, 30). Based on the results of these studies, we could draw the conclusion that the suitable selection method play a key role in classifier performance. As for our study, a relatively large number of parameters were extracted from different matrixes, increasing the chance in selecting the optimal features but also increasing the difficulty in selection. Therefore, three feature-selection methods (Distance Correlation, LASSO, and GBDT) were evaluated to select the one with best performance. The results of this study demonstrated that LASSO+LDA was the suitable discriminative model for glioblastoma from ANA with highest AUC in the testing group of 0.997. LASSO was proposed as a non-linear variable selection method for neural network in previous study with advantage in minimizing the common sum of squared errors. It could produce interpretable models (similar to the subset selection) when simultaneously exhibiting the stability of ridge regression. Previous study illustrated that it represented superior performance over other state-of-the-art variable selection methods (31). However, we must interpret the results carefully that the additional gain in information from comparing different machine learning techniques is quite limited, specifically given that all classifier/feature selection methods investigated seem perform quite comparably and variance in AUC maybe partially

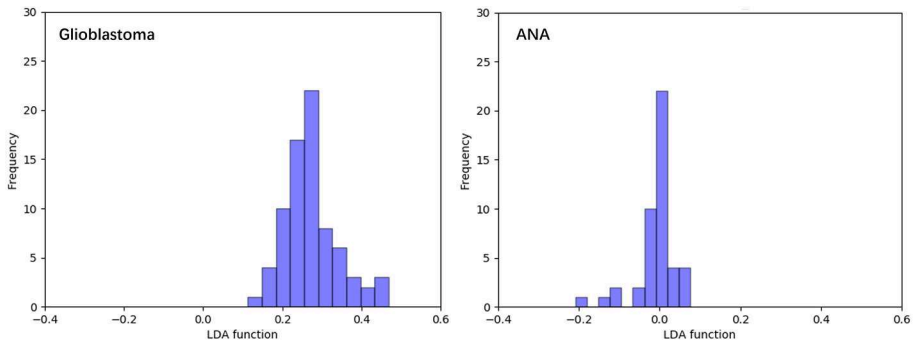


FIGURE 4 | Distribution of the direct LDA function determined for the glioblastoma and ANA for one of the 100 independent training cycles in the data analysis to illustrate the performance of the LASSO + LDA model. There were clear shifts of LDA function values, with left shift for ANA and right shift for glioblastoma. The minimal overlap is observed between the two groups and a strong qualitative similarity is apparent between the plots for cycles and triangles. LDA, linear discriminant analysis; ANA, anaplastic astrocytoma; LASSO, least absolute shrinkage and selection operator.

attributed due to the statistical group. Therefore, our study could only be regarded as hypothesis generation for future, larger studies.

There were some limitations of our study. First, as a retrospective single-center study, the bias in patient selection was inevitable. Second, the number of included patients was relatively small, and greater number of patients were required in further studies to validate the results. Third, ANA is now divided into three categories according to the 2016 World Health Organization Classification of Central Nervous System Tumors: IDH-mutant, IDH- wildtype, and NOS (32). The ability of machine learning in discriminating subtypes of ANA were required to be explored in future studies. Fourth, the machine learning models in our study were not actually validated in other datasets. We did not adopt other institution datasets because that texture features could be different when extracted from images acquired with various scanners or protocols. This could be regarded as a double-edged sword. On the one hand, a set of controlled variables could be provided; on the other hand, the results could not be guaranteed widely applied. The analysis protocol and image processing procedure were open-source packages and study with large population are required to validate and reproduce our results.

CONCLUSION

In this work, we extracted quantitative parameters from contrast-enhanced MR images and used three feature-selection methods to select the most discriminating parameters. Then we applied LDA algorithm to analyze the selected parameters. Our study shows that texture features has promising ability to discriminate glioblastoma from ANA. Multi-center studies with greater number of patients are warranted to confirm this preliminary result.

DATA AVAILABILITY

We are pleased to share our data to any qualified researchers without undue reservation. Please contact corresponding author if there is anything they need.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Sichuan University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

ZT, CC, JX, and XM contributed conception and design of the study. CC and XO enrolled eligible patients and obtained medical records and MRI images of each patient. YF and ZT did texture analysis. JW established the models and performed other statistical analysis. ZT wrote the first draft of the manuscript. CC wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00876/full#supplementary-material

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* (2015) 65:87–108. doi: 10.3322/caac.21262
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* (2016*)* 66:115–32. doi: 10.3322/caac.21338
3. Wen PY, Kesari S. Malignant gliomas in adults. *N Engl J Med.* (2008) 359:492–507. doi: 10.1056/NEJMra0708126
4. Grimm SA, Chamberlain CM. Anaplastic astrocytoma. *CNS Oncol.* (2016) 5:145–57. doi: 10.2217/cns-2016-0002
5. Batash R, Asna N, Schaffer P, Francis N, Schaffer M. Glioblastoma multiforme, diagnosis and treatment; recent literature review. *Curr Med Chem.* (2017) 24:3002–9. doi: 10.2174/0929867324666170516123206
6. Weller M, van den Bent M, Hopkins K, Tonn JC, Stupp R, Falini AE. et al. EANO guideline for the diagnosis and treatment of anaplastic gliomas and glioblastoma. *Lancet Oncol.* (2014) 15:e395–403. doi: 10.1016/s1470-2045(14)70011-7
7. Souhami L, Seiferheld W, Brachman D, Podgorsak EB, Werner-Wasik M, Lustig R, et al. Randomized comparison of stereotactic radiosurgery followed by conventional radiotherapy with carmustine to conventional radiotherapy with carmustine for patients with glioblastoma multiforme: report of Radiation Therapy Oncology Group 93-05 protocol. *Int J Radiat Oncol Biol Phys.* (2004) 60:853–60. doi: 10.1016/j.ijrobp.2004.04.011
8. Burger PC, Heinz ER, Shibata T, Kleihues P. Topographic anatomy and CT correlations in the untreated glioblastoma multiforme. *J Neurosurg.* (1988) 68:698–704. doi: 10.3171/jns.1988.68.5.0698
9. Wick W, Platten M, Meisner C, Felsberg J, Tabatabai G, Simon M, et al. Temozolomide chemotherapy alone versus radiotherapy alone for malignant astrocytoma in the elderly: the NOA-08 randomised, phase 3 trial. *Lancet Oncol.* (2012) 13:707–15. doi: 10.1016/s1470-2045(12)70164-x
10. Chen W, Silverman HD. Advances in evaluation of primary brain tumors. *Semin Nucl Med.* (2008) 38:240–50. doi: 10.1053/j.semnuclmed.2008.02.005
11. Burger PC, Vogel FS, Green SB, Strike AT. Glioblastoma multiforme and anaplastic astrocytoma. Pathologic criteria and prognostic implications. *Cancer.* (1985) 56:1106–11.
12. Omuro A, DeAngelis ML. Glioblastoma and other malignant gliomas: a clinical review. *JAMA.* (2013) 310:1842–50. doi: 10.1001/jama.2013.280319
13. Lakhman Y, Veeraraghavan H, Chaim J, Feier D, Goldman DA, Moskowitz CS, et al. Differentiation of uterine leiomyosarcoma from atypical leiomyoma: diagnostic accuracy of qualitative MR imaging features and feasibility of texture analysis. *Eur Radiol.* (2017) 27:2903–15. doi: 10.1007/s00330-016-4623-9
14. Skogen K, Schulz A, Dormagen JB, Ganeshan B, Helseth E, Server A. Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. *Eur J Radiol.* (2016) 85:824–9. doi: 10.1016/j.ejrad.2016.01.013
15. Gibbs P, Turnbull WL. Textural analysis of contrast-enhanced MR images of the breast. *Magn Reson Med.* (2003*)* 50:92–8. doi: 10.1002/mrm.10496
16. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clin Radiol.* (2004) 59:1061–9. doi: 10.1016/j.crad.2004.07.008
17. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging.* (2012) 3:573–89. doi: 10.1007/s13244-012-0196-6
18. Xiao DD, Yan PF, Wang YX, Osman MS, Zhao YH. Glioblastoma and primary central nervous system lymphoma: preoperative differentiation by using MRI-based 3D texture analysis. *Clin Neurol Neurosurg.* (2018) 173:84–90. doi: 10.1016/j.clineuro.2018.08.004
19. Itakura H, Achrol AS, Mitchell LA, Loya JJ, Liu T, Westbroek EM, et al. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. *Sci Transl Med.* (2015) 7:303ra138. doi: 10.1126/scitranslmed.aaa7582
20. Nioche C, Orlhac F, Boughdad S, Reuze S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* (2018) 78:4786–9. doi: 10.1158/0008-5472.Can-18-0125
21. Hu LS, Ning S, Eschbacher JM, Gaw N, Dueck AC, Smith KA, et al. Multi-parametric MRI and texture analysis to visualize spatial histologic heterogeneity and tumor extent in glioblastoma. *PLoS ONE.* (2015) 10:e0141506. doi: 10.1371/journal.pone.0141506
22. Florez E, Nichols T, Parker EE, Lirertte ST, Howard CM, Fatemi A. Multiparametric magnetic resonance imaging in the assessment of primary brain tumors through radiomic features: a metric for guided radiation treatment planning. *Cureus.* (2018) 10:e3426. doi: 10.7759/cureus.3426
23. Mahmoud-Ghoneim D, Alkaabi MK, de Certaines JD, Goettsche MF. The impact of image dynamic range on texture classification of brain white matter. *BMC Med Imaging.* (2008) 8:18. doi: 10.1186/1471-2342-8-18
24. Ahn SY, Park CM, Park SJ, Kim HJ, Song C, Lee SM, et al. Prognostic value of computed tomography texture features in non-small cell lung cancers treated with definitive concomitant chemoradiotherapy. *Invest Radiol.* (2015) 50:719–25. doi: 10.1097/rli.0000000000000174
25. Suh HB, Choi YS, Bae S, Ahn SS, Chang JH, Kang SG, et al. Primary central nervous system lymphoma and atypical glioblastoma: Differentiation using radiomics approach. *Eur Radiol.* (2018) 28:3832–9. doi: 10.1007/s00330-018-5368-4
26. Mouthuy N, Cosnard G, Abarca-Quinones J, Michoux N. Multiparametric magnetic resonance imaging to differentiate high-grade gliomas and brain metastases. *J Neuroradiol.* (2012) 39:301–7. doi: 10.1016/j.neurad.2011.11.002
27. Glotsos D, Spyridonos P, Cavouras D, Ravazoula P, Dadioti PA, Nikiforidis G. An image-analysis system based on support vector machines for automatic grade diagnosis of brain-tumour astrocytomas in clinical routine. *Med Inform Internet Med.* (2005) 30:179–93. doi: 10.1080/14639230500077444
28. Dellacasa Bellingegni A, Gruppioni E, Colazzo G, Davalli A, Sacchetti R, Guglielmelli E, et al. NLR, MLP, SVM, and LDA: a comparative analysis on EMG data from people with trans-radial amputation. *J Neuroeng Rehabil.* (2017) 14:82. doi: 10.1186/s12984-017-0290-6
29. Ion-Margineanu A, Van Cauter S, Sima DM, Maes F, Sunaert S, Himmelreich U, et al. Classifying glioblastoma multiforme follow-up progressive vs. responsive forms using multi-parametric MRI features. *Front Neurosci.* (2016) 10:615. doi: 10.3389/fnins.2016.00615
30. Fetit AE, Novak J, Peet AC, Arvanitits NT. Three-dimensional textural features of conventional MRI improve diagnostic classification of childhood brain tumours. *NMR Biomed.* (2015) 28:1174–84. doi: 10.1002/nbm.3353
31. Sun K, Huang SH, Wong DS, Jang SS. Design and application of a variable selection method for multilayer perceptron neural network with LASSO. *IEEE Trans Neural Netw Learn Syst.* (2017) 28:1386–96. doi: 10.1109/tnnls.2016.2542866
32. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* (2016) 131:803–20. doi: 10.1007/s00401-016-1545-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 27 September 2019
doi: 10.3389/fonc.2019.00973

Radiomics Facilitates Candidate Selection for Irradiation Stents Among Patients With Unresectable Pancreatic Cancer

Hai-Feng Zhou[1†], Yu-Qi Han[2,3†], Jian Lu[1†], Jing-Wei Wei[3,4†], Jin-He Guo[1†], Hai-Dong Zhu[1†], Ming Huang[5], Jian-Song Ji[6], Wei-Fu Lv[7], Li Chen[1], Guang-Yu Zhu[1], Zhi-Cheng Jin[1], Jie Tian[3,4,8,9] and Gao-Jun Teng[1*]*

[1] Center of Interventional Radiology and Vascular Surgery, Department of Radiology, Zhongda Hospital, Medical School, Southeast University, Nanjing, China, [2] School of Life Science and Technology, Xidian University, Xi'an, China, [3] Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China, [4] University of Chinese Academy of Sciences, Beijing, China, [5] Department of Minimally Invasive Interventional Radiology, Yunnan Tumor Hospital, The Third Affiliated Hospital of Kunming Medical University, Kunming, China, [6] Department of Radiology, Lishui Central Hospital, Wenzhou Medical University, Lishui, China, [7] Department of Interventional Radiology, Anhui Provincial Hospital, The First Affiliated Hospital of University of Science and Technology of China, Hefei, China, [8] Beijing Advanced Innovation Centre for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing, China, [9] Engineering Research Centre of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University, Xi'an, China

OPEN ACCESS

Edited by:
Rong Tian,
Sichuan University, China

Reviewed by:
Zhongxiang Ding,
Hangzhou First People's Hospital, China
Xiaohua Zhu,
Huazhong University of Science and Technology, China

***Correspondence:**
Jie Tian
tian@ieee.org
Gao-Jun Teng
gjteng@vip.sina.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 22 May 2019
Accepted: 13 September 2019
Published: 27 September 2019

Citation:
Zhou H-F, Han Y-Q, Lu J, Wei J-W, Guo J-H, Zhu H-D, Huang M, Ji J-S, Lv W-F, Chen L, Zhu G-Y, Jin Z-C, Tian J and Teng G-J (2019) Radiomics Facilitates Candidate Selection for Irradiation Stents Among Patients With Unresectable Pancreatic Cancer. Front. Oncol. 9:973. doi: 10.3389/fonc.2019.00973

Purpose: To develop a model to select appropriate candidates for irradiation stent placement among patients with unresectable pancreatic cancer with malignant biliary obstruction (UPC-MBO).

Methods: This retrospective study included 106 patients treated with an irradiation stent for UPC-MBO. These patients were randomly divided into a training group (74 patients) and a validation group (32 patients). A clinical model for predicting restenosis-free survival (RFS) was developed with clinical predictors selected by univariate and multivariate analyses. After integrating the radiomics signature, a combined model was constructed to predict RFS. The predictive performance was evaluated with the concordance index (C-index) in both the training and validation groups. The median risk score of progression in the training group was used to divide patients into high- and low-risk subgroups.

Results: Radiomics features were integrated with clinical predictors to develop a combined model. The predictive performance was better in the combined model (C-index, 0.791 and 0.779 in the training and validation groups, respectively) than in the clinical model (C-index, 0.673 and 0.667 in the training and validation groups, respectively). According to the median risk score of 1.264, the RFS was significantly different between the high- and low-risk groups ($p < 0.001$ for the training group, and $p = 0.016$ for the validation group).

Conclusions: The radiomics-based model had good performance for RFS prediction in patients with UPC-MBO who received an irradiation stent. Patients with slow progression should consider undergoing irradiation stent placement for a longer RFS.

Keywords: radiomics, pancreatic cancer, malignant biliary obstruction, irradiation stent, survival

INTRODUCTION

Pancreatic cancer is one of the leading causes of cancer-related death (1), and it has the lowest five-year relative survival rate among those with any type of cancer (approximately 8% for all stages) (2). Less than 20% of patients with pancreatic cancer are candidates for surgical resection (3, 4), and over half of them develop obstructive jaundice (5). Considering that patients with advanced pancreatic cancer have only a 6–10 month median survival, the general treatment is palliative care (6). Chemotherapy, radiotherapy, targeted therapy and immunotherapy are not always used for unresectable pancreatic cancer (UPC) patients due to poor performance status, limited effects and added toxicity (3, 7). Placement of a self-expanding metal stent is the standard palliative care for UPC patients with malignant biliary obstruction (MBO) (8–10). Intraluminal irradiation stents, which combined a self-expanding metal stent with brachytherapy to treat local obstructive lesions, were demonstrated to have better patency and be associated with longer survival than conventional stents (uncovered self-expanding metal stents) for unresectable MBO (11, 12). Although the subgroup analysis of overall survival according to tumor etiology showed better survival for biliary tract cancer, there did not appear to be a significant difference in patients with pancreatic cancer (12). Therefore, it is important to select appropriate candidates with pancreatic cancer to undergo irradiation stent placement, not only for individual and reasonable stent selection, but also for prolonged patency and improved survival.

Currently, different models have been developed to predict survival outcomes in patients with different stages of pancreatic cancer (13–16). A consensus statement also proposed clinical prognostic variables for UPC (17). Moreover, imaging-based or radiomic biomarkers have been reported to be available for the prognostic prediction of patients with pancreatic cancer, based on computed tomography (CT) (18–23), magnetic resonance imaging (24, 25), positron-emission tomography (25–27) and fluorescence microscopic imaging (28) findings. Radiomics, a novel method of in-depth feature analysis, is to quantify and extract the high-throughput imaging features from radiographic images (20). Radiomics, such as texture analysis, reflects different imaging phenotypes and tumor heterogeneity, which can be used to assess survival outcomes and predict treatment response (19, 21, 27). However, there are no tools to predict the survival benefits from irradiation stent placement in patients with UPC-MBO. A predictive model based on clinical and imaging features will offer an objective, convenient and non-invasive method for determining appropriate treatment options and making better clinical decisions, especially critical decisions in patients with UPC-MBO.

In this study, we proposed a novel model incorporating clinical biomarkers and CT radiomics features to predict restenosis-free survival (RFS) for individual patients with UPC-MBO who undergoing irradiation biliary stent placement. According to our proposed model, irradiation stent placement could be recommended for appropriate candidates with slow progression for a longer RFS.

MATERIALS AND METHODS

This multicenter retrospective study was approved by the institutional review boards at all participating centers. The need for informed consent was waived due to the study's retrospective nature.

Patients

Between January 2012 and December 2017, 106 patients (69 males, 37 females; mean age, 66 ± 12 years [standard deviation]; age range, 40–86 years) treated with irradiation stent placement for UPC-MBO from four centers were finally included and randomly divided into a training group (74 patients) and a validation group (32 patients). The sample size calculation is shown in **Appendix E1**. The study design and patient exclusion criteria are illustrated in **Figure 1**.

The inclusion criteria were as follows: (a) age 18 years or older; (b) clinical or histopathological diagnosis of UPC-MBO; (c) unresectable disease due to extensive lesions, metastases, a poor medical condition, or refusal to undergo surgery; (d) initial percutaneous transhepatic biliary stent placement; and (e) standard contrast-enhanced CT performed <2 weeks before stenting. The exclusion criteria were as follows: (a) history of surgical resection, (b) presence of severe infection or organ failure before stenting, (c) Eastern Cooperative Oncology Group score of 4 before stenting, (d) any other anticancer therapy except supportive treatment after stenting, (e) incomplete clinical or CT imaging data, or (f) loss to follow-up.

The following clinical characteristics were recorded: (a) demographics, including sex, age, and body mass index; (b) preprocedural status, including pain as assessed by a visual analog scale score, Eastern Cooperative Oncology Group performance status, prior biliary drainage, history of chemotherapy, history of radiotherapy, and degree of ascites; (c) preprocedural blood biochemical analysis, including total bilirubin, direct bilirubin, direct bilirubin/total bilirubin ratio, carbohydrate antigen (CA) 19-9, CA125, and carcinoembryonic antigen; and (d) parameters related to pancreatic cancer, such as the tumor stage according to the TNM classification system (American Joint Committee on Cancer, 8th ed., 2017) (29), liver metastasis, number of metastatic lesions, and length of obstruction.

A standard percutaneous transhepatic biliary stenting procedure was performed under fluoroscopic guidance with or without ultrasonographic guidance by interventional radiologists with more than 15 years of experience. The irradiation stent consisted of two overlapping parts, an outer 125I seed-loaded stent and an inner conventional uncovered self-expanding metal stent (Nanjing Micro-Tech Co., Ltd., Nanjing, China). The two parts were assembled in the biliary tract during the procedure. The 125I seeds (CIAE-6711; Chinese Atomic Energy Science Institution, Beijing, China) were preloaded into the sheaths that were attached to the outer surface of the stent immediately before the procedure. According to the Treatment Planning System (TPS, FTT Technology Ltd. Co., Beijing, China), the number, dosage, and distribution of the 125I seeds were calculated. The standards of radiation safety and management were performed after irradiation stent placement (30).

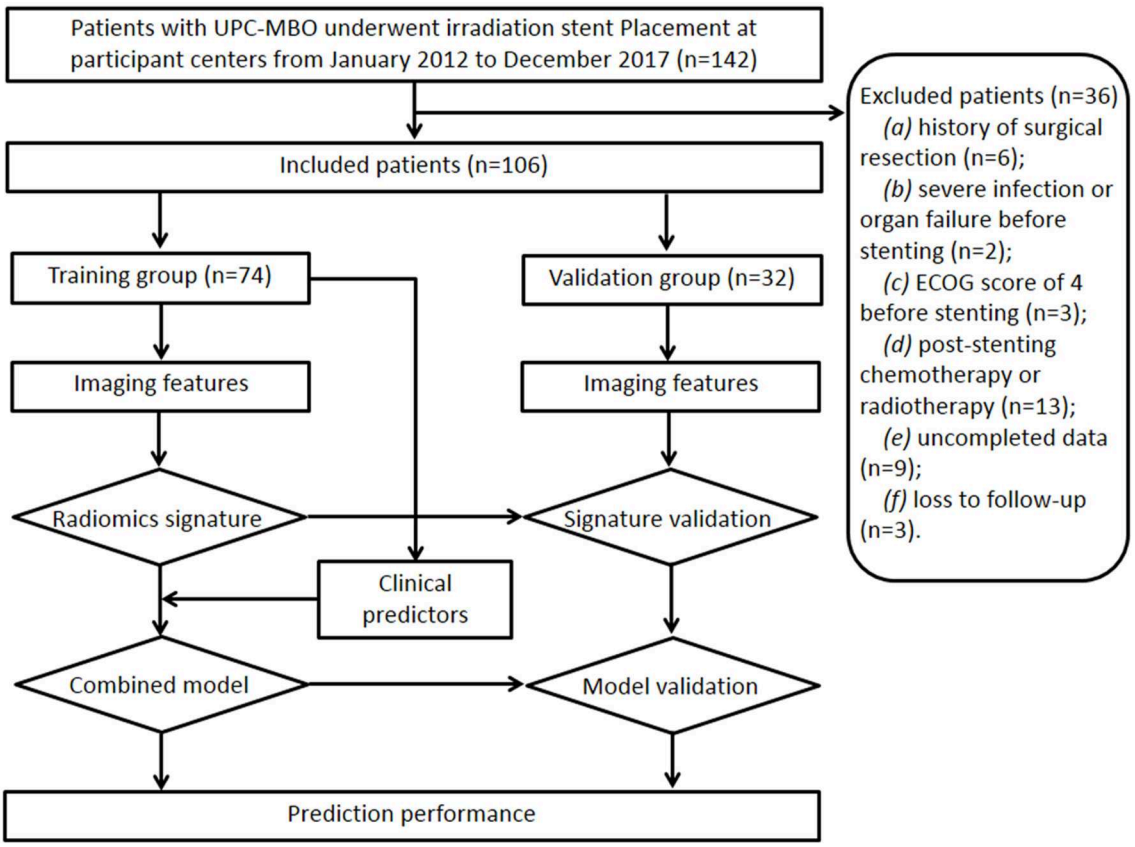


FIGURE 1 | Flowchart of the study design and patient exclusion.

Routine follow-up, including performance status, clinical signs, postprocedural treatment, blood biochemical analysis, and imaging, was performed 1 week after stent placement, monthly for 6 months, and then every 3 months. The endpoint of this study was the occurrence of restenosis or death. RFS was calculated from the date of stenting to the date of the endpoint, which equaled the duration of stent function. Restenosis indicated stent dysfunction, which was defined by clinical signs of recurrent jaundice with elevated bilirubin levels along with biliary reobstruction as evidenced on CT, ultrasound, magnetic resonance cholangiopancreatography, or percutaneous transhepatic cholangiography. RFS was censored at the date of the last follow-up visit for restenosis-free patients.

CT Image Acquisition and Tumor Segmentation

Imaging feature extraction was performed on each patient's CT images within 2 weeks before stent placement. The pancreatic CT scan included an arterial phase and a portal-venous phase, which were used to extract imaging features. The CT acquisition protocols and image preprocessing were described in **Appendix E2**. The region of interest (ROI) was drawn with ITK-SNAP software (version 3.4.0; www.itksnap.org) by an experienced radiologist (reader 1). Each two-dimensional CT image covering the visible tumor region was delineated along the tumor boundaries, and the overlap in the delineated areas was selected as the final ROI (**Figure E1**). The ROIs for the arterial and venous phases were annotated. The stability of radiomics features was verified from the ROI regions that were annotated by two radiologists (reader 1 and reader 2) separately through intra- and interobserver correlation coefficients. Correlation coefficients ranging from 0 to 1 were considered, and values $>$ 0.8 were considered almost perfect agreement (31).

Imaging Feature Extraction

Radiomics features were extracted from the ROIs using MATLAB (version R2018a; Mathworks; Natick, USA), including 25 non-texture features, 51 texture features and wavelet features in wavelet images decomposed on different scales. Non-texture features reflect the shape, size and intensity of tumor lesions, and texture features represent the inherent heterogeneity of tumors based on four textural matrices. In addition, a three-dimensional wavelet transform was applied to decouple the first-order statistical features and texture features for each CT image. Finally, we extracted 620 radiomics features from original CT images and wavelet decompositions in each phase from each patient. The details of these features are exhibited in **Appendix E3** and **Table E1**.

Image Feature Reduction and Radiomics Model Construction

For feature preselection, intra-, and inter-observer coefficients were used to detect the stability of features with a threshold of 0.8. Then, Pearson's correction analysis was applied to identify redundant and collinear features, and features with mutual correlation coefficients >0.9 were excluded. After initial selection, the least absolute shrinkage and selection operator (LASSO)-Cox regression approach was used to identify predictive factors for RFS in the training group (32). Ten-fold cross-validation was used to optimize the regression model to select the most reliable model. The minimum tuning parameter (lambda) was used in LASSO-Cox regression approach. A radiomics signature was constructed by a weighted linear combination of selected features in the arterial and portal-venous phases, separately. A radiomics model was constructed by both radiomics signatures of the two phases using the Cox proportional hazard regression method. The Harrel concordance index (C-index) was calculated to describe the performance of the radiomics model.

Clinical and Combined Models

Clinical and combined models were also built for comparison with the radiomics model. Univariate and multivariate Cox proportional hazard analyses were applied to identify effective clinical predictors. Based on the Cox proportional hazard regression model, a clinical model was constructed with clinical predictors, and the combined model integrated clinical predictors and the radiomics signature. In the combined model, the radiomics signature was calculated as the Rad-score for quantification. The C-index of the clinical and combined models was also calculated to illustrate their performance. The 3-month RFS rate of the combined model was assessed through receiver operating characteristic curve analysis along with the area under the curve. Decision curve analysis was used to compare the net benefit at different threshold probabilities from the clinical and combined nomograms.

Statistical Analysis

Continuous variable is described as mean ± standard deviation, and categorical variable is described as number and percentage. Baseline characteristics between two groups were compared by Student's *t*-test for continuous variables and by Pearson's chi squared or Fisher's exact test for categorical variables. With the R package (version 3.4.4; R Package for Statistical Computing; www.r-project.org), the nomograms were formulated in the training group based on the results of the multivariate analysis and by the Cox proportional hazard regression modeling strategies. Receiver operating characteristic curves were drawn and the area under the curve was calculated to evaluate the discrimination performance for 3-month RFS. Calibration curves were drawn to compare the 3-month RFS between the predicted and actual outcomes using the Hosmer–Lemeshow test. Decision curve analysis was used to evaluate the clinical utility of the nomogram by calculating the net benefit at different threshold probabilities. The combined model generated a risk score for RFS

TABLE 1 | Patient characteristics in the training and validation groups.

Characteristics	Total (n = 106)	Training (n = 74)	Validation (n = 32)	p-value
Age, mean ± SD, years	65.63 ± 11.95	66.41 ± 12.27	63.84 ± 11.71	0.313
Sex, n (%)				0.713
Male	69 (65.1)	49 (66.2)	20 (62.5)	
Female	37 (34.9)	25 (33.8)	12 (37.5)	
BMI, mean ± SD, kg/m^2	20.59 ± 3.07	20.39 ± 3.12	21.05 ± 2.94	0.312
Length of obstruction, mean ± SD, mm	37.67 ± 10.03	37.61 ± 9.91	37.81 ± 10.47	0.924
TB, mean ± SD, μmol/L	185.09 ± 134.44	179.60 ± 137.04	197.78 ± 129.45	0.525
DB, mean ± SD, μmol/L	139.33 ± 97.67	135.80 ± 99.51	147.48 ± 94.32	0.574
DB/TB ratio, mean ± SD	0.758 ± 0.110	0.756 ± 0.115	0.761 ± 0.100	0.829
Pain, n (%)				0.250
None	23 (21.7)	19 (25.7)	4 (12.5)	
Mild	63 (59.4)	43 (58.1)	20 (62.5)	
Moderate or severe	20 (18.9)	12 (16.2)	8 (25)	
T stage, n (%)				0.319
2	10 (9.4)	9 (12.2)	1 (3.1)	
3	11 (10.4)	8 (10.8)	3 (9.4)	
4	85 (80.2)	57 (77)	28 (87.5)	
N stage, n (%)				0.255
0	26 (24.5)	15 (20.3)	11 (34.4)	
1	68 (64.2)	51 (68.9)	17 (53.1)	
2	12 (11.3)	8 (10.8)	4 (12.5)	
M stage, n (%)				0.051
0	68 (64.2)	52 (70.3)	16 (50.0)	
1	38 (35.8)	22 (29.7)	16 (50.0)	
Liver metastasis, n (%)				0.361
No	76 (71.7)	55 (74.3)	21 (65.6)	
Yes	30 (28.3)	19 (25.7)	11 (34.4)	
Number of metastatic lesions, n (%)				0.099
0	68 (64.2)	50 (67.6)	18 (56.3)	
1	12 (11.3)	10 (13.5)	2 (6.3)	
≥2	26 (24.5)	14 (18.9)	12 (37.5)	
Ascites level, n (%)				0.541
None	85 (80.2)	61 (82.4)	24 (75)	
Mild	14 (13.2)	8 (10.8)	6 (18.8)	
Moderate or severe	7 (6.6)	5 (6.8)	2 (6.3)	
Radiotherapy, n (%)				0.137
No	101 (95.3)	72 (97.3)	29 (90.6)	
Yes	5 (4.7)	2 (2.7)	3 (9.4)	

(Continued)

TABLE 1 | Continued

Characteristics	Total ($n = 106$)	Training ($n = 74$)	Validation ($n = 32$)	p-value
Chemotherapy, n (%)				0.775
No	91 (85.8)	64 (86.5)	27 (84.4)	
Yes	15 (14.2)	10 (13.5)	5 (15.6)	
ECOG score, n (%)				0.774
0	3 (2.8)	2 (2.7)	1 (3.1)	
1	11 (10.4)	9 (12.2)	2 (6.3)	
2	60 (56.6)	40 (54.1)	20 (62.5)	
3	32 (30.2)	23 (31.1)	9 (28.1)	
Prior PTBD, n (%)				0.219
No	31 (29.2)	19 (25.7)	12 (37.5)	
Yes	75 (70.8)	55 (74.3)	20 (62.5)	
CA19-9, n (%)				0.349
<1,000 U/ml	57 (53.8)	42 (56.8)	15 (46.9)	
≥1,000 U/ml	49 (46.2)	32 (43.2)	17 (53.1)	
CA125, n (%)				0.660
<35 U/ml	33 (31.1)	24 (32.4)	9 (28.1)	
≥35 U/ml	73 (68.9)	50 (67.6)	23 (71.9)	
CEA, n (%)				0.870
<5 ng/ml	41 (38.7)	29 (39.2)	12 (37.5)	
≥5 ng/ml	65 (61.3)	45 (60.8)	20 (62.5)	

Continuous variable is described as mean ± SD, and categorical variable is described as number and percentage. Baseline characteristics between two groups were compared by Student's t-test for continuous variables and by Pearson's chi squared or Fisher's exact test for categorical variables. SD, standard deviation; BMI, body mass index; TB, total bilirubin; DB, direct bilirubin; ECOG, Eastern Cooperative Oncology Group; PTBD, percutaneous transhepatic biliary drainage; CA, carbohydrate antigen; CEA, carcinoembryonic antigen.

and dichotomized the patients into two groups with different risks of progression using the median risk score in the training group. Kaplan-Meier curves were generated to evaluate the ability of the risk score to stratify the patients, and log-rank tests were applied to assess the statistical significance with $p < 0.05$.

RESULTS

Patients

A total of 106 patients (69 males, 37 females; mean age, 66 ± 12 years [standard deviation]; age range, 40–86 years) were included in this study, including 74 patients in the training group and 32 patients in the validation group. The clinical characteristics showed no significant differences between the two groups (all $p > 0.05$, **Table 1**). During the mean follow-up time of 165.3 days, 99 of 106 (93%) patients reached the endpoint. There was no significant difference in the median RFS between the training group (139.5 days) and the validation group (120 days) ($p = 0.926$).

TABLE 2 | The C-indexes of clinical, radiomic, and combined models.

Models	Training		Validation	
	C-index	95% CI	C-index	95% CI
Clinical model	0.673	(0.594, 0.751)	0.667	(0.541, 0.793)
Arterial phase features	0.735	(0.559, 0.911)	0.719	(0.445, 0.994)
Portal-venous phase features	0.768	(0.523, 1)	0.788	(0.413, 1)
Radiomics signature	0.787	(0.542, 1)	0.796	(0.421, 1)
Combined model	0.791	(0.614, 0.967)	0.779	(0.504, 1)

C-index, concordance index; CI, confidence interval.

Radiomics Features

We extracted 620 features from the arterial and venous phases. After intra- and interobserver agreement analysis, 368 features from the arterial phase and 324 features from the portal-venous phase were retained for collinearity testing (**Figure E2**). A total of 61 features from the arterial phase and 49 features from the portal-venous phase were identified as independent after Pearson's correlation analysis (**Table E2**). The LASSO-Cox model identified that eight features from the arterial phase and six features from the portal-venous phase were most efficient for predicting RFS (**Figure E3**). The eight biomarkers from arterial phase were "glszm_LZHGE," "fos_median," "glszm_SZSE," "glcm_inverse_variance," "fos_minimum," "glcm_IMC2," "glszm_LGLZE," and "glszm_HGLZE." The six biomarkers from portal-venous phase were "glszm_ZSV," "fos_uniformity," "glrlm_SRHGLE," "glcm_correlation," "ngtdm_complexity," and "glszm_HGLZE." These radiomics biomarkers showed no significant difference between the training and validation groups (all $p > 0.05$, **Table E3**).

Radiomics Model

Regarding the LASSO-Cox model, the C-index in the arterial phase was 0.735 and 0.719 for the training and validation groups, respectively; the C-index in the portal venous phase was 0.768 and 0.788 for the training and validation groups, respectively. The radiomics model, which was developed by integrating the radiomics signatures of both phases, yielded higher C-indices of 0.787 and 0.796 for the training and validation groups, respectively (**Table 2**).

The arterial phase score for progression to the endpoint was calculated with the following formula.

$$\begin{aligned}\text{AP\_score} = exp\,(&-0.32651538+ 0.18816560\\ &\times \text{AP\_Coif1\_glszm\_LZHGE} - 0.01513133\\ &\times \text{AP\_Coif2\_fos\_median} - 0.05629135\\ &\times \text{AP\_Coif5\_glszm\_SZSE} - 0.03575725\\ &\times \text{AP\_Coif7\_glcm\_inverse\_variance}+0.10324552\\ &\times \text{AP\_Coif8\_fos\_minimum} - 0.06760264\\ &\times \text{AP\_Coif8\_glcm\_IMC2} - 0.22867284\\ &\times \text{AP\_Coif8\_glszm\_LGLZE}+0.03670083\\ &\times \text{AP\_Coif8\_glszm\_HGLZE})\end{aligned}$$

The portal-venous phase score for progression to the endpoint was calculated with the following formula.

$$\begin{aligned}\text{PP\_score} = exp\,(&-0.29073968+0.14320514\\ &\times\text{PP\_ori\_glszm\_ZSV}+0.05375225\\ &\times\text{PP\_Coif1\_fos\_uniformity}-0.26501636\\ &\times\text{PP\_Coif2\_glrlm\_SRHGLE}-0.03923173\\ &\times\text{PP\_Coif5\_glcm\_correlation}-0.01452813\\ &\times\text{PP\_Coif6\_ngtdm\_complexity}+0.01951407\\ &\times\text{PP\_Coif8\_glszm\_HGLZE})\end{aligned}$$

The total radiomics score for progression to the endpoint was calculated with the following formula.

$$\text{Rad-score} = exp\,(0.463 \times \text{AP\_score}+0.665 \times \text{PP\_score})$$

Clinical and Combined Models

After univariate and multivariate analysis, N stage (HR [95% CI], 1.663 [1.041–2.659]; $p = 0.033$), M stage (HR [95% CI], 2.861 [1.114–7.352]; $p = 0.029$), and CA19-9 (HR [95% CI], 1.898 [1.024–3.520]; $p = 0.042$) were ultimately selected as clinical predictors of RFS (**Table 3**). The C-index for the clinical model was 0.673 in the training group and 0.667 in the validation group. The performance of the combined model was increased when the radiomics signature was added to the model, with a C-index of 0.791 in the training group and 0.779 in the validation group (**Table 2**). The nomograms for the clinical and combined models are shown in **Figure 2**. The performance for predicting 3-month RFS as shown by the area under the receiver operating characteristic curve was better with the combined model than with the clinical model for both groups (**Figure 3**). The calibration curves for the combined model demonstrated good agreement between the predicted and observed probabilities of progression at 3 months with *p*-values of 0.823 for the training group and 0.329 for the validation group (**Figure 4**).

TABLE 3 | The univariate and multivariate analyses for clinical features in training group.

Characteristics	HR	95% CI	*p*-value
UNIVARIATE ANALYSIS			
Age	1.000	(0.980, 1.021)	0.990
Sex	0.948	(0.573, 1.570)	0.948
BMI	0.965	(0.895, 1.040)	0.347
Length of obstruction	0.983	(0.958, 1.010)	0.213
TB	1.000	(0.998, 1.002)	0.962
DB	1.000	(0.998, 1.003)	0.806
DB/TB ratio	1.747	(0.164, 18.626)	0.644
Pain	1.278	(0.853, 1.914)	0.234
T stage	1.251	(0.843, 1.857)	0.265
N stage	1.868	(1.238, 2.818)	0.003*
M stage	2.026	(1.194, 3.435)	0.009*
Liver metastasis	1.518	(0.858, 2.688)	0.152
Number of metastatic lesions	1.559	(1.131, 2.148)	0.007*
Ascites	1.602	(1.050, 2.444)	0.029*
Radiotherapy	1.489	(0.361, 6.146)	0.582
Chemotherapy	0.607	(0.276, 1.331)	0.213
ECOG score	1.096	(0.785, 1.529)	0.592
Prior PTBD	1.211	(0.706, 2.077)	0.487
CA19-9	2.442	(1.454, 4.102)	0.001*
CA125	2.230	(1.286, 3.865)	0.004*
CEA	1.410	(0.870, 2.287)	0.163
MULTIVARIATE ANALYSIS			
N stage	1.663	(1.041, 2.659)	0.033*
M stage	2.861	(1.114, 7.352)	0.029*
Number of metastatic lesions	0.666	(0.345, 1.285)	0.225
Ascites	1.328	(0.825, 2.139)	0.243
CA19-9	1.898	(1.024, 3.520)	0.042*
CA125	1.627	(0.877, 3.016)	0.123

**Data are statistically significant with $p < 0.05$. HR, hazard ratio; CI, confidence interval; BMI, body mass index; TB, total bilirubin; DB, direct bilirubin; ECOG, Eastern Cooperative Oncology Group; PTBD, percutaneous transhepatic biliary drainage; CA, carbohydrate antigen; CEA, carcinoembryonic antigen.*

Clinical Use

The risk score for progression to the endpoint was calculated with the following formula.

$$\begin{aligned}\text{risk score} = exp\,(&1.179075+0.931 \times \text{M}+0.753\\ &\times\text{N}+0.509 \times \text{CA19-9} + 1.139 \times \text{Rad-score})\end{aligned}$$

The median risk score for progression in the training group (score = 1.264) was used to divide patients into high- (score ≥1.264) and low-risk (score < 1.264) groups. Kaplan-Meier curves and the log-rank test indicated significant differences in RFS between the high- and low-risk groups (median RFS: 90 days vs. 198 days, $p < 0.001$ for the training group; and median RFS: 118 days vs. 265 days, $p = 0.016$ for the validation group, **Figure 5**). The risk score also showed satisfactory stratification ability when adjusting to the different subgroups (all $p < 0.05$, **Figure E4**). As shown in **Figure E5**, the decision curve analysis for the individualized nomograms shows the overall net benefit in predicting RFS for the combined model was not inferior to the clinical model, the treat-all-patients scheme, and the treat-none scheme if the threshold probability of a patient was >51.0%.

DISCUSSION

Although irradiation stents have been applied to manage malignant intraluminal obstructive diseases (11, 12, 33–38), it is necessary to optimize the selection of appropriate patients for personalized treatment. In this study, a CT radiomics signature was combined with clinical features to establish an objective, preprocedural, and non-invasive model to select appropriate patients with UPC-MBO for irradiation stent placement. The combined model performed better than the clinical model.

With the combined nomogram, the 3-, 6-, and 12-month RFS probabilities can be calculated for each individual undergoing irradiation stent placement. With the risk score formula, each

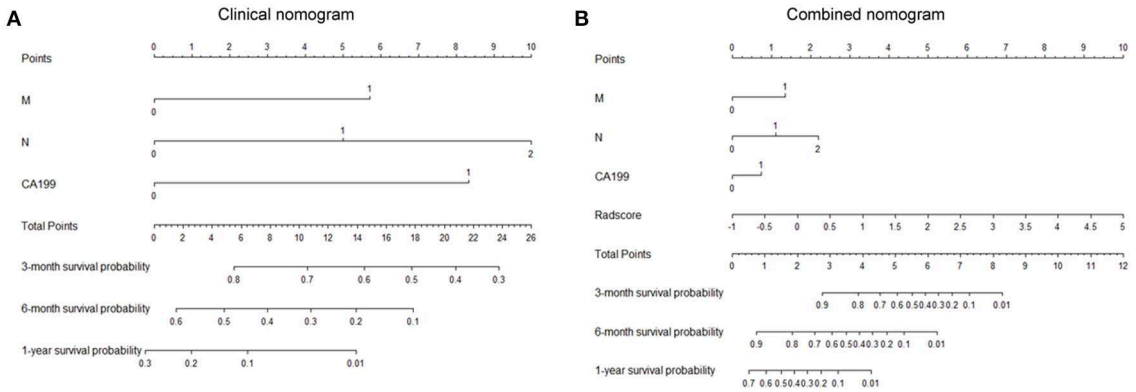


FIGURE 2 | Nomograms for the clinical and combined models. **(A)** Clinical nomogram based on three clinical predictors. **(B)** Combined nomogram based on three clinical predictors and the radiomics signature. To use these nomograms, the user locates an individual patient's value on each variable axis and draws a line up to determine the number of points received for each variable value. The sum of these numbers is located on the axis of total points, and three lines are drawn down to the risk axes to determine the 3-, 6-, and 12-month RFS probabilities.

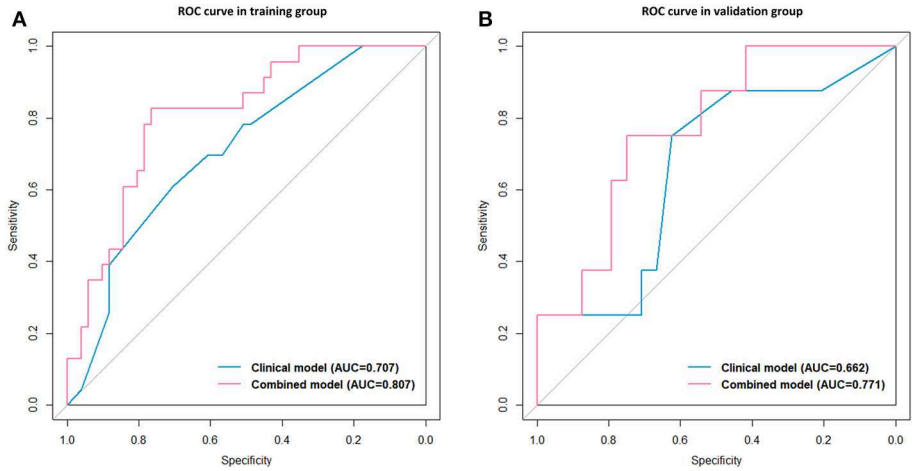


FIGURE 3 | Receiver operating characteristic (ROC) curves with the area under the curve (AUC) for the predictive performance for 3-month RFS. Clinical model vs. combined model in the training group **(A)** and the validation group **(B)**.

individual can be grouped into a low- or high-risk group. Two actual patients were classified using the combined model, as shown in **Appendix E4** and **Table E4**, which demonstrated accurate prediction. "Patient A" with an RFS of 347 days had predicted 3-, 6-, and 12-month RFS probabilities of 0.85, 0.55, and 0.10, respectively, and was grouped into the low-risk group "Patient B" with an RFS of 129 days had predicted 3-, 6-, and 12-month RFS probabilities of 0.65, 0.18 and < 0.01, respectively, and was grouped into the high-risk group. It seems useful for clinical decision making that "Patient A" should undergo irradiation stent placement, but "Patient B" should undergo alternative treatment.

For patients with UPC-MBO, few biomarkers or models with good discrimination have been reported for prognostic prediction. Clinical indexes, including the CA19-9 level and N and M stages, have been applied to develop a model to predict prognosis in this study. The clinical model had a moderate C-index for discrimination (0.673 and 0.667 in the training and validation groups, respectively), while the radiomics signature showed a better C-index (0.787 and 0.796 in the training and validation groups, respectively). This result indicated better predictive performance of radiomic biomarkers than of clinical biomarkers. Moreover, the combined model also performed well with C-indexes of 0.791 and 0.779 in the training and validation groups, respectively. The reason may be that radiomics features from the tumor can provide more information on the cancer phenotype and the tumor microenvironment (39, 40), but clinical characteristics are limited.

As shown in **Figure E4**, regardless of which subgroup the patient was included in, he or she had a longer RFS in the low-risk group than in the high-risk group. Male sex, age older than 65 years, and an abnormal carcinoembryonic antigen level seemed to have less influence on RFS. Recently, researchers have been interested in the role of CA125 in pancreatic cancer (41). Positive CA125 levels may indicate tumor-associated Treg enrichment, which promotes tumor cell escape from the immune system (42).

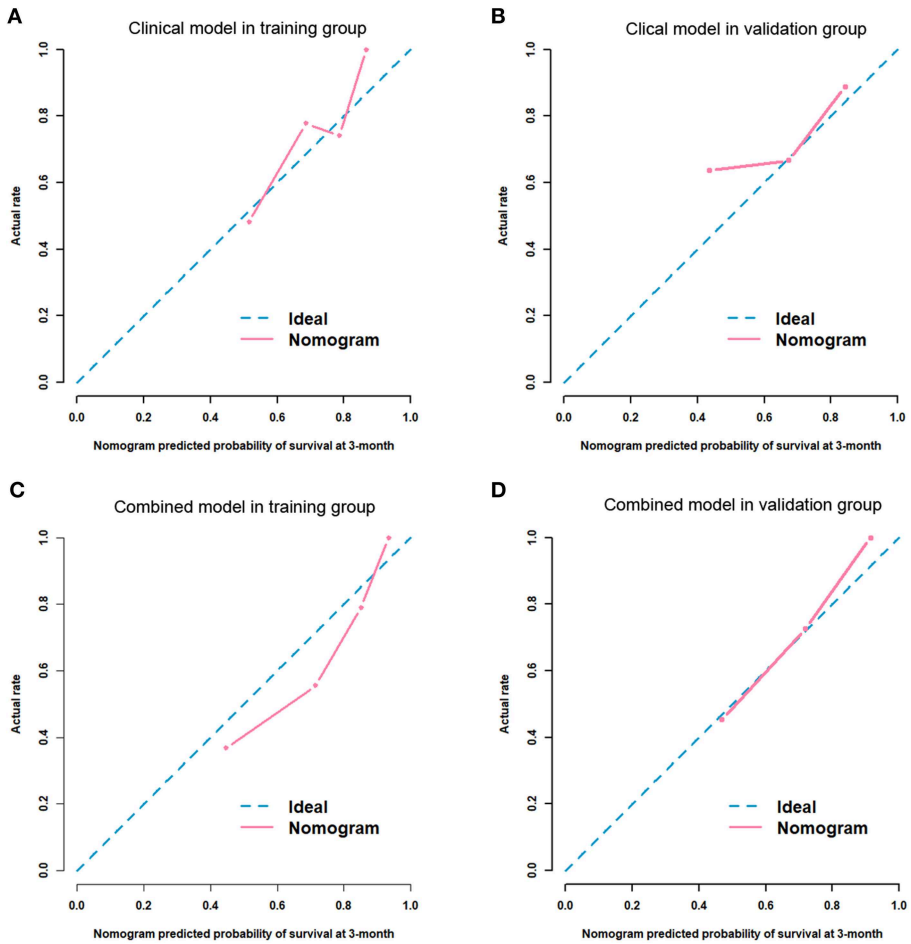


FIGURE 4 | Calibration curves for the predictive performance for 3-month RFS. Clinical model in the training group (**A**, $p = 0.105$) and the validation group (**B**, $p = 0.343$). Combined model in the training group (**C**, $p = 0.823$) and the validation group (**D**, $p = 0.329$).

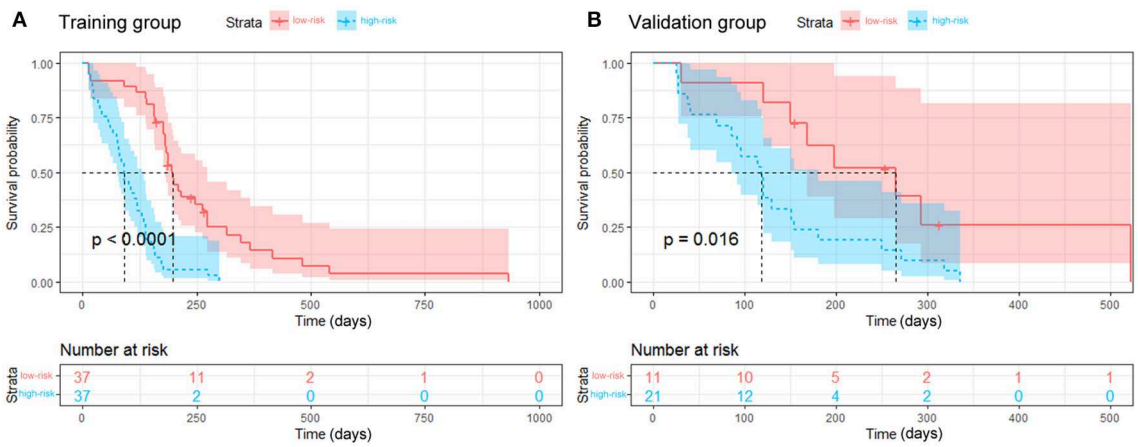


FIGURE 5 | Kaplan-Meier curves for the stratified groups. The low-risk group had a longer RFS than the high-risk group in the training group (**A**, $p < 0.001$) and the validation group (**B**, $p = 0.016$).

A high CA125 level is also associated with a high metabolic tumor burden (43) and poor prognosis (44–46). Although the CA125 level was a potential risk factor for RFS in the univariate analysis, this factor was not ultimately included in the predictive model developed with multivariate analysis (HR [95%]: 1.627 [0.877, 3.016]; $p = 0.123$). The prognostic importance of CA125 in pancreatic cancer should be further evaluated.

Imaging-based texture analysis is used to quantify intratumoral heterogeneity in patients with pancreatic cancer (21, 47, 48). Sandrasegaran et al. (19) demonstrated that contrast-enhanced CT-based radiomics features were associated with survival among patients with UPC, but only two-dimensional texture features from axial slices with maximum tumor dimensions were analyzed rather than features from multiple sections through the whole tumor. Cassinotto et al. (22) and Attiyeh et al. (23) evaluated only CT texture features in patients with surgically resectable pancreatic cancer. In our study, comprehensive radiomics features included intensity, shape, texture, and wavelet features that covered one-, two- and three-dimensional features in both the arterial and portal-venous phases. The radiomics signature based on both phases had good discrimination. In the arterial phase, the C-indexes were 0.735 and 0.719 in the training and validation groups, respectively. In the portal-venous phase, the C-indexes were 0.768 and 0.788 in the training and validation groups, respectively. Currently, few studies have explained the biological mechanisms of radiomics features for predicting treatment outcomes. However, this fact does not compromise the effectiveness and robustness of the proposed model for prognostic prediction.

This study has several limitations. First, as shown in **Table E5**, this study was a retrospective study with a small population. Second, evaluation of data from several independent centers for external validation is needed; however, this study was developed based on a limited sample. Third, the model was mainly used to choose appropriate patients for irradiation stent placement but was less able to predict the prognosis of patients who underwent placement of other stents or drainage mechanisms. Therefore, additional trials with large samples are needed to prospectively validate the findings in several independent centers. Radiogenomics-based studies are proposed for personalized treatment with radiotherapy or irradiation-related interventions for patients with pancreatic cancer.

CONCLUSIONS

In conclusion, the proposed model based on radiomics had good performance for RFS prediction in patients with UPC-MBO who underwent irradiation stent placement. Patients with slow progression should consider undergoing irradiation stent placement for a longer RFS. With further sufficient validation and future clinical trials, this model might be an important tool for clinical decision making in interventional oncology.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This study was approved by the institutional review boards at all participating centers, including IEC for Clinical Research of Zhongda Hospital, Southeast University, IEC for Clinical Research of Yunnan Tumor Hospital, IEC for Clinical Research of Lishui Central Hospital, and IEC for Clinical Research of Anhui Provincial Hospital.

AUTHOR CONTRIBUTIONS

G-JT, H-FZ, Y-QH, JL, J-HG, and H-DZ contributed to the study concept and design. MH, J-SJ, W-FL, and G-YZ contributed to perform stent placement. JL, LC, and Z-CJ contributed to the acquisition of clinical data. J-WW contributed to the statistical analysis. H-FZ, Y-QH, H-DZ, and J-HG contributed to data interpretation. H-FZ wrote the first draft of the manuscript. Y-QH wrote the sections of the manuscript. G-JT, JT, H-DZ, and J-HG supervised and oversaw the study. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was supported by National Key Scientific Instrument and Equipment Development Projects of China (81827805), Innovation Platform of Jiangsu Provincial Medical Center (YXZXA2016005), National Natural Science Foundation of China (81520108015, 81527805 and 81771924), National Key Research and Development Program of China (2017YFC1308700, 2017YFC1309100 and 2017YFA0205200), Chinese Academy of Sciences under Grant (GJJSTD20170004 and QYZDJ-SSW-JSC005). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

The authors thank Mr. Di Dong from Institute of Automation, Chinese Academy of Sciences, Beijing, China, for his assistance with research cooperation. The authors thank Dr. Qi Zhang, Dr. Yong Wang, and Dr. Jun-Ying Wang from Zhongda Hospital, Southeast University, Nanjing, China, for their work in the management of patients. The authors thank Mr. Hong-Yi Zhang, Mr. Peng Zeng, Ms. Rui Fan, Mr. Hai Zhou, and Mr. Sheng-Wei Li from Medical School, Southeast University, Nanjing, China, for their assistance with acquisition of data. The authors thank Dr. Dong Lu from Anhui Provincial Hospital, The First Affiliated Hospital of USTC, Hefei, China, Dr. Bo Peng from Yunnan Tumor Hospital, the Third Affiliated Hospital of Kunming Medical University, Kunming, China, and Dr. Jing-Jing Song from Lishui Central Hospital, Wenzhou Medical University, Lishui, China, for their efforts in the follow-up of patients.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00973/full#supplementary-material

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2018) 68:394–424. doi: 10.3322/caac.21492
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. (2018) 68:7–30. doi: 10.3322/caac.21442
3. Ryan DP, Hong TS, Bardeesy N. Pancreatic adenocarcinoma. *N Engl J Med*. (2014) 371:1039–49. doi: 10.1056/NEJMra1404198
4. Pu LZCT, Singh R, Loong CK, de Moura EGH. Malignant biliary obstruction: evidence for best practice. *Gastroent Res Pract*. (2016) 2016:3296801. doi: 10.1155/2016/3296801
5. Payne M, Burmeister EA, Waterhouse M, Jordan SJ, O'Connell DL, Merrett ND, et al. Biliary stenting in patients with pancreatic cancer: results from a population-based cohort study. *Pancreas*. (2018) 47:80–6. doi: 10.1097/MPA.0000000000000960
6. Eskander MF, Bliss LA, Tseng JF. Pancreatic adenocarcinoma. *Curr Prob Surg*. (2016) 53:107–54. doi: 10.1067/j.cpsurg.2016.01.001
7. Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. *Lancet*. (2016) 388:73–85. doi: 10.1016/S0140-6736(16)00141-0
8. NCCN. *NCCN Clinical Practice Guidelines in Oncology: Pancreatic Adenocarcinoma, Version 2*. (2018). Available online at: http://www.nccn.org/professionals/physician_gls/pdf/pancreatic.pdf (accessed September 15, 2018).
9. Ducreux M, Cuhna AS, Caramella C, Hollebecque A, Burtin P, Goere D, et al. Cancer of the pancreas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. (2015) 26 (Suppl. 5):v56–68. doi: 10.1093/annonc/mdv295
10. Kapoor BS, Mauri G, Lorenz JM. Management of biliary strictures: state-of-the-art review. *Radiology*. (2018) 2018:172424. doi: 10.1055/b-0038-162882
11. Zhu HD, Guo JH, Zhu GY, He SC, Fang W, Deng G, et al. A novel biliary stent loaded with (125)I seeds in patients with malignant biliary obstruction: preliminary results versus a conventional biliary stent. *J Hepatol*. (2012) 56:1104–11. doi: 10.1016/j.jhep.2011.12.018
12. Zhu HD, Guo JH, Huang M, Ji JS, Xu H, Lu J, et al. Irradiation stents vs. conventional metal stents for unresectable malignant biliary obstruction: a multicenter trial. *J Hepatol*. (2018) 68:970–7. doi: 10.1016/j.jhep.2017.12.028
13. Choi SH, Park SW, Seong J. A nomogram for predicting survival of patients with locally advanced pancreatic cancer treated with chemoradiotherapy. *Radiother Oncol*. (2018) 129:340–6. doi: 10.1016/j.radonc.2018.08.006
14. Nakamura T, Asano T, Okamura K, Tsuchikawa T, Murakami S, Kurashima Y, et al. A preoperative prognostic scoring system to predict prognosis for resectable pancreatic cancer: who will benefit from upfront surgery? *J Gastrointest Surg*. (2018) 23:990–6. doi: 10.1007/s11605-018-3972-x
15. Truty MJ, Kendrick ML, Nagorney DM, Smoot RL, Cleary SP, Graham RP, et al. Factors predicting pesponse, perioperative outcomes, and survival following total neoadjuvant therapy for borderline/locally advanced pancreatic cancer. *Ann Surg*. (2019) 2019:3284. doi: 10.1097/SLA.0000000000003284
16. Gargiulo P, Dietrich D, Herrmann R, Bodoky G, Ruhstaller T, Scheithauer W, et al. Predicting mortality and adverse events in patients with advanced pancreatic cancer treated with palliative gemcitabine-based chemotherapy in a multicentre phase III randomized clinical trial: the APC-SAKK risk scores. *Ther Adv Med Oncol*. (2019) 11:386143339. doi: 10.1177/1758835918818351
17. Ter Veer E, van Rijssen LB, Besselink MG, Mali R, Berlin JD, Boeck S, et al. Consensus statement on mandatory measurements in pancreatic cancer trials (COMM-PACT) for systemic treatment of unresectable disease. *Lancet Oncol*. (2018) 19:e151–60. doi: 10.1016/S1470-2045(18)30098-6
18. Amer AM, Zaid M, Chaudhury B, Elganainy D, Lee Y, Wilke CT, et al. Imaging-based biomarkers: changes in the tumor interface of pancreatic ductal adenocarcinoma on computed tomography scans indicate response to cytotoxic therapy. *Cancer-Am Cancer Soc*. (2018) 124:1701–9. doi: 10.1002/cncr.31251
19. Sandrasegaran K, Lin Y, Asare-Sawiri M, Taiyini T, Tann M. CT texture analysis of pancreatic cancer. *Eur Radiol*. (2019) 29:1067–73. doi: 10.1007/s00330-018-5662-1
20. Eilaghi A, Baig S, Zhang Y, Zhang J, Karanicolas P, Gallinger S, et al. CT texture features are associated with overall survival in pancreatic ductal adenocarcinoma—a quantitative analysis. *BMC Med Imaging*. (2017) 17:38. doi: 10.1186/s12880-017-0209-5
21. Chakraborty J, Langdon-Embry L, Cunanan KM, Escalon JG, Allen PJ, Lowery MA, et al. Preliminary study of tumor heterogeneity in imaging predicts two year survival in pancreatic cancer patients. *PLoS ONE*. (2017) 12:e188022. doi: 10.1371/journal.pone.0188022
22. Cassinotto C, Chong J, Zogopoulos G, Reinhold C, Chiche L, Lafourcade JP, et al. Resectable pancreatic adenocarcinoma: role of CT quantitative imaging biomarkers for predicting pathology and patient outcomes. *Eur J Radiol*. (2017) 90:152–8. doi: 10.1016/j.ejrad.2017.02.033
23. Attiyeh MA, Chakraborty J, Doussot A, Langdon-Embry L, Mainarich S, Gonen M, et al. Survival prediction in pancreatic ductal adenocarcinoma by quantitative computed tomography image analysis. *Ann Surg Oncol*. (2018) 25:1034–42. doi: 10.1245/s10434-017-6323-3
24. Nishiofuku H, Tanaka T, Marugami N, Sho M, Akahori T, Nakajima Y, et al. Increased tumour ADC value during chemotherapy predicts improved survival in unresectable pancreatic cancer. *Eur Radiol*. (2016) 26:1835–42. doi: 10.1007/s00330-015-3999-2
25. Chen BB, Tien YW, Chang MC, Cheng MF, Chang YT, Wu CH, et al. PET/MRI in pancreatic and periampullary cancer: correlating diffusion-weighted imaging, MR spectroscopy and glucose metabolic activity with clinical stage and prognosis. *Eur J Nucl Med Mol Imaging*. (2016) 43:1753–64. doi: 10.1007/s00259-016-3356-y
26. Hyun SH, Kim HS, Choi SH, Choi DW, Lee JK, Lee KH, et al. Intratumoral heterogeneity of 18F-FDG uptake predicts survival in patients with pancreatic ductal adenocarcinoma. *Eur J Nucl Med Mol I*. (2016) 43:1461–8. doi: 10.1007/s00259-016-3316-6
27. Cui Y, Song J, Pollom E, Alagappan M, Shirato H, Chang DT, et al. Quantitative analysis of (18)F-fluorodeoxyglucose positron emission tomography identifies novel prognostic imaging biomarkers in locally advanced pancreatic cancer patients treated with stereotactic body radiation therapy. *Int J Radiat Oncol Biol Phys*. (2016) 96:102–9. doi: 10.1016/j.ijrobp.2016.04.034
28. Kim H, Buchsbaum DJ, Zinn KR. A novel imaging biomarker extracted from fluorescence microscopic imaging of TRA-8/DR5 oligomers predicts TRA-8 therapeutic efficacy in breast and pancreatic cancer mouse models. *Mol Imaging Biol*. (2016) 18:325–33. doi: 10.1007/s11307-015-0913-x
29. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK, et al. *AJCC Cancer Staging Manual*. Oxford: Springer (2017). doi: 10.1007/978-3-319-40618-3
30. International CORP. Radiation safety aspects of brachytherapy for prostate cancer using permanently implanted sources. A report of ICRP Publication 98. *Ann ICRP*. (2005) 35:1–11. doi: 10.1016/j.icrp.2006.02.001
31. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. (1977) 33:159–74. doi: 10.2307/2529310
32. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. (1997) 16:385–95.
33. Lu J, Guo JH, Zhu HD, Zhu GY, Wang Y, Zhang Q, et al. Palliative treatment with radiation-emitting metallic stents in unresectable Bismuth type III or IV hilar cholangiocarcinoma. *ESMO Open*. (2017) 2:e242. doi: 10.1136/esmoopen-2017-000242
34. Lu J, Guo JH, Zhu HD, Zhu GY, Chen L, Teng GJ. Safety and efficacy of irradiation stent placement for malignant portal vein thrombus combined with transarterial chemoembolization for hepatocellular carcinoma: a single-center experience. *J Vasc Interv Radiol*. (2017) 28:786–94. doi: 10.1016/j.jvir.2017.02.014
35. Wang Y, Guo JH, Zhu GY, Zhu HD, Chen L, Lu J, et al. A novel self-expandable, radioactive airway stent loaded with (125)I seeds: a feasibility and safety study in healthy beagle dog. *Cardiovasc Intervent Radiol*. (2017) 40:1086–93. doi: 10.1007/s00270-017-1639-8
36. Wang Y, Lu J, Guo JH, Zhu GY, Zhu HD, Chen L, et al. A novel tracheobronchial stent loaded with 125 I seeds in patients with malignant airway obstruction compared to a conventional stent: a prospective randomized controlled study. *Ebio Med*. (2018) 33:269–75. doi: 10.1016/j.ebiom.2018.06.006

37. Guo JH, Teng GJ, Zhu GY, He SC, Fang W, Deng G, et al. Self-expandable esophageal stent loaded with 125I seeds: initial experience in patients with advanced esophageal cancer. *Radiology*. (2008) 247:574–81. doi: 10.1148/radiol.2472070999
38. Zhu HD, Guo JH, Mao AW, Lv WF, Ji JS, Wang WH, et al. Conventional stents versus stents loaded with (125)iodine seeds for the treatment of unresectable oesophageal cancer: a multicentre, randomised phase 3 trial. *Lancet Oncol*. (2014) 15:612–9. doi: 10.1016/S1470-2045(14)70131-7
39. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169
40. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
41. Liu L, Xiang J, Chen R, Fu D, Hong D, Hao J, et al. The clinical utility of CA125/MUC16 in pancreatic cancer: a consensus of diagnostic, prognostic and predictive updates by the Chinese Study Group for Pancreatic Cancer (CSPAC). *Int J Oncol*. (2016) 48:900–7. doi: 10.3892/ijo.2015.3316
42. Fan K, Yang C, Fan Z, Huang Q, Zhang Y, Cheng H, et al. MUC16 C terminal-induced secretion of tumor-derived IL-6 contributes to tumor-associated Treg enrichment in pancreatic cancer. *Cancer Lett*. (2018) 418:167–75. doi: 10.1016/j.canlet.2018.01.017
43. Shi S, Ji S, Qin Y, Xu J, Zhang B, Xu W, et al. Metabolic tumor burden is associated with major oncogenomic alterations and serum tumor markers in patients with resected pancreatic cancer. *Cancer Lett*. (2015) 360:227–33. doi: 10.1016/j.canlet.2015.02.014
44. Yang C, Cheng H, Luo G, Lu Y, Guo M, Jin K, et al. The metastasis status and tumor burden-associated CA125 level combined with the CD4/CD8 ratio predicts the prognosis of patients with advanced pancreatic cancer: a new scoring system. *Eur J Surg Oncol*. (2017) 43:2112–8. doi: 10.1016/j.ejso.2017.07.010
45. Xu H, Liu L, Xiang J, Wang W, Qi Z, Wu C, et al. Postoperative serum CEA and CA125 levels are supplementary to perioperative CA19-9 levels in predicting operative outcomes of pancreatic ductal adenocarcinoma. *Surgery*. (2017) 161:373–84. doi: 10.1016/j.surg.2016.08.005
46. Luo G, Xiao Z, Long J, Liu Z, Liu L, Liu C, et al. CA125 is superior to CA19-9 in predicting the resectability of pancreatic cancer. *J Gastrointest Surg*. (2013) 17:2092–8. doi: 10.1007/s11605-013-2389-9
47. Al-Hawary MM, Francis IR, Chari ST, Fishman EK, Hough DM, Lu DS, et al. Pancreatic ductal adenocarcinoma radiology reporting template: consensus statement of the Society of Abdominal Radiology and the American Pancreatic Association. *Gastroenterology*. (2014) 146:291–304. doi: 10.1053/j.gastro.2013.11.004
48. Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. *Radiology*. (2013) 269:8–14. doi: 10.1148/radiol.13122697

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 27 September 2019
doi: 10.3389/fonc.2019.00934

Prediction of Anal Cancer Recurrence After Chemoradiotherapy Using Quantitative Image Features Extracted From Serial ^{18}F-FDG PET/CT

*Jiahui Wang[1], Hao Zhang[1], Michael Chuong[1,2], Kujtim Latifi[3], Shan Tan[1,4], Wookjin Choi[1,5], Sarah Hoffe[3], Ravi Shridhar[3] and Wei Lu[1,5]**

[1] Department of Radiation Oncology, University of Maryland Baltimore, Baltimore, MD, United States, [2] Miami Cancer Institute, Baptist Hospital of Miami, Miami, FL, United States, [3] Department of Radiation Oncology, Moffitt Cancer Center, Tampa, FL, United States, [4] School of Automation, Huazhong University of Science and Technology, Wuhan, Hubei, China, [5] Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, United States

OPEN ACCESS

Edited by:
Chunxiao Guo,
University of Texas MD Anderson Cancer Center, United States

Reviewed by:
Seunghwan Moon,
Samsung Medical Center, South Korea
Laurence Gluch,
The Strathfield Breast Centre, Australia

***Correspondence:**
Wei Lu
luw@mskcc.org

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 25 June 2019
Accepted: 06 September 2019
Published: 27 September 2019

Citation:
Wang J, Zhang H, Chuong M, Latifi K, Tan S, Choi W, Hoffe S, Shridhar R and Lu W (2019) Prediction of Anal Cancer Recurrence After Chemoradiotherapy Using Quantitative Image Features Extracted From Serial ^{18}F-FDG PET/CT. Front. Oncol. 9:934. doi: 10.3389/fonc.2019.00934

We extracted image features from serial ^{18}F-labeled fluorodeoxyglucose (FDG) positron emission tomography (PET) / computed tomography (CT) scans of anal cancer patients for the prediction of tumor recurrence after chemoradiation therapy (CRT). Seventeen patients (4 recurrent and 13 non-recurrent) underwent three PET/CT scans at baseline (Pre-CRT), in the middle of the treatment (Mid-CRT) and post-treatment (Post-CRT) were included. For each patient, Mid-CRT and Post-CRT scans were aligned to Pre-CRT scan. Comprehensive image features were extracted from CT and PET (SUV) images within manually delineated gross tumor volume, including geometry features, intensity features and texture features. The difference of feature values between two time points were also computed and analyzed. We employed univariate logistic regression model, multivariate model, and naïve Bayesian classifier to analyze the image features and identify useful tumor recurrent predictors. The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the accuracy of the prediction. In univariate analysis, six geometry, three intensity, and six texture features were identified as significant predictors of tumor recurrence. A geometry feature of Roundness between Post-CRT and Pre-CRT CTs was identified as the most important predictor with an AUC value of 1.00 by multivariate logistic regression model. The difference of Number of Pixels on Border (geometry feature) between Post-CRT and Pre-CRT SUVs and Elongation (geometry feature) of Post-CRT CT were identified as the most useful feature set (AUC = 1.00) by naïve Bayesian classifier. To investigate the early prediction ability, we used features only from Pre-CRT and Mid-CRT scans. Orientation (geometry feature) of Pre-CRT SUV, Mean (intensity feature) of Pre-CRT CT, and Mean of Long Run High Gray Level Emphasis (LRHGLE) (texture feature) of Pre-CRT CT were identified as the most important feature set (AUC = 1.00) by multivariate logistic regression model. Standard deviation (intensity

feature) of Mid-CRT SUV and difference of Mean of LRHGLE (texture feature) between Mid-CRT and Pre-CRT SUVs were identified as the most important feature set (AUC = 0.86) by naïve Bayesian classifier. The experimental results demonstrated the potential of serial PET/CT scans in early prediction of anal tumor recurrence.

Keywords: serial PET/CT, chemoradiation therapy, recurrence prediction, image analysis, anal cancer

INTRODUCTION

Anal cancer is a relatively uncommon malignancy. In the United States, the National Cancer Institute estimated 8,580 new cases and 1,160 deaths from anal cancer in 2018[1]. Chemoradiation therapy (CRT) is preferred over abdominoperineal resection for the treatment of anal cancer patients because of sphincter preservation, although surgery can be an effective salvage option (1–4). After CRT, early detection of tumor recurrence is important for initiating salvage surgery and preventing the spread of disease to distant sites (5, 6). Current guideline recommendations for treatment response evaluation after CRT includes serial digital rectal examination with biopsy of clinically progressive lesions, beginning 8–12 weeks after therapy is completed. However, early detection of residual and progressive disease can sometimes be challenging because of treatment-related mucositis and dermatitis that may limit adequate physical examination (5). Alternatively, as a non-invasive evaluation tool, anatomical imaging techniques (CT, ultrasound, and MRI) have been widely used in the tumor staging and treatment response evaluation. Because the region of anal tumors has similar intensity to the surrounding normal structures in the anatomical images and tumor margins may blend with surrounding normal tissues (5), these techniques may fail to accurately assess the presence of tumor.

Positron emission tomography (PET) scans provide metabolic information of tumors and can assist in differentiating recurrent tumors from surrounding tissue. As such, ^{18}F-FDG PET scans obtained after CRT has been increasingly used for the anal cancer recurrence prediction (7, 8) alongside CT scans (^{18}F-FDG PET/CT), which provide corresponding anatomic information. However, there are few reports in the literature about the usage of interim PET/CT scans (Mid-CRT) obtained during CRT to assess treatment response for anal cancer. In a recent paper of Hong et al. (9), they reported a anal cancer chemoradiation treatment evaluation study using pretreatment and interim PET/CT scans. Some commonly used standardized uptake value (SUV) based image features were evaluated, including maximum SUV (SUV_{max}), mean SUV (SUV_{mean}), metabolic tumor volume (MTV), and total lesion glycolysis (TLG). However, no correlation between tumor recurrence and relative change of those image features was detected.

Recent studies have shown that spatial PET/CT features are more informative than the commonly used SUV based measures (10, 11). In this study, instead of conventional SUV image features, comprehensive image features from both CT and SUV of serial PET/CT scans (Pre-CRT, Mid-CRT, and Post-CRT) were analyzed to identify useful image features for the prediction of anal cancer recurrence. In particular, we investigated the image features from Pre-CRT and Mid-CRT PET/CT scans only to explore their potential in the early prediction of tumor recurrence. The image features we used in this study include geometry, intensity, and texture features.

MATERIALS AND METHODS

Patients and PET/CT Scans

IRB approval was obtained from both institutions involved in the analysis. The PET/CT scans used in this study were collected from Department of Radiation Oncology of one major cancer center in the United States for patients with non-metastatic squamous cell carcinoma of the anal canal treated with definitive CRT between 2008 and 2010. Seventeen patients were included in this study, including 4 recurrent and 13 non-recurrent. Clinical characteristics of the patients were shown in **Table 1**. One of the recurrent patients developed a locoregional recurrence and the other three were diagnosed with distant metastasis. The typical approach and clinical outcomes for managing anal cancer with CRT from this institution have been previously published (12). Patients received CRT for 32–50 total elapsed days (median 43 days) with initial prescription gross tumor volume (GTV) dose of 36–50 Gy and GTV boost dose of 0–22 Gy. The median GTV total dose was 56 Gy (range 50–62.5 Gy). Pre-CRT PET/CT imaging was performed 6–46 days before CRT (median 20 days), Mid-CRT PET/CT was performed 22–38 days (median 32 days) after the starting of CRT, and Post-CRT imaging was performed 42–141 days (median 88 days) after completion of CRT. All PET/CT scans were acquired with a Discovery-VCT (DVCT) scanner (GE Medical System, Milwaukee, USA). Each patient fasted for a minimum of 6 h before intravenous injection of 10 mCi ^{18}F-FDG. Whole-body PET and CT imaging was started 90 min after tracer injection. The CT scans were acquired with 110 mAs and 120 kVp. Each slice had a matrix size of 512 × 512 pixels; the pixel was 1.37 × 1.37 mm with a 12-bit gray-level in Hounsfield Units (HU). The slice thickness was 3.27 mm. PET images were attenuation corrected with a matrix size of 128 × 128 pixels; the pixel was 5.47 × 5.47 mm. The slice thickness was 3.27 mm. To compare the PET-based image features between patients, standardized uptake value (SUV) was calculated on a voxel-by-voxel basis and a SUV image was created for each PET scan.

An experienced radiation oncologist (MC) contoured the GTVs in the Pre-, Mid-, and Post-CRT PET/CT scans, respectively, using mainly CT while referring to PET. In general,

[1]http://seer.cancer.gov/statfacts/html/anus.html

TABLE 1 | Clinical characteristics of patients.

	Non-recurrent patients	Recurrent patients
n	13	4
Gender		
Male	5	1
Female	8	3
Age		
Median	53	49
Range	36 ∼ 78	45 ∼ 76
BMI	20.6 ∼ 35.7	19.8 ∼ 28.2
HIV+	2	0
Weight decrease during CRT (%)	−1.7 ∼ 12.9	−1.9 ∼ 22.4
ECOG performance status		
0	10	3
1	3	1
T stage		
1	3	0
2	6	1
3	4	1
4	0	2
N stage		
0	10	2
1	0	1
2	2	1
3	1	0
AJCC stage (TNM stage)		
2	3	0
3	7	1
4	0	2
5	3	1

anal tumors have similar intensity to the surrounding normal structures in CT scans. Thus, it would be hard for the radiation oncologists to contour the tumors accurately. FDG uptake of the tumors in PET scans could help in identify the tumor regions. However, high FDG uptake caused by non-tumor lesions, such as tissue inflammation, could affect the accuracy of delineated GTVs. In this study, we did not find such difficult cases.

Image Registration

A rigid image registration followed by a B-Spline deformable image registration was used to align the Mid-CRT CT and Post-CRT CT to the Pre-CRT CT, respectively by maximizing their normalized cross correlation. To achieve higher registration accuracy in the tumor area, registration was constrained within a cuboid region, excluding irrelevant structures. The region ranged from the top of femoral head to the inferior pubic ramus, from the right lateral aspect of the right femoral head to the left lateral aspect of the left femoral head, and from the anterior border of the pubis to the most posterior border of sacrum.

The registration results were visually evaluated, and no obvious misalignments were observed. The resulting registration transform was applied to warp the manually delineated GTVs from Mid-CRT and Post-CRT scans to Pre-CRT scan, respectively. The Mid-CRT and Post-CRT SUV images were similarly warped to the Pre-CRT SUV image using the same transform as above, respectively. The following image analysis was performed in the same frame of reference, i.e., the Pre-CRT coordinate system.

Extraction of Image Features

We used the Insight Segmentation and Registration Toolkit (ITK, National Library of Medicine; Bethesda, MD) to extract image features. Nineteen geometry features, nine intensity features, eight texture features based on co-occurrence matrix, and ten texture features based on run-length matrix were computed within the GTVs in the Pre-, Mid-, and Post-CRT CT and SUV images, respectively. The difference or change of feature values between two time points were also computed, including Diff1 = Mid-CRT - Pre-CRT, Diff2 = Post-CRT - Mid-CRT, and Diff3 = Post-CRT - Pre-CRT. The detailed definition of these features was described in **Appendix A**.

Geometry Features

Geometry features described the shape, size, or relative position of a tumor. Nineteen geometry features were computed (13, 14), including volume, major axis length, minor axis length, eccentricity, elongation, orientation, bounding box volume, oriented bounding box volume, equivalent spherical perimeter, equivalent spherical radius, ferret diameter, number of lines, number of pixel on border, perimeter, perimeter on border, perimeter on border ratio, physical size, region elongation, and roundness. For instance, Roundness (R) is defined by

$$R = A/v$$

where v is the surface area of the GTV, A is the surface area of the hyper-sphere with the same volume of the GTV.

Intensity Features

Nine intensity features were computed based on the intensity (CT number in CT images and SUV in PET images) of all voxels within the GTV, including minimum, maximum, mean, standard deviation, sum, median, skewness, kurtosis, and variance.

Texture Features

Texture features quantify the spatial patterns of tumor from images (15). In each CT or SUV image, the intensity was first normalized into 64 gray levels. The texture features were computed based on the gray level co-occurrence matrix (GLCM) (16–18) and gray level run-length matrix (GLRM) (18, 19).

An element of a GLCM measures the number of two specified gray levels separated by a given distance in a specified direction (16–18). After the construction of the GLCM, the following eight frequently used features were computed (16–18): Energy, entropy, correlation, inverse difference moment, inertia, cluster shade, cluster prominence, Haralick correlation. Each GLCM feature was computed in 13 directions (in 3D) with a distance of one voxel between the pair of voxels. The feature was then averaged over the 13 directions. The standard

deviation of each feature over the 13 directions was computed as well.

An element of a GLRM measures the number of "runs" with a specific length and specific gray level (18, 19), where a "run" is defined as a block of consecutive voxels with the same gray level in a specific direction. In this study, we set the maximum allowed length of runs to the length of the minimum bounding box of the GTV. After the construction of the GLRM, the following ten frequently used features were computed: short run emphasis (SRE), long run emphasis (LRE), gray level non-uniformity (GLN), run length non-uniformity (RLN), low gray level run emphasis (LGLRE), high gray level run emphasis (HGLRE), short run low gray level emphasis (SRLGLE), short run high gray level emphasis (SRHGLE), long run low gray level emphasis (LRLGLE), long run high gray level emphasis (LRHGLE). Each GLRM feature was also computed in 13 directions. The feature was then averaged over the 13 directions. The standard deviation of each feature over the 13 directions was computed as well.

Predictive Model Construction

For this study the binary response variable of interest is the recurrence of anal cancer, coded 1 = recurrence and 0 = non-recurrence. The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the accuracy of the prediction.

Firstly, we used a univariate logistic regression model (20) to identify individually significant image features for predicting recurrence. The AUC of each image feature was obtained and p-value was calculated with Wilcoxon rank-sum test (21). A cut-off p-value of 0.05 was used to identify significant tumor recurrence predictors, i.e., an image feature was identified as a significant predictor, if its p-value was lower than 0.05.

Secondly, we used a multivariate logistic regression model (20) to select the most significant feature set (as single feature or multiple features) for predicting recurrence. The initial null hypothesis was that there was no relationship between the image features and recurrence. When multivariate logistic regression was used, null hypotheses of adding image features to the multiple logistic regression do not improve the prediction accuracy any more than expected by chance were tested. Again, p-value of 0.05 from the null hypotheses was used to select the most significant feature set.

Lastly, we used an advanced pattern classification framework of naïve Bayesian classifier (22) for predicting recurrence using the identified feature set as input. Due to the small patient cohort, leave-one-out cross-validation was used. Let C be the outcome (recurrence) class, which is modeled as a random variable, and let X be a vector of random variables denoting the input features. Further, let c and x represent particular class of C and particular observed value of X. Our model uses Bayes' rule to compute the probability of each class given the observed values as,

$$p(C = c \,|X = x) = \frac{p(C = c)\, p(X = x \,|C = c)}{p(X = x)} \qquad (1)$$

Because in naïve Bayesian classifier the features are assumed to be conditionally independent, we have

$$p(X = x \,|C = c) = \prod_i p(X_i = x_i \,|C = c)$$

which is simple to estimate from training data as well as to compute for test data. For example, for each recurrent class and continuous image feature, we will estimate the mean and standard deviation of the feature given the class. Traditionally a single Gaussian distribution assumption was used when estimating the mean and standard deviation. Here this assumption was eliminated in favor of kernel density estimation, but still maintaining the independence assumption (22). Finally, the probability computed from (1) is used to determine the most probable class.

RESULTS

Univariate Analysis of Logistic Regression Model

Six geometry features, three intensity features, and six texture features (three co-occurrence matrix features and three run-length matrix features) were identified as individually significant predictors ($p \leq 0.05$) to differentiate recurrence and non-recurrence using the univariate logistic regression model. The identified predictors were listed in **Table 2**.

Multivariate Analysis of Logistic Regression

By applying the multivariate logistic regression model, the Diff3 of roundness of CT, i.e., the difference in tumor roundness

TABLE 2 | Selected anal cancer recurrence predictors from all the image features by univariate logistic regression model.

Features	[a]Association	AUC	p-value
Diff3 Roundness	+	1.00	0.00
Post-CRT Roundness	−	0.96	0.00
Diff2 Roundness	+	0.90	0.01
Diff3 Perimeter on Border Ratio	−	0.77	0.02
Diff3 CT Minimum	+	0.85	0.02
Post-CRT CT SD of Correlation	−	0.77	0.03
Post-CRT Major Axis Length	−	0.81	0.03
Diff3 CT Mean of Inverse Difference Moment	−	0.83	0.03
Post-CRT CT Elongation	−	0.83	0.04
Diff3 CT Mean of Short Run Emphasis	+	0.83	0.04
Post-CRT CT Minimum	+	0.62	0.04
Post-CRT CT Mean of Inverse Difference Moment	−	0.63	0.05
Post-CRT SUV SD of Cluster Shade	+	0.69	0.05
Diff3 CT Mean	+	0.79	0.05
Diff1 CT SD of Long Run High Gray Level Emphasis	−	0.83	0.05

[a]*Association = "+" indicates the larger a feature, the more likely tumor recurrent; Association = "−" indicates the larger a feature, the less likely tumor recurrent.*

between Post-CRT and Pre-CRT, was identified as the most powerful predictor with an AUC of 1.0. Using Pre-CRT and Mid-CRT scans only, Orientation of Pre-CRT SUV and LRHGLE of Pre-CRT CT were selected as the most important feature set with a high AUC of 1.00. The results were shown in **Table 3**.

Naïve Bayesian Classifier Based Analysis

By applying the naïve Bayesian classifier, Diff3 of Number of Pixels on Border of SUV and Elongation of Post-CRT CT were identified as the most useful feature set with AUC = 1.00. Using Pre-CRT and Mid-CRT scans only, standard deviation of Mid-CRT SUV, Diff1 of Mean of LRHGLE of SUV were selected as the most important feature set with an AUC of 0.86. The results were shown in **Table 4**.

DISCUSSION

Almost all the predictors identified by univariate logistic regression model, multivariate logistic regression model, and naïve Bayesian classifier were derived from Post-CRT scans or from Diff3 (Post-CRT - Pre-CRT). To investigate the early prediction ability of the features, we applied the multivariate logistic regression model and naïve Bayesian classifier by using features from Pre-CRT and Mid-CRT scans and Diff1 only. As shown in **Table 3**, Orientation of Pre-CRT SUV, Mean of Pre-CRT CT, and Mean of LRHGLE of Pre-CRT CT were identified as the most useful feature set by the multivariate logistic regression model with an AUC of 1.0. As shown in **Table 4**, standard deviation of Mid-CRT SUV and Diff1 of Mean of LRHGLE of SUV were identified as the most useful feature set by the naïve Bayesian classifier with an AUC of 0.86. These results demonstrated the potential of Pre-CRT and Mid-CRT PET/CT scans for the early predication of anal cancer recurrence.

TABLE 3 | Selected anal cancer recurrence predictors (correlation to the recurrence in parentheses) by multivariate logistic regression model.

	Using Pre-CRT, Mid-CRT, and Post-CRT	Using Pre-CRT and Mid-CRT
Features	Diff3 CT Roundness (0.83)	Pre-CRT SUV Orientation (−0.31), Pre-CRT CT Mean (−0.15), Pre-CRT CT Mean of Long Run High Gray Level Emphasis (0.41)
AUC	1.00	1.00

TABLE 4 | Selected anal cancer recurrence predictors (correlation to the recurrence in parentheses) by naïve Bayesian classifier.

	Using Pre-CRT, Mid-CRT, and Post-CRT	Using Pre-CRT and Mid-CRT
Features	Diff3 SUV Number of Pixels on Border (−0.07) and Post-CRT CT Elongation (−0.28)	Mid-CRT SUV Standard Deviation (−0.15), Diff1 SUV Mean of Long Run High Gray Level Emphasis (0.16)
AUC	1.00	0.86

One geometry feature Diff3 of roundness, has been identified as the most useful predictor by both univariate and multivariate logistic regression models. Roundness measures how similar the shape of a tumor is to a sphere with range [0, 1]. A larger value of roundness means higher similarity to a sphere. As shown in **Figure 1** and **Table 5** the roundness of all four recurrent tumors increased from Pre-CRT to Post-CRT by 0.05 or more, with a mean increase of 0.08, whereas the roundness of the 13 non-recurrent tumors either decreased (11 tumors), or did not change (2 tumors), or increased slightly by 0.01 (1 tumor), with a mean decrease of 0.06. Therefore, by using Diff3 of roundness only, we were able to correctly predict all cases with an AUC of 1.0.

As shown in **Table 5**, the roundness of many of the non-recurrent tumors decreased from Pre-CRT to Post-CRT and the roundness of recurrent tumors increased from Pre-CRT to Post-CRT. The difference in roundness may reflect that the normal anal canal has a low roundness value, which would be consistent with decreased anal tumor burden; to the contrary, higher gross tumor burden would have a higher roundness value. **Figure 2** shows an example of the comparison between Pre-CRT and Post-CRT of a non-recurrent tumor. Its roundness decreased from 0.70 to 0.58. The tumor regressed significantly in coronal direction. However, it enlarged in axial direction on CT. We further investigated all the tumors in our dataset. For non-recurrent patients, the changes of tumor size were mainly in superior-inferior direction. However, the changes were not consistent, i.e., some tumors decreased, and other tumors increased in superior-inferior direction. On the other hand, the changes in axial plane were generally quite small. For recurrent patients, changes could be observed in both axial plane and superior-inferior direction. However, we did not find a consistent pattern in the trend of size changes either. It is important to note that the radiation oncologist who delineated the tumor volumes was blinded to the prior contours while contouring the follow up scans, which could have affected the consistency of the volumes over time including perceived enlargement of delineated tumor regions in follow up scans of non-recurrent patients. Therefore, the recurrence prediction purely based on the geometry measurements, such as roundness, volume, etc., may not be reliable enough. Intensity and texture-based imaging features would be useful complementary to the geometry measurements in the recurrence prediction. In addition, this experimental result could raise a hypothesis—tumors may regress in a non-uniform manner after CRT. The tumors with positive response to the treatment may have regressed asymmetrically, which also may have contributed to their roundness measurement decreasing. We were unable to find other published literature about directional tumor regression after CRT and therefore warrant further evaluation.

In addition to roundness, some other features were identified as recurrence predictors by multivariate logistic regression model (**Table 3**) and by naïve Bayesian classifier (**Table 4**). Each of these features had low correlation (<0.50) to tumor recurrence and was weak classifier by itself. However, the performance can be improved significantly by systematically combining a number of weak classifiers (23) and using well-designed training procedure, such as the multivariable logistic regression model

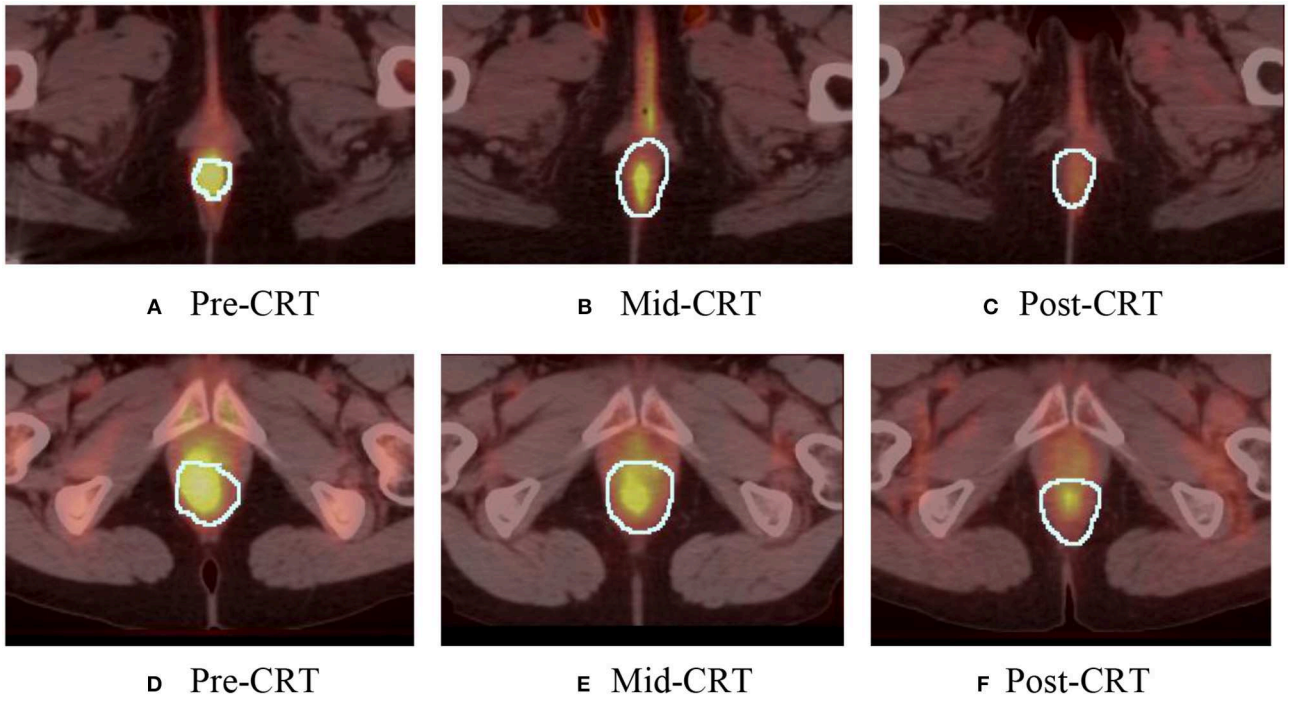


FIGURE 1 | Manually delineated tumor contour (white) in Pre-CRT, Mid-CRT, and Post-CRT scans of a non-recurrent patient **(A–C)** and a recurrent patient **(D–F)** Post-CRT. Roundness showed different changing patterns (in Diff3 = Post-CRT – Pre-CRT) between recurrent and non-recurrent groups.

TABLE 5 | The roundness of each patient at Pre-CRT, Mid-CRT, and Post-CRT and their differences.

Recurrent Status (1 = recurrent, 0 = non-recurrent)	Pre	Mid	Post	Diff1	Diff2	Diff3
0	0.78	0.70	0.67	−0.08	−0.04	−0.12
0	0.68	0.64	0.69	−0.04	0.05	0.01
0	0.80	0.61	0.75	−0.19	0.14	−0.05
0	0.72	0.73	0.70	0.01	−0.04	−0.03
0	0.80	0.64	0.70	−0.15	0.05	−0.10
0	0.72	0.70	0.62	−0.01	−0.08	−0.09
0	0.70	0.78	0.58	0.08	−0.20	−0.12
0	0.84	0.67	0.76	−0.17	0.09	−0.09
0	0.78	0.73	0.78	−0.04	0.04	0.00
0	0.73	0.80	0.70	0.06	−0.10	−0.03
0	0.88	0.76	0.77	−0.12	0.01	−0.11
0	0.75	0.72	0.74	−0.02	0.02	0.00
0	0.76	NA	0.72	NA	NA	−0.04
1	0.68	0.66	0.76	−0.02	0.10	0.08
1	0.78	0.69	0.83	−0.09	0.14	0.05
1	0.67	0.71	0.78	0.04	0.07	0.11
1	0.72	0.73	0.80	0.01	0.07	0.08
Mean of non-recurrent	0.76	0.71	0.71	−0.06	0.00	−0.06
Mean of recurrent	0.71	0.70	0.79	−0.01	0.10	0.08

Diff1 = Mid-CRT – Pre-CRT; Diff2 = Post-CRT – Mid-CRT; Diff3 = Post-CRT – Pre-CRT.

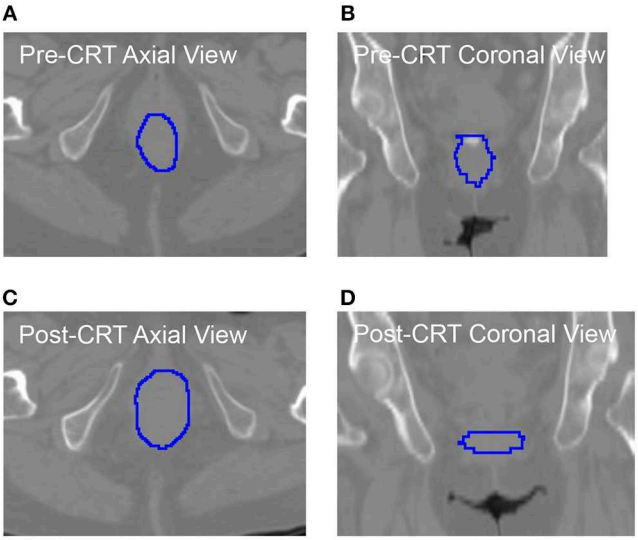


FIGURE 2 | The CT images with manually delieated tumor contour of anal cancer patient with no tumor recurrence. The tumor regressed in coronal directions **(B)** vs. **(D)**. However, it progressed in axial direction **(A)** vs. **(C)**. The roundness of this tumor was changed from 0.70 (Pre-CRT) to 0.58 (Post-CRT).

and the naïve Bayesian classifier in this study. By use of these combined features, relatively high performance in the tumor recurrence prediction was achieved (high AUC values) as shown in **Tables 3**, **4**.

Traditional naïve Bayesian classifier assumes that numeric features are generated by a single Gaussian distribution. It is a reasonable approximation to many real-world applications, but not always the best. Our approach extended the traditional naïve Bayesian by eliminating the single Gaussian assumption, which certainly could be violated in cancer recurrence prediction. Another advantage of the method is that it does not suffer from the high dimensionality of the model. This is because features are assumed to be conditionally independent in Naïve Bayesian classifier, so that the curse-of-dimensionality can be avoided by allowing the join distribution to be decomposed. Naïve Bayesian classifier is also closely related to the widely used logistic regression classifier. While naïve Bayesian classifier fits a probability that optimizes the joint likelihood, logistic regression fits the same probability model that optimizes the conditional probability. It has been shown that in some practical cases naïve Bayesian can outperform logistic regression because it converges faster (24).

One limitation of this study was that this was a retrospective analysis of a small patient cohort ($n = 17$), particularly only 4 patients with recurrence. This was a small, unbalanced dataset for reliable prediction. Although we used cross-validations to avoid potential over fitting, the predictive accuracy and stability of the model should be validated in a larger and independent patient cohort. Another limitation is the lack of standard dose prescription for each patient. The total dose a patient received was based subjectively on tumor response as per Mid-CRT PET. Finally, it was hard to provide biological or clinic explanations for why the extracted image features were important for recurrence prediction.

CONCLUSIONS

Early prediction of tumor persistence or recurrence using PET/CT scans obtained prior to or during CRT for anal cancer may be possible through analysis of quantitative imaging features. Additional study is warranted in a larger patient population to confirm our findings. A future study to investigate the correlation between clinical characteristics (e.g., T staging, N staging, radiation dose, etc.) and the image features is needed as well.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Department of Radiation Oncology, University of Maryland Baltimore and Department of Radiation Oncology, Moffitt Cancer Center. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

WL, MC, and JW contributed conception and design of the study. MC, KL, SH, and RS collected experimental data, evaluated the tumor recurrence status, and created clinical indices of the dataset. MC delineated gross tumor volume for each patient. JW, WC, and ST contributed to the computation of image features. HZ and WC performed statistical analysis. JW, HZ, MC, WC, and WL wrote the draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported in part through the NIH/NCI Grant R01CA172638 and the NIH/NCI Cancer Center Support Grant P30 CA008748.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00934/full#supplementary-material

REFERENCES

1. Dewdney A, Rao S. Metastatic squamous cell carcinoma of the anus: time for a shift in the treatment paradigm? *ISRN oncol.* (2012) 2012:756591. doi: 10.5402/2012/756591
2. Das P, Bhatia S, Eng C, Ajani JA, Skibber JM, Rodriguez-Bigas MA, et al. Predictors and patterns of recurrence after definitive chemoradiation for anal cancer. *Int J Radiat Oncol Biol Phys.* (2007) 68:794–800. doi: 10.1016/j.ijrobp.2006.12.052
3. Nguyen BT, Joon DL, Khoo V, Quong G, Chao M, Wada M, et al. Assessing the impact of FDG-PET in the management of anal cancer. *Radiother Oncol.* (2008) 87:376–82. doi: 10.1016/j.radonc.2008.04.003
4. Delhorme JB, Severac F, Waissi W, Romain B, Antoni D, Freel KC, et al. Surgery is an effective option after failure of chemoradiation in cancers of the anal canal and anal margin. *Oncology.* (2017) 93:183–90. doi: 10.1159/000475758
5. Grigsby PW.Grigsby PW. FDG-PET/CT: new horizons in anal cancer. *Gastroenterol Clin Biol.* (2009) 33:456–8. doi: 10.1016/j.gcb.2009.01.012
6. Saboo SS, Zukotynski K, Shinagare AB, Krajewski KM, Ramaiya N. Anal carcinoma: FDG PET/CT in staging, response evaluation, and follow-up. *Abdom Imaging.* (2013) 38:728–35. doi: 10.1007/s00261-012-9958-3
7. Vercellino L, Montravers F, de Parades V, Huchet V, Kerrou K, Bauer P, et al. Impact of FDG PET/CT in the staging and the follow-up of anal carcinoma. *Int J Colorectal Dis.* (2011) 26:201–10. doi: 10.1007/s00384-010-1080-9
8. Lieng H, Kneebone A, Hayden AJ, Christie DRH, Davis BJ, Eade TN, et al. Prognostic utility of pre-and post-treatment FDG-PET parameters in anal squamous cell carcinoma. *Radiother Oncol.* (2019) 136:21–8. doi: 10.1016/j.radonc.2019.03.014
9. Hong JC, Cui Y, Patel BN, Rushing CN, Faught AM, Eng JS, et al. Association of interim FDG-PET imaging during chemoradiation for squamous anal canal carcinoma with recurrence. *Int J Radiat Oncol Biol Phys.* (2018) 102:1046–51. doi: 10.1016/j.ijrobp.2018.04.062
10. Xu B, Ma J, Jiang G, Wang Y, Ma Q. Diagnostic value of positron emission tomography (PET) and PET/computed tomography in recurrent/metastatic ovarian cancer: a meta-analysis. *J Obstet Gynaecol Res.* (2017) 43:378–86. doi: 10.1111/jog.13222

11. Li J, Yan R, Lei J, Jiang C. Comparison of PET with PET/CT in detecting peritoneal carcinomatosis: a meta-analysis. *Abdom Imaging*. (2015) 40:2660–6. doi: 10.1007/s00261-015-0418-8
12. Chuong MD, Hoffe SE, Weber J, Freilich JM, Lomas H, Almhanna K, et al. Outcomes of anal cancer treated with definitive IMRT-based chemoradiation. *J Radiat Oncol*. (2012) 1:165–72. doi: 10.1007/s13566-012-0029-8
13. Lehmann G. Label object representation and manipulation with ITK. *Insight J*. (2007) 8:1–31. Available online at: http://hdl.handle.net/1926/584
14. Padfield D, Miller J. A label geometry image filter for multiple object measurement. *Insight J*. (2008) 301:1–13. Available online at: http://hdl.handle.net/1926/1493
15. Tan S, Kligerman S, Chen W, Lu M, Kim G, Feigenberg S, et al. Spatial-temporal [18F] FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy. *Int J Radiat Oncol Biol Phys*. (2013) 85:1375–82. doi: 10.1016/j.ijrobp.2012.10.017
16. Haralick RM. Statistical and structural approaches to texture. *Proc IEEE*. (1979) 67:786–804. doi: 10.1109/PROC.1979.11328
17. Mayerhoefer ME, Riedl CC, Kumar A, Gibbs P, Weber M, Tal I, et al. Radiomic features of glucose metabolism enable prediction of outcome in mantle cell lymphoma. *Eur J Nucl Med Mol Imaging*. (2019). doi: 10.1007/s00259-019-04420-6. [Epub ahead of print].
18. Chen SW, Shen WC, Hsieh TC, Liang JA, Hung YC, Yeh LS, et al. Textural features of cervical cancers on FDG-PET/CT associate with survival and local relapse in patients treated with definitive chemoradiotherapy. *Sci Rep*. (2018) 8:11859. doi: 10.1038/s41598-018-30336-6
19. Galloway MM. *Texture Analysis Using Grey Level Run Lengths. NASA STI/Recon*. Technical Report N, 1974. 75.
20. Sperandei S. Understanding logistic regression analysis. *Biochem Med*. (2014) 24:12–8. doi: 10.11613/BM.2014.003
21. Hogg RV, Tanis EA, Zimmerman DL. *Probability and Statistical Inference*. Vol. 993. New York, NY: Macmillan (1977).
22. John GH, Langley P. Estimating continuous distributions in Bayesian classifiers. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. Montreal, QC: Morgan Kaufmann Publishers Inc (1995).
23. Kleinberg E. Stochastic discrimination. *Ann Math Artif Intell*. (1990) 1:207–39. doi: 10.1007/BF01531079
24. Ng AY, Jordan MI. On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes. In: *Advances in Neural Information Processing Systems*. Vancouver, BC (2002).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 09 October 2019
doi: 10.3389/fonc.2019.01007

Multiplanar MRI-Based Predictive Model for Preoperative Assessment of Lymph Node Metastasis in Endometrial Cancer

Xiaojuan Xu[1†], Hailin Li[2,3,4†], Siwen Wang[2,3], Mengjie Fang[2,3], Lianzhen Zhong[2,3], Wenwen Fan[1], Di Dong[2,3], Jie Tian[2,5*] and Xinming Zhao[1*]*

[1] Department of Diagnostic Imaging, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2] CAS Key Lab of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China, [3] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China, [4] School of Automation, Harbin University of Science and Technology, Harbin, China, [5] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing, China

OPEN ACCESS

Edited by:
Chunxiao Guo,
University of Texas MD Anderson Cancer Center, United States

Reviewed by:
Zaiyi Liu,
Guangdong Provincial People's Hospital, China
Anna Colleen Crouch,
University of Texas MD Anderson Cancer Center, United States

***Correspondence:**
Di Dong
di.dong@ia.ac.cn
Jie Tian
jie.tian@ia.ac.cn
Xinming Zhao
xinmingzh@sina.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 28 July 2019
Accepted: 18 September 2019
Published: 09 October 2019

Citation:
Xu X, Li H, Wang S, Fang M, Zhong L, Fan W, Dong D, Tian J and Zhao X (2019) Multiplanar MRI-Based Predictive Model for Preoperative Assessment of Lymph Node Metastasis in Endometrial Cancer. Front. Oncol. 9:1007. doi: 10.3389/fonc.2019.01007

Introduction: Assessment of lymph node metastasis (LNM) is crucial for treatment decision and prognosis prediction for endometrial cancer (EC). However, the sensitivity of the routinely used magnetic resonance imaging (MRI) is low in assessing normal-sized LNM (diameter, 0–0.8 cm). We aimed to develop a predictive model based on magnetic resonance (MR) images and clinical parameters to predict LNM in normal-sized lymph nodes (LNs).

Materials and Methods: A total of 200 retrospective patients were enrolled and divided into a training cohort (n = 140) and a test cohort (n = 60). All patients underwent preoperative MRI and had pathological result of LNM status. In total, 4,179 radiomic features were extracted. Four models including a clinical model, a radiomic model, and two combined models were built. Area under the receiver operating characteristic (ROC) curves (AUC) and calibration curves were used to assess these models. Subgroup analysis was performed according to LN size. All patients underwent surgical staging and had pathological results.

Results: All of the four models showed predictive ability in LNM. One of the combined models, Model^{CR1}, consisting of radiomic features, LN size, and cancer antigen 125, showed the best discrimination ability on the training cohort [AUC, 0.892; 95% confidence interval [CI], 0.834–0.951] and test cohort (AUC, 0.883; 95% CI, 0.786–0.980). The subgroup analysis showed that this model also indicated good predictive ability in normal-sized LNs (0.3–0.8 cm group, accuracy = 0.846; <0.3 cm group, accuracy = 0.849). Furthermore, compared with the routinely preoperative MR report, the sensitivity and accuracy of this model had a great improvement.

Conclusions: A predictive model was proposed based on MR radiomic features and clinical parameters for LNM in EC. The model had a good discrimination ability, especially for normal-sized LNs.

Keywords: endometrial cancer, lymph node, metastasis, magnetic resonance imaging, radiomics

INTRODUCTION

Endometrial cancer (EC) is the most common gynecological malignancy in industrialized countries (1, 2). In China, EC is the second most common malignancy of the female genital tract with patients steadily increasing, especially in high urbanization areas (3). Lymph node metastasis (LNM) is an important risk factor for EC prognosis. Systematic lymphadenectomy is routinely performed according to International Federation of Gynecology and Obstetrics (FIGO). However, there is long-term controversy regarding whether it is necessary for low-risk or stage IA disease (4), as the incidence of LNM is very low in these patients (5). In addition, indiscriminative lymphadenectomy may lead to overtreatment and increase in post-operative complications, including chronic lymphedema, lymphocysts, infection, and nerve/vascular injuries (6).

Several histopathological findings, such as histological subtype, depth of myometrial invasion (DMI), primary tumor diameter (PTD), lymphovascular space invasion, and tumor grade, are known to be risk factors for LNM (4, 5), and researchers proposed various risk-classification models (4, 7). However, most of them are only available post-operatively. Sentinel lymph nodes mapping was proposed to evaluate LNM intraoperatively (8), but the technological dependence on experienced surgeons and relatively high false-negative rates limited its clinical application. Accurate preoperative and non-invasive evaluation of LNM is crucial, which can provide valuable information for prognosis prediction and treatment decision, especially in determining the extent of lymphadenectomy.

Magnetic resonance imaging (MRI) is a routinely used imaging modality for preoperative evaluation of EC. It plays an important role in assessing DMI (9), but its value for LNM assessment remains unsatisfactory, with reported sensitivities of 25–50% (10, 11). Radiomics, as a novel data mining technique, could extract high-dimensional quantitative features from medical images and select reliable features for the establishment of prediction models that could be used in computer-assisted decision support. Some recent researches showed that radiomics had the potential to evaluate therapeutic effects, predict the recurrence and metastasis, predict survival time (12–14), and aid the differential diagnosis of cancers (15). Currently, radiomic investigations in preoperative prediction of LNM showed encouraging achievement (16–18). However, to our knowledge, there is no literature that has determined whether a radiomics-based study would render superior prediction of metastasis in different size groups of LNs, and there has been no study on EC.

The purpose of this study was to investigate the efficacy of multiplanar enhanced MRI-based radiomics for preoperative prediction of metastasis in normal-sized (diameter 0–0.8 cm on MRI) LNs in EC patients.

Abbreviations: LNM, lymph node metastasis; EC, endometrial cancer; LN, lymph node; AUC, area under the curve; MRI, magnetic resonance imaging; IGO, International Federation of Gynecology and Obstetrics; DMI, depth of myometrial invasion; A125, cancer antigen 125; PTD, primary tumor diameter; WHO, World Health Organization; DCE, dynamic contrast enhanced; VOI, volume of interest; ICC, intraclass correlation coefficient; mRMR, minimum redundancy/maximum relevance; RFE, recursive feature elimination; HGLE, high gray-level emphasis; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; PV, positive predictive value; TP, true positive; FN, false negative; FP, false positive; pN+, pathologically LN positive; 3D-iso-LAVA-XV, three-dimensional liver acquisition with volume acceleration DCE with isotropy scanning; CI, confidence interval; ADC, apparent diffusion coefficient; DWI, diffusion-weighted MR imaging; HE4, human epididymis secretory protein 4; GLCM, gray-level cooccurrence matrix.

MATERIALS AND METHODS

Study Design and Participants

This retrospective study with anonymous data was approved by the Ethics Committee of our hospital, and the informed consent requirement was waived.

Two hundred consecutive patients with EC who had been treated between January 2011 and December 2017 were enrolled. **Figure 1** shows the patient recruitment pathway. Patients were divided into two independent cohorts: 140 patients treated between January 2011 and March 2016 in the training cohort, and 60 patients treated between April 2016 and December 2017 in the test cohort.

As shown in **Table 1**, clinical parameters including age, blood serum cancer antigen 125 (CA125) level, preoperative histological type, and differentiation were derived from medical records.

All MR imaging data were reviewed together by two board-certified radiologists (reader 1 and reader 2) specialized in gynecological tumor imaging with 6 and 20 years of experience. The PTD, DMI, involvement of the cervix, cornua, adnexa, parametrium, and LN status including the size and positive or negative were recorded. Maximal short-axis diameter of LN was measured on delayed phase of dynamic contrast enhanced (DCE) sequence at axial–sectional images (see details below). Patients with pelvic LN $>$ 8 mm or abdominal LN $>$ 10 mm, or with non-homogeneous enhancement and central necrosis on DCE images were regarded as MR report LN-positive (19). The consistency between the two radiologists was assessed by calculating the Cohen's kappa coefficients. Any disagreement was resolved by consultation. Note that LN status was defined by case.

MR Image Acquisition, Region of Interest Segmentation, and Radiomic Feature Extraction

Before receiving standard FIGO surgical staging, all patients underwent pelvic DCE MRI on two 3.0-T MR scanners (Signa HDxt and Discovery MR750, GE Medical Systems) with 8-channel phased array body coils. Two non-enhanced and one enhanced sequence were obtained and collected for analysis. Detailed scanning parameters are listed in **Table 2**.

Tumor volume of interest (VOI), covering the whole tumor volume on each MR image, were manually segmented by reader 1 using ITK-SNAP software (www.itksnap.org, version 3.6.0). Radiomic feature extraction was performed with algorithms implemented in Python (www.python.org, version 3.6.5) (20). Three-dimensional radiomic features were extracted from the

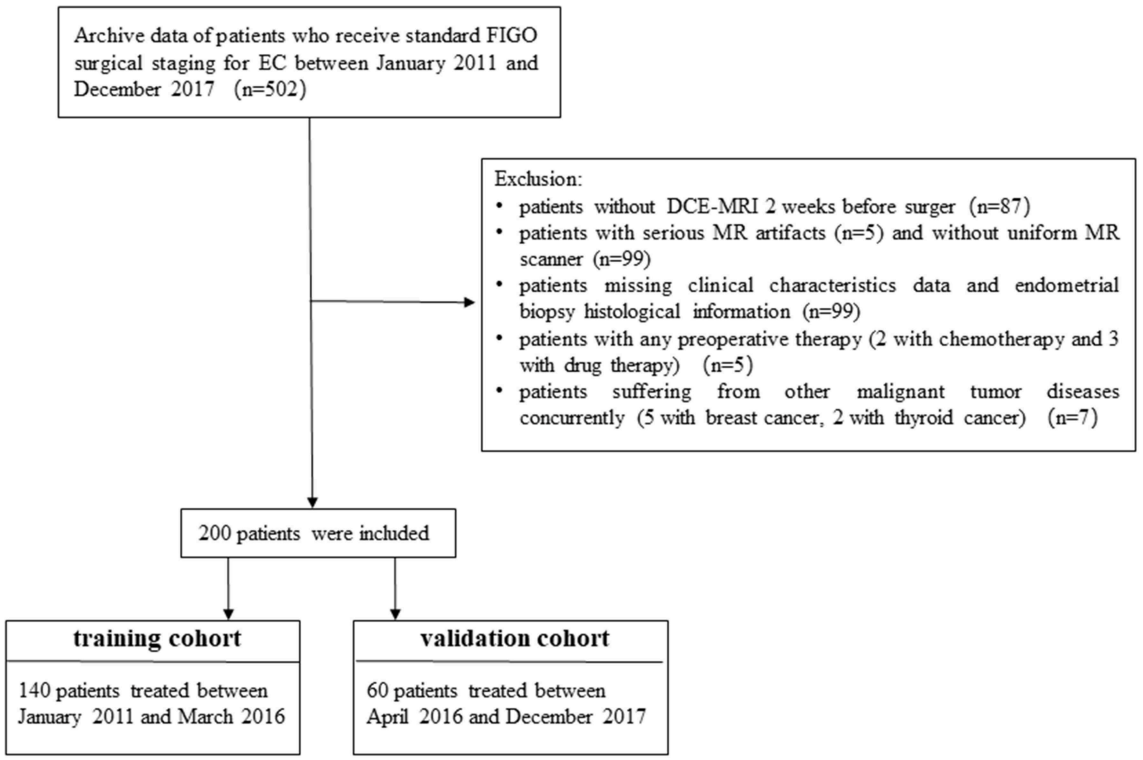


FIGURE 1 | Recruitment pathway for patients in this study.

corresponding VOIs, including first-order statistics, shape-based, and texture features. More information about the radiomic feature extraction methodology can be found in **Supplementary Method 1**.

Surgery and Histopathologic Work-Up

All patients underwent FIGO surgical staging, and accepted template systematic lymphadenectomy. All lymph node specimens were processed and evaluated according to a standard protocol. Histologic analysis of each template lymph node dissection specimen included the following parameters: total number of histologically detected lymph nodes and number of positive nodes in each region as follows: external iliac, internal iliac and obturator, and common iliac. Note that the histopathologic LN status was still considered by case level in our analysis. The 2014 World Health Organization (WHO) classification (21) and the 2009 revised FIGO staging criteria for EC (22) were used for histological diagnosis, grading, and pathological staging.

DATA ANALYSIS

Feature Selection and Model Construction

Stability analysis of radiomic features between inter-/intra-reader segmentations was firstly carried out. Thirty patients were randomly chosen, and all of their images were segmented separately by the two radiologists, thereinto, reader 1 then re-segmented these images 1 week later. The intraclass correlation coefficients (ICCs) are usually adopted to assess the stability of radiomic features extracted from VOIs delineated by different readers or segmented by the same reader at different times. The radiomic features with ICC >0.75 were retained since they had good agreement between different segmentations.

Then, stability analysis between different versions of MR scanners on radiomic features was carried out. With all the patients randomly assigned to two MR scanners, Mann–Whitney *U* test was used to find out whether a radiomic feature showed statistical difference between different versions of MR scanners in the training cohort. We removed the radiomic features that had significant differences in the two versions of MR scanners, which would improve the generalization capability of our classifier.

Figure 2 shows the workflow of model development and decision-making process for model selection. Four models were constructed, including a clinical model with only clinical parameters ($Model^{C}$), a radiomic model with only radiomic features ($Model^{R}$), and two combined model ($Model^{CR1}$ and $Model^{CR2}$). After model evaluation, the final model was selected to be visualized as a clinical useful preoperative nomogram. The detailed construction processes of the four models were as follows.

$Model^{C}$

The original feature set of the $Model^{C}$ consisted of all of the 10 clinical parameters, including age, CA125, tumor pathologic type

TABLE 1 | Baseline characteristics of the training and test cohorts.

	Training cohort ($n = 140$)			Test cohort ($n = 60$)			
	pN(+)	pN(–)	*P*	pN(+)	pN(–)	*P*	*P**
	$n = 52$	$n = 88$		$n = 15$	$n = 45$		
Age, years			0.840			0.017	0.077
Mean ± SD	55.271 ± 7.936	55.723 ± 8.382		57.403 ± 6.926	51.730 ± 9.111		
Median (range)	56.000 (28.000–68.000)	56.000 (26.000–80.000)		59.000 (45.000–67.000)	53.000 (35.000–76.000)		
CA125 level (ng/ml), Mean ± SD	86.740 ± 133.348	24.962 ± 23.559	0.002	84.491 ± 100.066	26.772 ± 32.407	0.044	0.539
MR-reported DMI			0.001			0.033	0.524
Less than 50%	19	66		5	35		
More than 50%	33	22		10	10		
MR-reported PTD (mm), Mean ± SD	3.802 ± 2.435	3.929 ± 1.994	0.735	3.758 ± 2.341	3.072 ± 1.535	0.300	0.031
MR-reported tumor staging			<0.001			<0.001	0.659
I	16	68		3	38		
II	3	10		2	2		
III	32	10		10	5		
IV	1	0		0	0		
MR-reported LN status			<0.001			0.367	0.104
cN(+)	17	6		2	2		
cN(–)	35	82		13	43		

pN(+), pathologically LN positive; pN(–), pathologically LN negative; SD, standard deviation; CA125, cancer antigen 125; DMI, depth of myometrial invasion; PTD, primary tumor diameter; LN, lymph node; cN(+), clinically LN positive; cN(–), clinically LN negative.
CA125 level was acquired within 1 week before surgery with a threshold value between 0 and 35 U/ml.
The P was derived from the univariate association analyses between each clinical parameter and different cohort.*

TABLE 2 | Detailed acquired parameters in two MR scanners.

	GE signa excite HD 3.0T			GE discovery HD750 3.0T		
	Axial T2-fs-FSE[#]	Sagittal T2-FSE	Axial 3D-iso-LAVA-XV*	Axial T2-fs-FSE[#]	Sagittal T2-FSE	Axial 3D-iso-LAVA-XV*
TR/TE	5900/121	3300/130	4.1/1.8	5541/85	4633/120	7.9/4.1
FOV (cm)	40.0	22.0	35.0	40.0	22.0	35.0
Matrix	Freq 320/Phase 256	Freq 320/Phase 256	Freq 350/Phase 350	Freq 320/Phase 256	Freq 320/Phase 256	Freq 350/Phase 350
Slice thickness (mm)	5.0	4.0	1.0	5.0	4.0	1.0
Slice gap	1.0	1.0	0	1.0	0.4	0

[#]*T2-weighted fat-suppressed fast spin echo (T2-fs-FSE).*
**Three-dimensional liver acquisition with volume acceleration DCE with isotropy scanning (3D-iso-LAVA-XV).*
TR, repetition time; TE, echo time; FOV, field of view.
Enhanced scan was done by injecting gadopentetate dimeglumine (Omniscan, GE Healthcare) into the upper limb vein by using a high-pressure syringe, with a flow rate at 2.0 ml/s and a total dose of 0.2 mmol/kg body weight. A total of 15 phases were obtained post-drug injection with a time interval of 15 s in the sagittal plane, followed by a delayed phase with isotropy axial scanning.

and differentiation by biopsy, tumor long-axis diameter, DMI, the ratio of tumor infiltration depth to myometrium depth, LN size, and adnexa or other organ involvement, which were all observed on MR images. After feature selection via Mann–Whitney *U* test and the least absolute shrinkage and selection operator (LASSO) method, six features were retained to fit the Model^{C}. The logistic regression model was constructed to examine the ability of the clinical parameters in classifying LNM.

Model^{R}

The stable original feature set of the Model^{R} consisted of 3,040 radiomic features that were dimensionally reduced by stability analysis. Univariate analysis (Mann–Whitney *U* test and chi-square test) was performed to evaluate the difference in LNM status, and the distribution of the *p*-values for the radiomic features is shown in **Supplementary Figure 1**. The retained significant features were then ranked by minimum

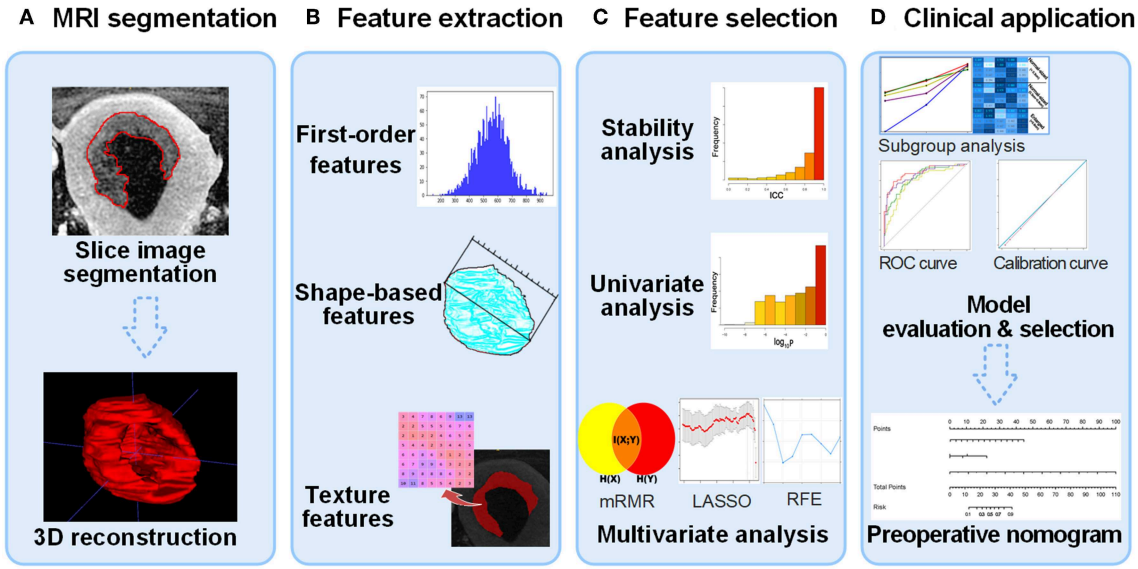


FIGURE 2 | Flow diagram of radiomic model construction. **(A)** MR images segmentation. The tumor region in each MRI slice was manually segmented, and then the whole tumor volume was reconstructed in order to extract 3D radiomic feature. **(B)** Radiomic feature extraction. Three types of radiomic features were extracted from tumor volume. **(C)** Feature selection process including stability, univariate analysis, and multivariate analysis. The construction of Model^{C} starts from univariate analysis. **(D)** Clinical application. After evaluating the four models, an optimal model was selected to plot nomogram for clinical computer-assisted decision support. MRI, magnetic resonance image; 3D, three-dimensional.

redundancy/maximum relevance (mRMR), and the most redundant features were removed. The LASSO method was used to prevent overfitting. Two radiomic features were finally selected to fit the Model^{R}. Similarly, logistic regression model was constructed to examine the ability of the radiomic features in classifying LNM.

Model^{CR1}

All of the clinical parameters and retained 3,040 radiomic features formed the stable feature set. Univariate analysis was performed on this feature set. After removing features with p-values > 0.05, we computed the mRMR ranking for the feature set based on the concordance index (23, 24), and the top 5% features were retained. Then, the recursive feature elimination (RFE) method was performed to further select the LNM-related features. The RFE algorithm repeatedly constructed the model and removed the features, depending on the root mean square error of the model by a cross-validation in the training cohort.

Model^{CR2}

The stable original feature set of the Model^{CR2} consisted of stable radiomic features and clinical parameters except LN size. Univariate analysis was first performed on this feature set. Then, in the multivariable analysis, mRMR and the LASSO method were performed successively. Logistic regression model was constructed to examine the classification ability of the combination of radiomic features and clinical primary lesion information.

Assessment and Validation of Model Performance

The receiver operating characteristic (ROC) curves were plotted to assess the performance of the four models in both cohorts. Area under ROC curve (AUC) was calculated for quantitative comparison. The model with the highest AUC was selected as the final model. Delong test was used to compare AUCs between the training cohort and test cohort, and a p-value > 0.05 indicated that there was no significant difference in AUCs, which ensured that the model had an enough low risk of over-fitting. Calibration curve was plotted to evaluate the agreement between prediction result and gold standard.

In previous research, good effects were gained in predicting the metastasis of an EC-LN larger than 1 cm (25). However, there is no study that ever focused on metastasis prediction on different sized LNs. So, we carried out a subgroup analysis on LN size. Patients were divided into three subgroups according to the LN size measured on MRI, including enlarged LNs with diameter larger than 0.8 cm (>0.8 cm), normal-sized LNs with diameter between 0.3 and 0.8 cm (0.3–0.8 cm), and normal-sized LNs with diameter smaller than 0.3 cm (<0.3 cm). F-score ($F_1 = \frac{2Recall \times Precision}{Recall + Precision}$) was calculated in these subgroups, assuming that recall (equivalently, sensitivity, $\frac{TP}{TP+FN}$) and precision (equivalently, PPV, positive predictive value, $\frac{TP}{TP+FP}$) are of equal importance, where TP, FN, and FP represent true positive, false negative, and false positive, respectively. The higher F-score synthetically reflects higher sensitivity and higher PPV.

TABLE 3 | Pathological characteristics of the patients in our study.

	Training cohort ($n = 140$)	Test cohort ($n = 60$)	*P*
Surgically histological type, *n* (%)			0.602
Endometrioid	112 (80%)	47 (78.33%)	
Non-endometrioid	28 (20%)	13 (21.67%)	
Histological grade, *n* (%)			0.041
Well-differentiated	67 (47.86%)	17 (28.33%)	
Moderately differentiated	52 (37.14%)	37 (61.67%)	
Poorly differentiated	21 (15.00%)	6 (10.00%)	
Pathological N stage, *n* (%)			0.133
pN–	88 (62.86%)	45 (75.00%)	
pN+	52 (37.14%)	15 (25.00%)	
Pathological staging, *n* (%)			0.250
pI	68 (48.57%)	38 (63.33%)	
pII	16 (11.43%)	6 (10.00%)	
pIII	50 (35.71%)	15 (25.00%)	
pIV	6 (4.29%)	1 (1.67%)	

Clinical Utility of the Final Model

In order to determine the clinical significance of the final model, decision curves were plotted by quantifying the net benefits in the training and test cohort. For the convenience of clinical application, a visualized preoperative nomogram was developed based on the formula exported by the logistic regression of the final model.

Statistical Analysis

In this study, statistical analysis programs were completed by R software (version 3.5.0; https://www.r-project.org). All statistical hypothesis tests were two-sided, and p-values < 0.05 were considered significant.

RESULTS

Patient Characteristics

The clinical and pathological characteristics in the two cohorts are shown in **Tables 1**, **3**, respectively. Pathologically LN positive (pN+) patients formed 37.14% (52/140) and 25.00% (15/60) of the training and test cohorts, respectively, and there was no significant difference between them (p-value $= 0.133$, χ^2 test). The clinical parameters age and CA125 had no differences between the two cohorts (p-value $= 0.077$ and 0.539 respectively, Mann–Whitney U-test). In total, sensitivity and specificity were 28.36% (19/67) and 93.98% (125/133) according to the MR report LN status after consensus within two radiologists in our study. Also, the judgments by two radiologists on MRI were basically stable (sensitivity was 0.642 and 0.552, and the specificity was 0.917 and 0.940, respectively). The inconsistency of judgment was resolved by consultation. The Cohen's kappa coefficients to test consistency of the main MR indicators evaluated by the two radiologists are listed in **Supplementary Table 1**.

Feature Selection and Model Construction

In total, 1,393 radiomic features were extracted from each of the three MR scanning sequences. Then, 4,179 radiomic features were reduced to 3,040 by stability analysis.

In Model^{C}, six clinical parameters were selected including CA125, tumor differentiation by biopsy, DMI, the ratio of tumor infiltration depth to myometrium depth, LN size, and adnexa involvement, which were all observed on MR. In Model^{R}, two radiomic features were selected including correlation and HGLE. In Model^{CR1}, four risk factors including two clinical parameters (CA125 and LN size) and two radiomic features (correlation and HGLE) were used to build the prediction model (**Figure 3**) (13). The two radiomic features were extracted from the delayed phase of the 3D-Iso-LAVA and sagittal T2WI FSE, respectively. In Model^{CR2}, the LN size was removed and the same other three indicators (CA125, correlation, and HGLE) were selected. The detailed calculation formulas for Model^{CR1} and Model^{CR2} were given in **Supplementary Method 2**.

Assessment of Predictive Models

Model^{CR1} showed a significant ability in detecting pN+ with an AUC of 0.892 [95% confidence interval [CI]: 0.834–0.951] in the training cohort and an AUC of 0.883 (95% CI: 0.786–0.980) in the test cohort. Nomogram (**Figure 4C**) was established for Model^{CR1}. The p values calculated from Delong tests were 0.875, 0.8416, 0.7008, and 0.5865 for Model^{CR1}, Model^{CR2}, Model^{R}, and Model^{C}, respectively, indicating that there were no significant differences in AUCs between the training cohort and test cohort for each model. Performances of the four models in the training and test cohort are shown in **Figures 4A,B**. Based on the threshold determined by Youden's index in the training cohort, we used net reclassification index (NRI) to analyze the improvement brought by Model^{CR1} compared with other models. The results showed that Model^{CR1} outperformed Model^{R} (NRI $= 0.306$, $P < 0.001$), Model^{C} (NRI $= 0.134$, $P = 0.010$), and Model^{CR2} (NRI $= 0.090$, $P = 0.077$). Meanwhile, Model^{CR1} also significantly surpassed MR reports by radiologists (NRI $= 0.489$, $P = 0.006$). Besides, the calibration curves were plotted in both cohorts for further performance evaluation of Model^{CR1} (**Figures 5A,B**). Calibration curves show good fitness for probability of LNM (Hosmer–Lemeshow test, p-value $=$ 0.961 in the training cohort, 0.803 in the test cohort). **Figure 5C** shows patients' risk scores calculated from Model^{CR1}, intuitively indicating its high classification ability.

As shown in **Figure 6A**, in the subgroup of enlarged LNs, Model^{CR1} achieved the highest sensitivity of 0.970, equal to that predicted by MR report. In the subgroup of normal-sized LNs (0.3–0.8 cm), Model^{CR1} displayed the highest accuracy of 0.846 and a sensitivity of 0.647, which far surpassed the MR report (accuracy, 0.785; sensitivity, 0.235). In the subgroup of normal-sized LNs (<0.3 cm), ModelCR1 showed the best accuracy of 0.849 and a moderate sensitivity of 0.471, however, still greatly outperforming the MR report (accuracy, 0.817; sensitivity, 0.000). Meanwhile, F-score and accuracy in three subgroups are shown in **Figures 6B,C**, respectively. The highest F-score and most powerful accuracy of ModelCR1 were reflected among the five predictive models.

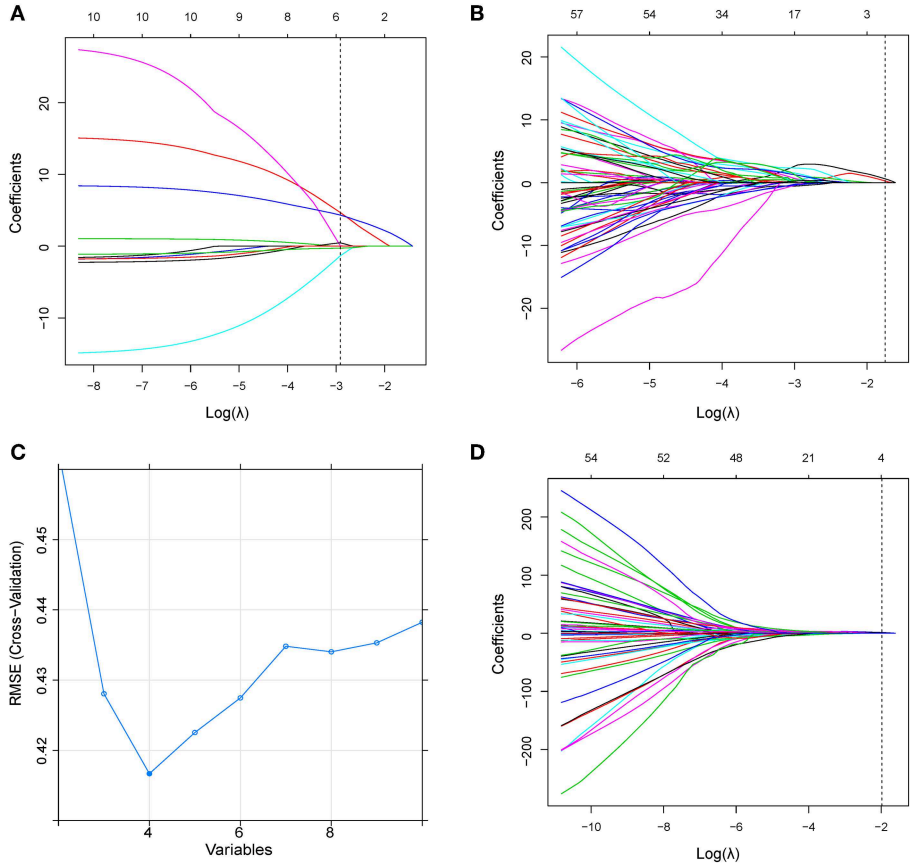


FIGURE 3 | (A) LASSO coefficient profiles of the clinical parameters in Model^{C}. According to the 1 standard error of the minimum criteria (the 1-SE criteria), the dotted line was plotted at the selected log(λ) (−2.914) via 10-fold cross-validation. **(B)** LASSO coefficient profiles of the radiomic features in Model^{R}. A log(λ) value of −1.750 was chosen (10-fold cross-validation, 1-SE criteria). **(C)** Feature selection using the RFE method in Model^{CR1}. The rank of feature importance was obtained using the random forest method; RFE built the model continuously by eliminating the lower ranking feature. The RMSE was used to select the optimal feature set in a 10-fold cross-validation. **(D)** LASSO coefficient profiles of the combined feature set in Model^{CR2}. A log(λ) value of −1.983 was chosen (10-fold cross-validation, 1-SE criteria). RFE, recursive feature elimination; RMSE, root mean square error.

DISCUSSION

In the present study, we developed four predictive models based on multiplanar DCE MR images and clinical parameters for LNM in EC patients. Model^{CR1}, which consisted of radiomic features, LN size, and CA125, showed the best discrimination ability, especially in patients with normal-sized LNs (diameter, 0–0.8 cm on MRI) and the sensitivity was greatly improved compared with the routine MR reports. The high *F*-scores indicated that while the sensitivity increased significantly, the PPV remained high.

A non-invasive and convenient preoperative assessment for LNM is crucial for EC treatment decision and prognosis prediction. Patients' data from preoperative procedures such as MRI, biopsy, and CA125 have been studied to assess LNM in recent years. MRI still remained the cornerstone in LN assessment in EC, showing satisfactory specificity but relatively low sensitivity. The combination of relative apparent diffusion coefficient (ADC) value and LN size was reported to result in a significant increase in sensitivity from 25 to 83% compared with conventional MRI (26); however, there have been conflicting reports in the literature regarding the detection of LNM at diffusion-weighted MR imaging (DWI). Nakai et al. (27) used 1.5-T MRI to evaluate nodal ADC values in gynecologic malignancies and were unable to differentiate benign from malignant LNs. Wang et al. (28) proposed a tumor biomarker predictive method by combining human epididymis secretory protein 4 (HE4) and CA125, achieving a high sensitivity of 94.1% but a low specificity of 30.7%. Notably, there is no clearly defined HE4 cutoff value for EC at present. In our study, we incorporated CA125 in our models, which was more generally accepted than HE4. Kang et al. (29) developed a low-risk prediction model for LNM based on MRI and serum CA125 data in endometrioid-type EC patients, and obtained sensitivity and specificity of 84.9 and 55.5%, respectively. Here, three MRI parameters including DMI, LN enlargement, and extension beyond uterine corpus were identified to be independent risk factors for LNM. In our study, we obtained CA125 and MR report LN size as risk factors for EC LNM prediction, which was similar to that result, and

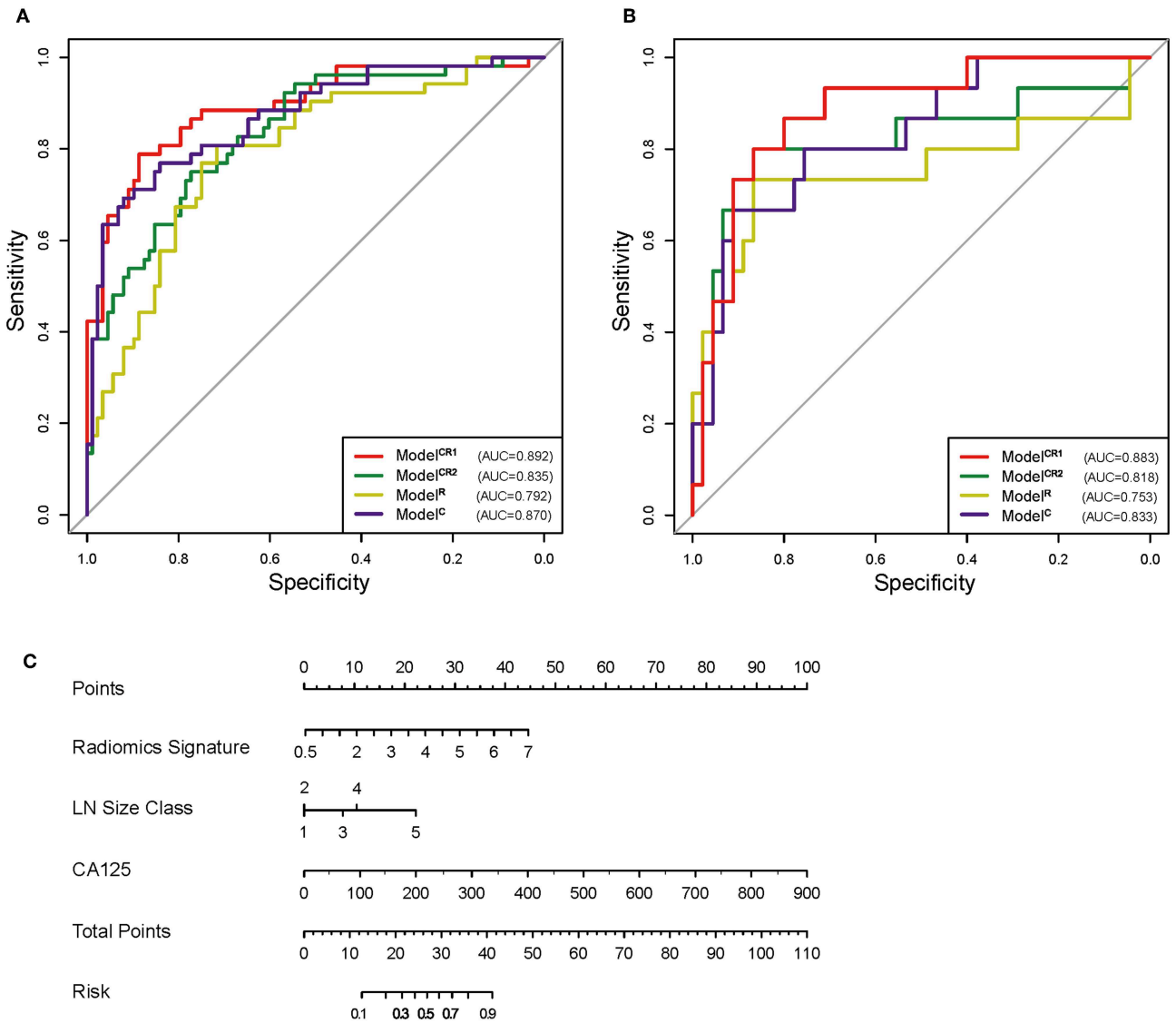


FIGURE 4 | (A) ROC of the four models in training cohort. **(B)** ROC of the four models in test cohort. **(C)** Preoperative nomogram of Model^{CR1}. ROC, receiver operating characteristic.

showed good discrimination ability on both cohorts and different LN size subgroups, especially for those normal-sized LNs, which previous researches had not yet focused on.

In our study, we collected and analyzed all available preoperative clinical parameters and established four prediction models. We aimed to determine the prediction efficiency of different models compared with the MR report in different sized LN subgroups. MRI uses several common morphological criteria in differentiating benign from malignant nodes (30) but nodal size still remains the commonly accepted standard. Low sensitivity is a recognized limitation when nodal size criteria are used on cross-sectional imaging, especially for normal-sized nodes due to limited spatial resolution. In this study, with node size gradually decreasing, the MR report and Model^{C} showed a decreasing sensitivity, whereas the Model^{R} and Model^{CR2} were more stable because of the high sensitivity in each sized LN subgroup (**Supplementary Figure 2**). The performance of the above classifiers confirmed our thoughts: When the LN was normal sized on MRI, combining LN size in classifiers could improve prediction accuracy but greatly reduce sensitivity. It is already accepted that normal-sized LNs may also contain metastases (31). The results of the MR report rely too much on LN size so that when LN size is normal on MRI, the sensitivity becomes very low. The concept remains the same when LN size is enlarged (>0.8 cm), then the specificity becomes very low. This can be due to the fact that it is usually difficult to differentiate enlarged nodes because of benign pathology, such as infection, granulomatous disease, and reactive hyperplasia vs. malignant

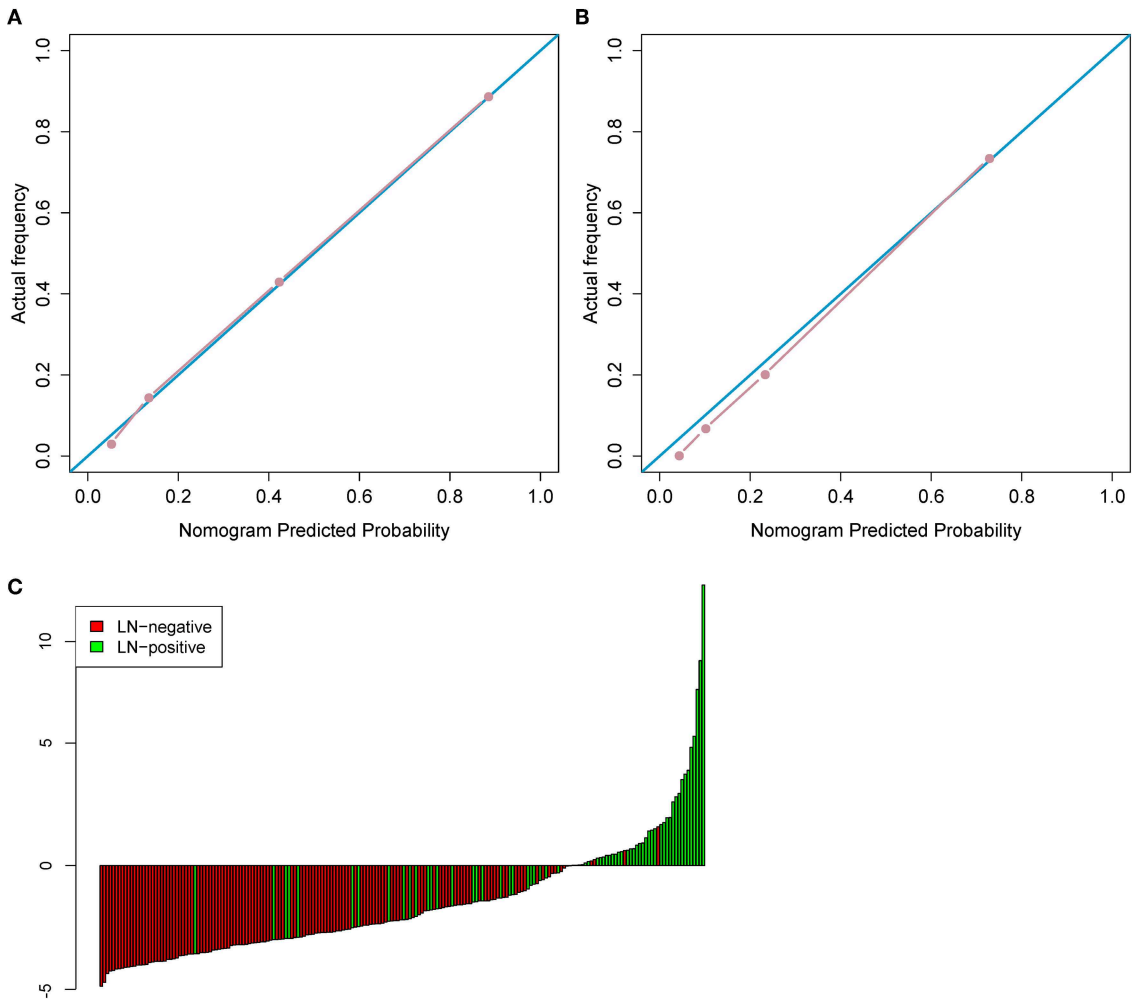


FIGURE 5 | (A) The calibration curve in the training cohort. **(B)** The calibration curve in the test cohort. **(C)** Patient risk score output by Model^{CR1}, while red bars show scores for those who were pathologically LN(–).

disease (30). LN size was not a significant predictor for the fusion model based on radiomics, although it could improve the predictive accuracy. This may indicate that there was enough information contained in the primary tumor region that could detect LNM. It is feasible to predict LNM status without the dependence on information of LNs.

Although Model^{CR1} showed a slightly lower sensitivity than Model^{R} and Model^{CR2} in the normal-sized group, its accuracy in each group is the highest, and the *F*-score with normal-sized LNs is greatly improved. Therefore, it was proposed as the optimal prediction model. To our knowledge, this is the first subgroup analysis on different sized LNs with preoperative nomogram study in EC.

Due to a variety of MR scanner parameter settings and scanner models, it is difficult to guarantee different scanners with exactly the same imaging quality, thus making it difficult to ensure the stability of radiomic features. By eliminating the radiomic features sensitive to scanner models and parameter settings in the training cohort, the radiomic model generalization ability can be improved. The radiomic texture features (correlation and HGLE) selected in Model^{CR1}, Model^{CR2}, and Model^{R} reflected two kinds of heterogeneity of VOI with a Pearson correlation coefficient of 0.095. Correlation shows the linear dependency of gray-level values to the corresponding voxels in the gray-level cooccurence matrix (GLCM) of MRI. HGLE is a measure of the proportion of areas with higher gray values in the tumor. These two radiomic features indicated that the extent of heterogeneity of tumor is associated with LNM. The more heterogeneous the tumor, the higher the risk of LNM.

The limitations of the present study include two aspects. First, there was no external validation. Multicenter investigation with a larger dataset was needed to further validate the generalization ability of our model. Second, genomic information was not yet incorporated into our models. A combination of gene marker panels and radiomic features will be promising in evaluation of EC.

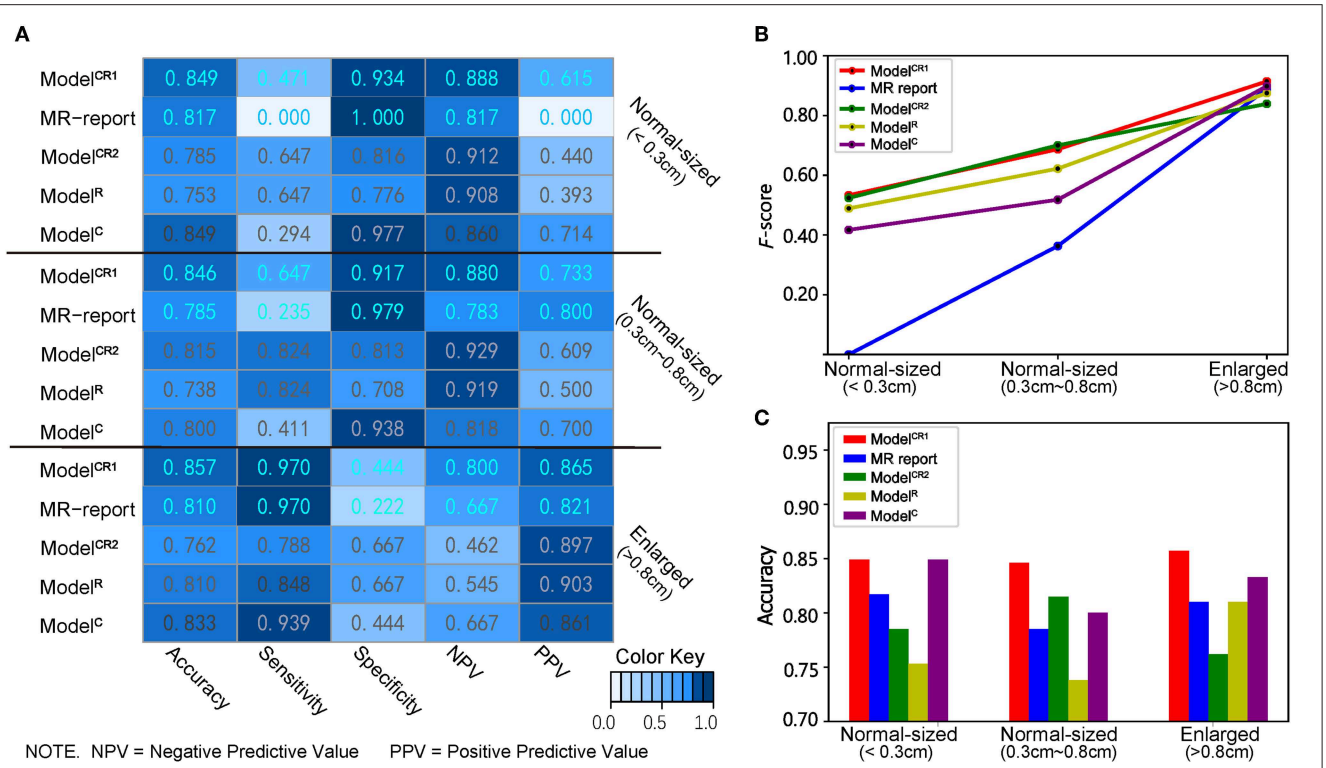


FIGURE 6 | (A) Heatmap of showing the classifiers' performance in different LN size subgroups. A deeper blue indicates a larger value. **(B)** Line chart of *F*-score, **(C)** histogram of accuracy in three subgroups. Red, blue, green, yellow, and purple lines and boxes, respectively, represent Model^{CR1}, MR report, Model^{CR2}, Model^{R}, and Model^{C}.

In conclusion, our study presented a predictive model based on multiplanar contrast enhanced MR images and incorporated both the radiomic features and clinical parameters, which showed good predictive accuracy for preoperative LNM in EC, especially in patients with normal-sized LNs.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

XX, HL, XZ, DD, JT, MF, and SW: conception and design. XX, HL, WF, and LZ: collection and assembly of data. HL, XX, DD, XZ, and MF: data analysis and interpretation. All authors: manuscript writing and final approval of manuscript.

FUNDING

This work was supported by the Beijing Hope Run Special Fund of Cancer Foundation of China (grant LC2016B01), the National Natural Science Foundation of China (81971776, 81771924, 81501616, 81671851, 81527805, and 61671449), National Key R&D Program of China (2017YFA0205200, 2017YFC1308700, 2017YFC1308701, 2017YFC1309100, and 2016YFC0103803), the Beijing Municipal Science and Technology Commission (Z171100000117023, Z161100002616022), the Beijing Natural Science Foundation (L182061), and the Youth Innovation Promotion Association CAS (2017175).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01007/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. (2018) 68:7–30. doi: 10.3322/caac.21442
2. Ghanem AI, Khan NT, Mahan M, Ibrahim A, Buekers T, Elshaikh MA. The impact of lymphadenectomy on survival endpoints in women with early stage uterine endometrioid carcinoma: a matched analysis. *Eur J Obstet Gynecol Reprod Biol*. (2017) 210:225–30. doi: 10.1016/j.ejogrb.2016.12.033
3. Chen W, Zheng R, Zhang S, Zeng H, Zuo T, Xia C, et al. Cancer incidence and mortality in China in 2013: an analysis based on urbanization level. *Chin J Cancer Res*. (2017) 29:1–10. doi: 10.21147/j.issn.1000-9604.2017.01.01
4. Korkmaz V, Meydanli MM, Yalçin I, Sari ME, Sahin H, Kocaman E, et al. Comparison of three different risk-stratification models for predicting lymph node involvement in endometrioid endometrial cancer clinically confined to the uterus. *J Gynecol Oncol*. (2017) 28:e78. doi: 10.3802/jgo.2017.28.e78
5. Papathemelis T, Scharl S, Kronberger K, Gerken M, Scharl A, Pauer A, et al. Survival benefit of pelvic and paraaortic lymphadenectomy in high-grade endometrial carcinoma: a retrospective population-based cohort analysis. *J Cancer Res Clin Oncol*. (2017) 143:2555–62. doi: 10.1007/s00432-017-2508-1
6. Biglia N, Librino A, Ottino MC, Panuccio E, Daniele A, Chahin A. Lower limb lymphedema and neurological complications after lymphadenectomy for gynecological cancer. *Int J Gynecol Cancer*. (2015) 25:521–5. doi: 10.1097/IGC.0000000000000341
7. Colombo N, Preti E, Landoni F, Carinelli S, Colombo A, Marini C, et al. ESMO Guidelines Working Group. Endometrial cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. (2013) 24(Suppl. 6):vi33–8. doi: 10.1093/annonc/mdt353
8. Khoury-Collado F, St Clair C, Abu-Rustum NR. Sentinel lymph node mapping in endometrial cancer: an update. *Oncologist*. (2016) 21:461–6. doi: 10.1634/theoncologist.2015-0473
9. Taufiq M, Masroor I, Hussain Z. Diagnostic accuracy of diffusion weighted magnetic resonance imaging in the detection of myometrial invasion in endometrial carcinoma. *J Coll Physicians Surg Pak*. (2016) 26:13–7.
10. Kim HJ, Cho A, Yun M, Kim YT, Kang WJ. Comparison of FDG PET/CT and MRI in lymph node staging of endometrial cancer. *Ann Nucl Med*. (2016) 30:104–13. doi: 10.1007/s12149-015-1037-8
11. Teng F, Zhang YF, Wang YM, Yu J, Lang X, Tian WY, et al. Contrast-enhanced MRI in preoperative assessment of myometrial and cervical invasion, and lymph node metastasis: diagnostic value and error analysis in endometrial carcinoma. *Acta Obstet Gynecol Scand*. (2015) 94:266–73. doi: 10.1111/aogs.12570
12. Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol*. (2019) 30:431–8. doi: 10.1093/annonc/mdz001
13. Liang W, Xu L, Yang P, Zhang L, Wan D, Huang Q, et al. Novel nomogram for preoperative prediction of early recurrence in intrahepatic cholangiocarcinoma. *Front Oncol*. (2018) 8:360. doi: 10.3389/fonc.2018.00360
14. Song J, Shi J, Dong D, Fang M, Zhong W, Wang K, et al. A new approach to predict progression-free survival in stage IV EGFR-mutant NSCLC patients with EGFR-TKI therapy. *Clin Cancer Res*. (2018) 24:3583–92. doi: 10.1158/1078-0432.CCR-17-2507
15. Li L, Wang K, Ma X, Liu Z, Wang S, Du J, et al. Radiomic analysis of multiparametric magnetic resonance imaging for differentiating skull base chordoma and chondrosarcoma. *Eur J Radiol*. (2019) 118:81–7. doi: 10.1016/j.ejrad.2019.07.006
16. Lu W, Zhong L, Dong D, Fang M, Dai Q, Leng S, et al. Radiomic analysis for preoperative prediction of cervical lymph node metastasis in patients with papillary thyroid carcinoma. *Eur J Radiol*. (2019) 118:231–8. doi: 10.1016/j.ejrad.2019.07.018
17. Dong Y, Feng Q, Yang W, Lu Z, Deng C, Zhang L, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI. *Eur Radiol*. (2018) 28:582–91. doi: 10.1007/s00330-017-5005-7
18. Shen C, Liu Z, Wang Z, Guo J, Zhang H, Wang Y, et al. Building CT radiomics based nomogram for preoperative esophageal cancer patients lymph node metastasis prediction. *Transl Oncol*. (2018) 11:815–24. doi: 10.1016/j.tranon.2018.04.005
19. Meissnitzer M, Forstner R. MRI of endometrium cancer—How we do it. *Cancer Imaging*. (2016) 16:11. doi: 10.1186/s40644-016-0069-1
20. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
21. Kurman RJ, Carcangiu ML, Herrington CS, Young RH. *WHO Classification of Tumors of Female Reproductive Organs [M]*. 4th edn. Lyon: IARC Press (2014).
22. Adapted from FIGO Committee on Gynecologic cancers. FIGO classification of cancer of the vulva, cervix, and corpus uteri. *Int J Gynecol Obstetr*. (2014) 115:97–8. doi: 10.1016/j.ijgo.2014.02.003
23. Harrel FE Jr, Lee KL, Mark DB. Tutorial in biostatistics: multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing error. *Stat Med*. (1996) 15:361–87. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
24. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med*. (2004) 23:2109–23. doi: 10.1002/sim.1802
25. Koplay M, Dogan NU, Erdogan H, Sivri M, Erol C, Nayman A, et al. Diagnostic efficacy of diffusion-weighted MRI for pre-operative assessment of myometrial and cervical invasion and pelvic lymph node metastasis in endometrial carcinoma. *J Med Imaging Radiat Oncol*. (2014) 58:538–46. doi: 10.1111/1754-9485.12209
26. Lin G, Ho KC, Wang JJ, Ng KK, Wai YY, Chen YT, et al. Detection of lymph node metastasis in cervical and uterine cancers by diffusion-weighted magnetic resonance imaging at 3T. *J Magn Reson Imaging*. (2008) 28:128–35. doi: 10.1002/jmri.21412
27. Nakai G, Matsuki M, Inada Y, Tatsugami F, Tanikake M, Narabayashi I, et al. Detection and evaluation of pelvic lymph nodes in patients with gynecologic malignancies using body diffusion-weighted magnetic resonance imaging. *J Comput Assist Tomogr*. (2008) 32:764–8. doi: 10.1097/RCT.0b013e318153fd43
28. Wang Y, Han C, Teng F, Bai Z, Tian W, Xue F. Predictive value of serum HE4 and CA125 concentrations for lymphatic metastasis of endometrial cancer. *Int J Gynaecol Obstet*. (2017) 136:58–63. doi: 10.1002/ijgo.12010
29. Kang S, Nam JH, Bae DS, Kim JW, Kim MH, Chen X, et al. Preoperative assessment of lymph node metastasis in endometrial cancer: a Korean Gynecologic Oncology Group study. *Cancer*. (2017) 123:263–72. doi: 10.1002/cncr.30349
30. Lai G, Rockall AG. Lymph node imaging in gynecologic malignancy. *Semin Ultrasound CT MR*. (2010) 31:363–76. doi: 10.1053/j.sult.2010.07.006
31. Thoeny HC, Froehlich JM, Triantafyllou M, Huesler J, Bains LJ, Vermathen P, et al. Metastases in normal-sized pelvic lymph nodes: detection with diffusion-weighted MR imaging. *Radiology*. (2014) 273:125–35. doi: 10.1148/radiol.14132921

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 11 October 2019
doi: 10.3389/fonc.2019.01045

Superpixel-Based Conditional Random Fields (SuperCRF): Incorporating Global and Local Context for Enhanced Deep Learning in Melanoma Histopathology

Konstantinos Zormpas-Petridis[1,2†], *Henrik Failmezger*[3†], *Shan E Ahmed Raza*[3], *Ioannis Roxanis*[3,4], *Yann Jamin*[1,2] *and Yinyin Yuan*[3*]

[1] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, United Kingdom, [2] The Royal Marsden NHS Trust, Surrey, United Kingdom, [3] Division of Molecular Pathology, The Institute of Cancer Research, London, United Kingdom, [4] Royal Free London NHS Foundation Trust, London, United Kingdom

OPEN ACCESS

Edited by:
Lei Deng,
Jacobi Medical Center, United States

Reviewed by:
Sandra Avila,
Campinas State University, Brazil
Sara Hosseinzadeh Kassani,
University of Saskatchewan, Canada

***Correspondence:**
Yinyin Yuan
yinyin.yuan@icr.ac.uk

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 11 June 2019
Accepted: 25 September 2019
Published: 11 October 2019

Citation:
Zormpas-Petridis K, Failmezger H, Raza SEA, Roxanis I, Jamin Y and Yuan Y (2019) Superpixel-Based Conditional Random Fields (SuperCRF): Incorporating Global and Local Context for Enhanced Deep Learning in Melanoma Histopathology. Front. Oncol. 9:1045. doi: 10.3389/fonc.2019.01045

Computational pathology-based cell classification algorithms are revolutionizing the study of the tumor microenvironment and can provide novel predictive/prognosis biomarkers crucial for the delivery of precision oncology. Current algorithms used on hematoxylin and eosin slides are based on individual cell nuclei morphology with limited local context features. Here, we propose a novel multi-resolution hierarchical framework (SuperCRF) inspired by the way pathologists perceive regional tissue architecture to improve cell classification and demonstrate its clinical applications. We develop SuperCRF by training a state-of-art deep learning spatially constrained- convolution neural network (SC-CNN) to detect and classify cells from 105 high-resolution (20×) H&E-stained slides of The Cancer Genome Atlas melanoma dataset and subsequently, a conditional random field (CRF) by combining cellular neighborhood with tumor regional classification from lower resolution images (5, 1.25×) given by a superpixel-based machine learning framework. SuperCRF led to an 11.85% overall improvement in the accuracy of the state-of-art deep learning SC-CNN cell classifier. Consistent with a stroma-mediated immune suppressive microenvironment, SuperCRF demonstrated that (i) a high ratio of lymphocytes to all lymphocytes within the stromal compartment ($p = 0.026$) and (ii) a high ratio of stromal cells to all cells ($p < 0.0001$ compared to $p = 0.039$ for SC-CNN only) are associated with poor survival in patients with melanoma. SuperCRF improves cell classification by introducing global and local context-based information and can be implemented in combination with any single-cell classifier. SuperCRF provides valuable tools to study the tumor microenvironment and identify predictors of survival and response to therapy.

Keywords: deep learning, machine learning, conditional random fields, digital pathology, cell classification, melanoma, tumor microenvironment

INTRODUCTION

Cancer is a highly complex, non-autonomous disease. The interactions between microenvironmental selective pressures and cancer cells dictate how cancer progresses and evolves. Accurate and spatially explicit characterization of the tumor microenvironmental landscape including how cancer cells interact with the extra-cellular matrix and other cellular players such as stromal cells and immune cells within the tumoral niche, is needed to understand the context in which cancer evolves, and may also provide robust predictor of cancer behavior for risk-stratification (1). More specifically the recent success of cancer immunotherapy including the spectacular response observed in patients with previously incurable melanoma, a highly aggressive form of skin cancer, calls for a better understanding of the cancer-immune interface.

In the new era of digital pathology, advanced image analysis can objectively, consistently, and quantitatively characterize the different components of the tumor and how they spatially interact, and as a result assist pathologists in tasks such as tumor grading (2). Algorithms for cell detection and classification are key components of this process. Machine learning, and more recently deep learning algorithms, both exploiting the phenotypic differences in nuclear morphology between each cell type, revolutionized the field yielding significantly better cell detection, segmentation, and classification results (3–9).

However, even state-of-the-art deep learning algorithms can underperform especially in cases where different cell types appear morphologically similar. Current computed pathology tools focus on individual cell nuclei morphology with limited abstract local context features, whereas pathologists incorporate regional tissue architecture (in practice, by zooming in/out), together with cell morphological features to accurately classify cells.

Here, we hypothesize that robust tumor regional classification from lower resolution images can provide the contextual information that is key to further improve single cell classification algorithms. Our aim is to introduce dependencies on global tissue context and cell neighborhood and enhance learning results for cell classification from deep convolution neural networks (CNNs). Probabilistic graphical models have successfully been applied to improve cell classification in time-lapse imaging by taking into account the temporal context of a cell (10–15). Probabilistic graphical models have also been used successfully in histopathology images for pathology detection and segmentation (16–19), disease and tissue staging (20, 21), and nuclei segmentation (22). In our study, instead of time dependency, we apply graphical models to introduce the spatial context of a cell as additional information to improve single-cell classification. A multi-resolution hierarchical framework was proposed to mirror the way pathologists perceive tumor architecture, and applied to whole-slide images (WSI) hematoxylin and eosin (H&E)-stained slides of melanoma skin cancer (**Figure 1A**). We demonstrated that our new system is computationally efficient and significantly improves single cell classification. The increased accuracy in cell classification further enabled us to shed new light on the understanding of cancer-immune-stroma interface of melanoma.

MATERIALS AND METHODS

Datasets

In total, 105 full-face, H&E stained section images from formalin-fixed, paraffin-embedded (FFPE) diagnostic blocks of melanoma skin cancer from The Cancer Genome Atlas (TCGA) were used. We scaled all digitized (Aperio ImageScope) histology images to 20, 5, and 1.25× magnification with pixel resolution 0.504, 2.016, and 8.064 μm, respectively, using Bio-Formats (https://www.openmicroscopy.org/bio-formats/). WSIs at 20× magnification (representative size: 30,000 × 30,000 pixels), were split into sub-images (tiles) of 2,000 × 2,000 pixels each, for computational efficiency.

For the purpose of training and testing the different parts of our system we divided the dataset into sub-datasets, namely single-cell classification dataset, 5× sub-dataset, 1.25× sub-dataset and discovery sub-dataset (**Table 1**, also see **Supplementary Tables 1–4**).

Single-Cell Classification Using a Spatially Constrained Convolutional Neural Network

We used a Spatially Constrained Convolutional Neural Network (SC-CNN) (6) for single cell classification (**Figure 1E**). SC-CNN uses spatial regression in order to predict the probability of a pixel being the center of the nucleus. The nucleus is classified by a neighboring ensemble predictor (NEP) in conjunction with a standard softmax CNN. We randomly initialized the network's layers as we have found that to perform better than transfer learning from real-world datasets in our experiments with pathological samples.

Superpixel-Based Tumor Region Classification

A machine learning superpixel-based framework was implemented in Matlab (23) to classify tumor tissue regions and was subsequently applied to low resolution (5 and 1.25×) images. Reinhard stain normalization (24) was applied separately on each of the 5 and 1.25× sub-datasets to account for stain variabilities that could affect the classification (25).

Downscaled images were segmented using the simple linear iterative clustering (SLIC) superpixels algorithm (26), which is designed to provide roughly uniform superpixels. Choosing the optimal number of superpixels is important to ensure that the superpixels capture homogeneous areas and adhere to image boundaries. With our pathologist's input, we visually identified a size of superpixels that met these criteria and chose the number of superpixels automatically based on each image's size (Equation 1).

$$N_i = ceiling\left(\frac{S_i}{U}\right) \quad (1)$$

where N_i is the number of superpixels in the ith image, S_i is the size of image i in pixels, and U (here $U = 1{,}250$) is a constant held across all images that defined a desired size of the superpixels. This means, on average, a superpixel occupies an area of approximately 35 × 35 pixels, equivalent to 280 × 280 mm^2. We identified the superpixels belonging to each area by

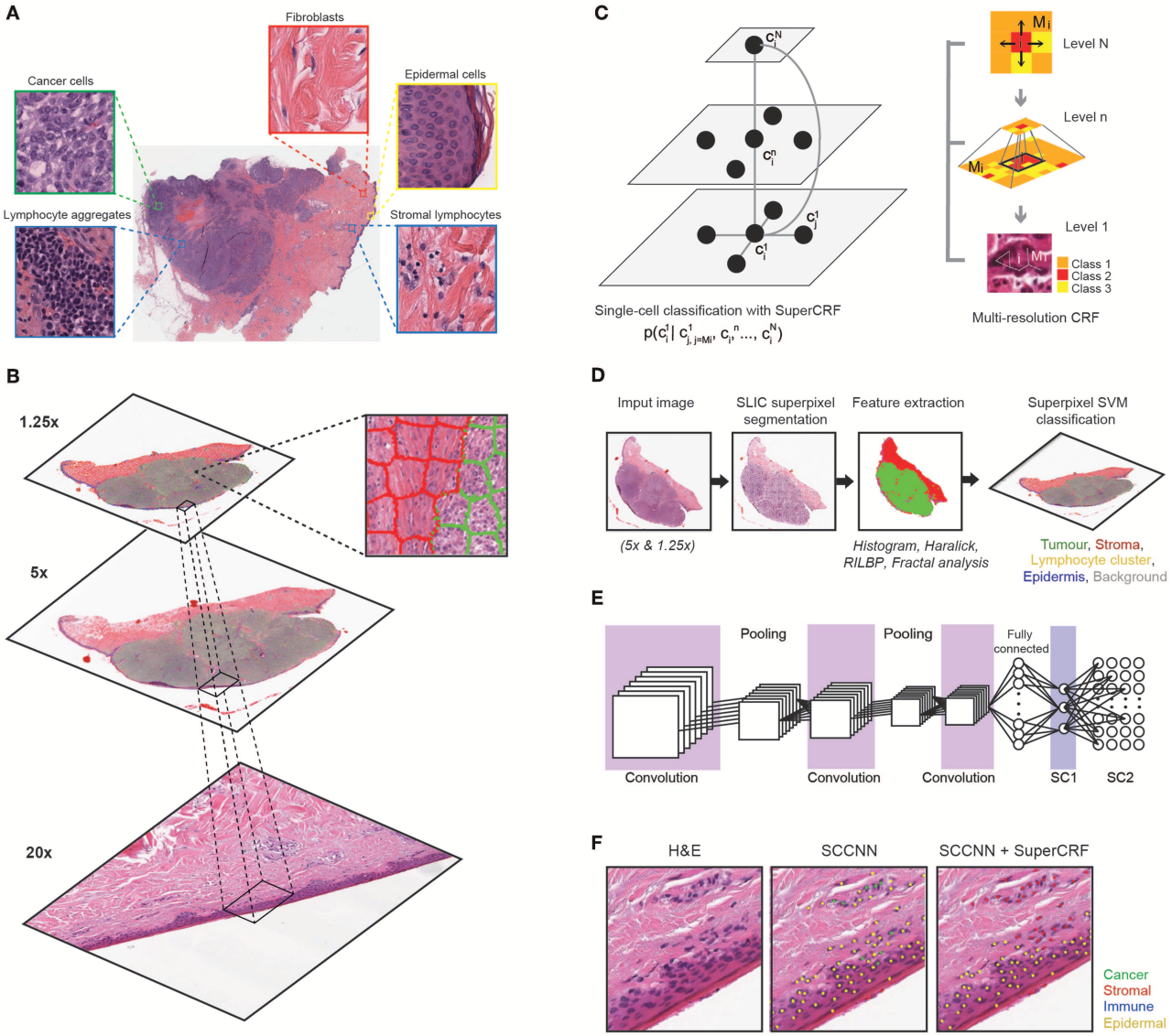


FIGURE 1 | Overview of the SuperCRF framework for analyzing H&E-stained pathological images of melanoma. **(A)** Major histological features of melanoma architecture. **(B)** Projection of regional classification results using superpixels from various scales to the 20× magnification for the improvement of single-cell classification. **(C)** Graphical representation of node dependencies (cells and superpixels) across different scales. **(D)** Region classification scheme using a superpixel based machine-learning method in whole-slide images (5× and 1.25× magnification) **(E)** Single-cell classification using a state-of-the-art spatially constrained-convolution neural network (SC-CNN) classifier **(F)** representative results of the SC-CNN cell classifier alone and combined with our SuperCRF system. Note the misclassification of various stromal cells by the SC-CNN, which are corrected by our model.

determining whether their central points fell within the regions annotated by the pathologist.

Overall, for the 1.25× training sub-dataset, we found 15,477 superpixels belonging in tumor areas, 6,989 in stroma areas, 141 in epidermis and 691 in lumen/white space, while for the 5× training sub-dataset we found 1,193 superpixels belonging in tumor areas, 1,324 in stroma areas, 360 in epidermis, 506 in lymphocyte clusters and 830 in lumen/white space.

Next, we extracted four types of features, 85 in total, from each superpixel, including seven histogram features (mean values of hue, saturation, and brightness, sum of intensities, contrast, standard deviation, and entropy), and well-established texture features [12 Haralick features (27), 59 rotation-invariant local binary patterns (RILBP), 7 segmentation-based fractal texture analysis (SFTA) features (28)]. Features were standardized into z-scores. The mean values and standard deviation of the features from the training set were used for the normalization of the test set. A support vector machine (SVM) with a radial basis function (RBF, $\gamma = 1/\text{number\_of\_features}$) was trained with these

TABLE 1 | Summary of the data used to train and test the different parts of the SuperCRF system, as well as study the cancer-immune-stroma interface (also, see **Supplementary Tables 1–4**).

Name	Number of WSIs	Purpose
Single-cell classification sub-dataset	**8** Training SC-CNN: 3 (348 tiles) Training SuperCRF: 2 (84 tiles) Testing: 3 (290 tiles)	Single-cell classification into four categories: cancer cells, lymphocytes, stromal cells, epidermal cells
5x sub-dataset	**16** Training: 10 Testing: 6	Region classification into five categories: tumor, normal stroma, lymphocyte cluster, normal epidermis, lumen/white space
1.25x sub-dataset	**58** Training: 21 Testing: 37	Region classification into four categories: tumor, normal stroma, normal epidermis, lumen/white space
Discovery dataset	**97**	Study of the tumor-stroma interface. To accelerate the analysis, 50 tiles (2,000 × 2,000 pixels) containing tumors were randomly sampled from every whole-slide image (WSI)

The values are bold for visual (illustration) purposes.

features to classify superpixels into different biologically meaningful categories.

For the 5× sub-dataset, superpixels were classified into five categories: tumor area, normal stroma, normal epidermis, lymphocytes cluster, and lumen/white space. We increased the penalty in the cost function for the epidermis and lumen/white space classes by a factor of 10 when training the SVM, to account for class imbalance. For the 1.25× sub-dataset superpixels classification consisted of four categories: tumor area, normal stroma, normal epidermis, and lumen/white space. We randomly selected a subset of 5,000 cancer and stroma superpixels and increased the penalty in the cost function for the epidermis and lumen/white space classes by a factor of 10, again to account for class imbalance (**Figure 1D**).

SuperCRF

Single-cell based classification approaches often assign a class label based on the morphology of -individual cells, regardless of their neighboring cells. However, these spatial relationships provide important information that is used by pathologists. Conditional random fields (CRF) are undirected graphical models that represent efficient ways to model dependences, by factorizing the probability density into a specific set of conditional dependence (29). Therefore, the tumor microenvironment can be modeled by a CRF by introducing nodes for cells and superpixels, as well as edges whenever there is a spatial relationship between nodes.

We excluded lymphocytes from the CRF assumption that neighboring cells have a higher probability to share the same class labels, since they infiltrate, in an inconsistent manner ranging from sparse to highly dense, in tumor as well as stromal tissue. Therefore, lymphocytes kept their label as assigned by the SC-CNN.

Let n be the total number of cells (besides lymphocytes) in the image and $c_i \in \{stromal, cancer, epidermis\}$, $i = 1, 2, \ldots, n$ the input labels of the cells as assigned by the SC-CNN. Let s_i, be the corresponding superpixel for a cell c_i with $s_i \in \{stromal, cancer, epidermis, white\ space\}$ for 1.25× superpixels and $s_i \in \{stromal, cancer, epidermis, lymphocyte, whitespace\}$ for 5× superpixels. $\mathbf{x} \in \{\mathbf{c}, \mathbf{s}\}$ comprises the nodes of the CRF. The CRF assigns output labels $y_i \in \{stromal, cancer, epidermis, lymphocyte, white\ space\}$ based on the input data. The joint probability distribution over input data and output labels, $p(y_1, y_2, \ldots, y_n \lfloor x_1, x_2, \ldots, x_n)$ can be modeled by factorizing the probability density into a specific set of conditional dependence relationships (**Figure 1C**).

$$p(y|\,x) = \prod_{n}^{i=1} p(y_i\,|x_i) = \frac{1}{Z}\exp(\sum E(x_i, y_i, xN_i, yN_i)) \quad (2)$$

where Z is a normalizing constant, w is a weight vector and

$$E(x_i, y_i, xN_i, yN_i,) = \sum \Phi(x_i, y_i) + \sum \psi_c(xN_i, yN_i) \quad (3)$$

defines the energy function of the CRF.

The node potentials $\Phi(x_i, y_i)$ represent the evidence that a cell i, with the input label x_i takes the class label y_i. The node potential can be defined as $\Phi(x_i, y_i) = f(x_i, y_i)$+b, with $f(x_i, y_i) = \begin{cases} 1 & if\ y_i = x_i \\ 0 & else \end{cases}$ and b representing the bias.

The edge potentials $\psi_c(xNi, yNi)$ model the probability that neighboring cells take a similar cell label. N_i is the neighborhood of cell i, defined as all the cells that can be found in a defined distance. The edge potentials are defined as: $\psi_c(x_i, y_i, xN_i, yN_i) = f(x_i, y_i) * f(xN_i, yN_i) + b$.

The CRF was trained with stochastic gradient descent and the decoding was applied using loopy belief propagation. The toolbox of M. Schmidt was used to train and decode the CRF (30).

The source code for the study is available at Github (https://github.com/Henrik86/SuperCRF).

Survival Analysis

We evaluated the prognosis value of the abundance of stromal cells and location of lymphocytes in our discovery sub-dataset. The ratio of stromal cells to all cells, the ratio of lymphocytes in cancer areas to all lymphocytes, and the ratio of lymphocytes in stroma areas to all lymphocytes were calculated for each patient.

Patients were divided into high and low ratio groups, split at the median value of all scores. Patients with a ratio of lymphocytes being high inside the tumor area and low in the stroma were categorized as the "immune infiltration" group whereas patients with a ratio of lymphocytes being low in the tumor area and high in the stroma were categorized as "immune excluded," based on the recent classification of the main immune phenotypes of anticancer immunity that predict response to immunotherapy (31). The number of patients belonging to neither of these two groups (high/high $n = 6$ and low/low $n = 5$) was too small to perform the survival analysis. Non-parametric Kaplan-Meier estimation was used to analyze overall survival in 94 patients. Differences between survival estimates were assessed with the log-rank test. Finally, Cox regression models were adjusted,

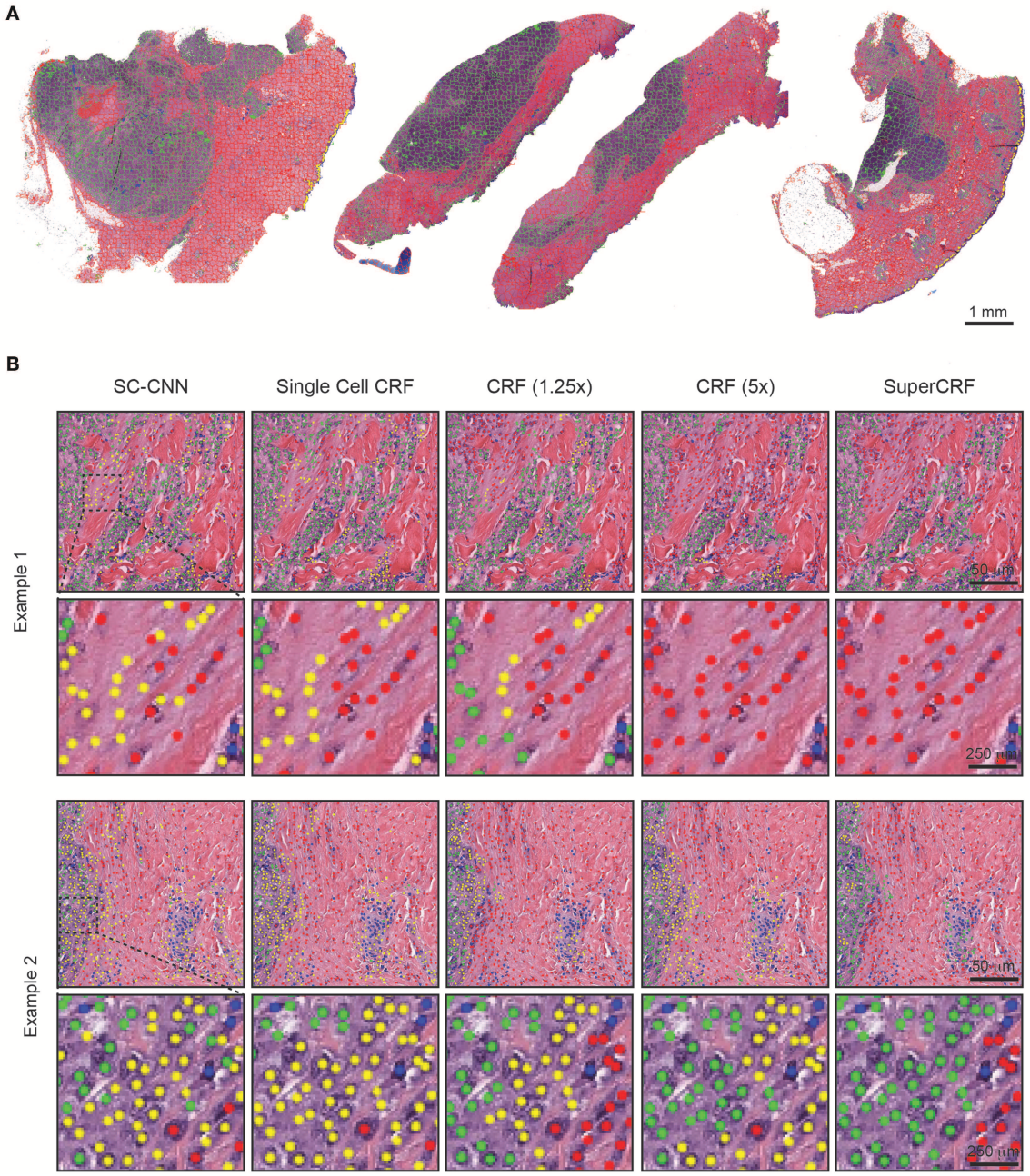


FIGURE 2 | Representative examples of both superpixel and single-cell classification with or without SuperCRF. **(A)** Superpixels-based regional classification on representative whole slide images (5× magnification) of melanoma. Green: tumor area, Red: stroma area, Blue: normal epidermis, Yellow: lymphocyte cluster. **(B)** Representative images showing cell classification using a state-of-the-art spatially constrained-convolution neural network (SC-CNN) and four conditional random fields (CRF) models. Note the mislabeling of many cancer and stromal cells as epidermis cells when using the SC-CNN and the gradual increase in classification accuracy with the best accuracy achieved with the SuperCRF. Green, cancer cells; Red, stromal cells; Blue, lymphocytes; Yellow, epidermis cells.

testing for the independent prognostic relevance of our risk scores. To test if Breslow-thickness (the distance between the upper layer of the epidermis and the deepest point of tumor penetration) was contributing to a high ratio of stromal cells, we created a multivariate model containing both stromal cells ratio and Breslow-thickness, as well as two univariate models containing the covariates separately. Pearson's correlation was used to test for linear relation between the two variables.

RESULTS

SuperCRF Improves Accuracy of Cell Classification

First, we trained the state-of-the-art deep learning method, spatially-constrained CNN (SC-CNN) algorithm, to detect and classify cells in high resolution (20×) WSI into four categories: cancer cells, stroma cells, lymphocytes, and epidermis cells. The SC-CNN network yielded an accuracy of 84.63% over 4,059 cells in the independent test set (**Table 1**, **Supplementary Table 5**). Visual inspection revealed that the majority of false positives were misclassification of stromal and cancer cells as epidermis, which confirmed our initial motive for the incorporation of regional and spatial information to improve classification.

Subsequently, we trained a conditional random field (CRF) by combining the cellular neighborhood with tumor region classification (cancer area, normal stroma, normal epidermis, lymphocyte cluster, and lumen/white space) from low resolution images (5 and 1.25×, **Figure 1B**), given by the superpixel-based machine-learning framework. The SLIC superpixels algorithm has previously been shown to be computationally efficient, requiring only 3s on average to segment a single downscaled image of 2,500 × 2,500 pixels using a 2.9 GHz Intel core i7 processor. Performance of classification using individual and various combinations of feature sets was tested and the use of all 85 features, yielded the highest accuracy (23). It was then applied on the two datasets of 1.25 and 5× magnification (**Figure 2A**) and achieved high accuracy in regional classification (1.25× sub-dataset: Overall accuracy 97.7% in the training set using 10-fold cross validation and 95.7% in 2,997 superpixels annotated in the 37 images of the independent test set; 5× sub-dataset: Overall accuracy 97.1% in the training set using 10-fold cross validation and 95.2% in 1,798 superpixels annotated in the six images of the independent test set).

To train SuperCRF, we first introduced dependencies on cell neighborhood. Cells were considered neighbors in the CRF, if they were located in a spatial proximity of 15 μm (or 30 pixels), which resulted in an average of 1.3 neighbors per cell. Subsequently, we integrated this local neighborhood with global context by connecting the CRF single-cell nodes to the regional classification results from superpixels. To determine the best configuration, we trained four different CRFs and compared their performance in terms of single-cell classification on a test set, including three samples, 290 tiles and 4,059 single-cell annotations (1,527 cancer cells, 676 lymphocytes, 837 normal epidermis cells, 1,019 stromal cells).

In detail, for the first CRF we did not use any context classification, just cell neighborhood dependencies, i.e., the only edges of the CRF were between neighboring cells (singleCellCRF). For the second and third CRF we introduced superpixel nodes. Now, single-cell nodes are not only connected to neighboring cells but every single-cell node is also connected to a superpixel node. We trained a CRF for 5× superpixel classification (CRF5×) and 1.25× superpixel classification (CRF1.25×). Furthermore, we trained a CRF in which every single-cell node was connected to two superpixel nodes in 5 and 1.25× resolution (SuperCRF). Already the singleCellCRF (Accuracy: 87.6%, Precision: 89.7%, Recall: 89.5%, **Table 2**) improves the classification accuracy compared to the SC-CNN (84.6%, Precision: 87.6%, Recall: 88.1%, **Table 2**). However, the use of contextual information by the introduction of superpixel nodes, markedly improves the classification metrics (Accuracy 90.8%, Precision: 92.5%, Recall: 91.1%, **Table 2**) for CRF1.25× and (Accuracy 91.7%, Precision: 93%, Recall: 91.3%, **Table 2**) for the CRF5×. The SuperCRF, using nodes from superpixels in both 5 and 1.25× resolution images, as well as the neighboring cells, resulted in the highest classification outcome (Accuracy 96.5%, Precision: 96.4%, Recall: 96.3%, **Table 2**, **Figures 1F**, **2B**, **Supplementary Tables 5–9**).

TABLE 2 | Evaluation of different conditional random fields (CRF) versions and a state-of-the-art spatially constrained-convolution neural network (SC-CNN) deep learning cell-classifier.

Method	Accuracy (%)	Precision	Recall
SC-CNN	84.63	0.8756	0.8808
singleCellCRF	87.61	0.8973	0.8946
CRF1.25×	90.79	0.9248	0.9110
CRF5×	91.70	0.9298	0.9126
SuperCRF	**96.48**	**0.9644**	**0.9629**

The values are bold to indicate the highest achieved accuracy, precision and recall.

SuperCRF's Increased Accuracy of Cell Classification Improves Confidence in Stromal Cell Ratio as a Predictive Feature of Survival in Melanoma

The crosstalk between cancer cells and stromal cells play an active role in tumor invasion and metastasis, and controlling immune infiltration and is increasingly recognized as a hallmark of cancer (32). Tumor-stromal cell ratio has been shown to hold prognostic and predictive information in patient with solid tumors (31, 33, 34). Here, we demonstrate that a high stromal cell ratio is also a predictor of poor prognosis in melanoma using both values derived from the multivariate models of SC-CNN and SuperCRF in our discovery sub-dataset. Yet SuperCRF yields a significantly higher confidence in the predictive value of the stromal cell ratio (SuperCRF: $p < 0.0001$, Coxph-Regression (discretized by median): HR = 4.1, $p = 0.006$; SC-CNN: $p = 0.039$, Coxph-Regression (discretized by median): HR = 2.4, $p = 0.05$, **Figure 3A**).

Similar regression coefficients for both stromal cells ratio and Breslow-thickness covariates were observed between the multivariate and the two univariate survival models (1.404 and

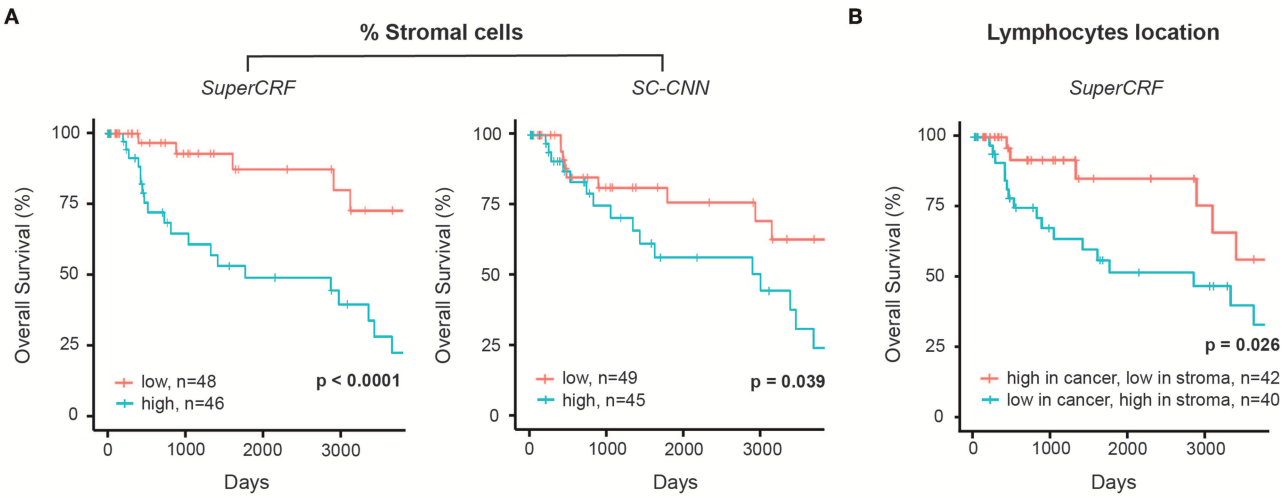


FIGURE 3 | Associations between survival outcomes and SuperCRF-define risk groups in the Cancer Genome Atlas (TCGA) cohorts of patients with melanoma. **(A)** Kaplan-Meier Survival curves for patients in the high-risk group (blue) and low risk group classified by stromal cells ratio derived from SuperCRF (left) and using only the SC-CNN classifier. Note the difference in the p-value using the two methods. **(B)** Kaplan-Meier Survival curves for patients in the high-risk group (blue) and low risk group classified by immune phenotype based on spatial distribution of lymphocytes in different tumor compartments derived from SuperCRF.

0.171, respectively, for the multivariate model and 1.633 and 0.179 for the univariate models) of the SuperCRF. Pearson's correlation showed no correlation between stromal cells ratio and Breslow-thickness ($r = -0.05$), overall indicating that stromal cells ratio is independent to Breslow-thickness.

Combining Cell and Region Classification: Location of the Immune Infiltrate Is Predictive of Survival in Melanoma

There is increasing evidence of the value of immune infiltration to provide prognostic information and predictors of response in patient with melanoma [recently reviewed in (35)]. The spatial compartmentalization of immune cells afforded by our SuperCRF (by the cell and region classification results) was used to define the recently-described main immune phenotypes of anticancer immunity that predict response to immunotherapy (31). Patients with a classified "immune excluded" phenotype, defined by a low lymphocyte ratio inside the tumor area and high inside the stroma area, was associated with a significantly worse prognosis compared to "inflamed" tumors characterized by a high ratio of lymphocytes inside the tumor and a low ratio inside the stroma ($p = 0.026$, Cox PH –regression: HR = 2.57, $p = 0.032$, **Figure 3B**). Taken together, our data is consistent with the model of a stroma-mediated immune suppressive microenvironment that exclude T cells from the vicinity of cancer cells.

DISCUSSION

In this study, we implemented a framework which fuses traditional machine learning with deep learning to model the way pathologists incorporate large-scale tissue architecture and context across spatial scales, to improve single-cell classification in large whole-section slide images. Using this approach, we demonstrated a marked 11.85% overall improvement in the accuracy of the state-of-art deep learning SC-CNN cell classifier. Also, the similar values of both precision and recall and their simultaneous increase in every step show the unbiased nature of our approach.

Computational pathology algorithms, typically exploit the inter-cell phenotypic differences for cell classification, yet even state-of-art deep learning algorithms tend to underperform in this task, mainly due to the disproportional numbers of cells sharing similar nuclear morphological features, or due to intra-class diversity, seen for example in tumor stroma (fatty tissue, necrosis, vessels, muscle, fibroblasts, and associated collagen). Whilst computers can quantify morphological differences in a considerably more complex way, pathologists still generally outperform computers in cell classification. An essential reason is that they incorporate key contextual information such as heterogeneous tissue architecture, together with cell morphological features.

The idea that a cancer cell is dependent on its neighboring cells and global context is comparable to the fundamental concept in landscape ecology that a living population depends on the existing habitats and is not equally spread on the terrain. A particular habitat could favor the development of specific organisms. In practice, landscape ecologists denote the habitats from satellite images and then "ground-truth" them by detailed small-scale sampling of the habitats of interest (36). This inspired the design of our framework by introducing CRF dependencies between (i) the cells and their neighbors and (ii) the cells and to the global context (i.e., habitats from low resolution captured by the classified superpixels).

Our proposed framework connects deep learning and classical image processing using probabilistic graphical models. All the information was combined using a CRF graphical model, which have been widely applied in image analysis for pathological images, yet mainly for semantic segmentation (16, 17, 37, 38). Here, (1) we introduce a new way to capture high-level spatial context using superpixels, (2) propose a new CRF model that introduces dependences over space and across different spatial scales, thereby modeling multiple cells and their associated superpixels simultaneously for more accurate classification, (3) introduce the concept of context-specific CRF modeling, given that the strength of dependence can be variable according to tumor compartments. There is an increasing interest in combining deep learning with different strategies, or "umbrella approaches," such as the use of traditional machine learning to spatially explicit context used in this study, with the aim to, not only refine and improve the overall existing deep learning network (17, 39–41), but also facilitate biological interpretation compared to the "black-box"-like approach of deep-learning-only methods. However, optimizing and inventing new and refined deep learning networks is of equal importance, as during experimentation we observed that the better we made our single-cell classifier baseline, the more effective our SuperCRF approach became.

We also showed that combining cell classification with the global context given by the region classification (both inherent parts of the SuperCRF architecture) can open new avenues to study the cancer microenvironment from histopathological slides. For example, the spectacular response observed in clinical trials of immunotherapy in patients with incurable melanoma calls for a better understanding of the tumor microenvironment and in particular the cancer-immune-stroma interface. Here, our approach and its ability to look at lymphocytes within their cellular and global context can predict melanoma patient survival and potentially provide biomarker stratification for immunotherapeutic approaches, by identifying the three main types of tumor immunophenotypes including (i) inflamed tumors which are characterized by infiltrated T Cells within the tumor, and associated with a generally good prognosis (ii) immune-excluded tumors, in which T cells are present but prevented to infiltrate the tumor due to stromal interaction, and associated with worse prognosis (and obviously (iii) immune desert tumors). This could also potentially be extended to provide quantitative biomarkers to characterize the immune infiltrating response to immunotherapy. We also demonstrated that in accordance with the immune-excluded phenotype, tumors rich in stromal cells had a marked poorer prognosis in patients with melanoma. With p-value lower by two orders of magnitude, our method provide stronger predictive power than by using deep-learning only method for cell classification.

In the future, we plan to extend our framework and include an upward optimization step for the superpixels which may include additional classes for cells, regions and structures in order to provide a complete characterization of the tumor microenvironment. This may include deriving further classes from higher resolution images as we did for lymphocyte clusters in this study which were difficult to visualize in 1.25× resolution images. Incorporating additional deep learning methods should also be explored to perfect the classification of superpixels, for example by incorporating features extracted from a DCNN or a deep autoencoder, or to provide a potential alternative to superpixels, which may not be appropriate for the characterization of complicated structures, such as glands (42).

The primary aim of this study was to demonstrate proof-of-principle that the introduction of global and local context as cell dependencies using a probabilistic graphical model as a post-processing step, like an "umbrella," can significantly improve the performance of deep learning or classical machine learning cell classifiers based only on cell-morphology and abstract local context information. We chose the SC-CNN architecture as our primary cell classification step due to its state-of-the-art performance in cell detection and classification compared to other well-established deep learning and classical machine learning approaches (6). Alternatively, other promising deep learning networks could potentially be used including Inception v3 (43), Inception v4 (44), or a VGG architecture (45).

Overall, our vision is to establish a network which will provide a complete characterization of every component of the tumor microenvironment where all the parts will interact with each other like an ecological landscape. Such system has immense potential and can be virtually transferred to any cancer type, to provide a better understanding of the cancer-immune cell interface, cell-stroma interactions, and predictive biomarkers of response to novel therapies, including immunotherapy, which has radically changed melanoma patient survival.

CONCLUSION

The novel general framework SuperCRF improves cell classification by introducing global and local context-based information much like pathologists do. SuperCRF can be implemented in combination with any single-cell classifier and represent valuable tools to study the cancer-stroma-immune interface, which we used to identify predictors of survival in melanoma patients from conventional H&E stained histopathology.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.

AUTHOR CONTRIBUTIONS

KZ-P, HF, SR, IR, YJ, and YY: substantial contributions to the conception or design of the work; or the method development, analysis, or interpretation of data for the work; drafting the work or revising it critically for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

Cancer Research UK to the Cancer Imaging Centre at ICR, in association with the MRC and Department of Health (England) (C1060/A16464), NHS funding to the NIHR Biomedicine Research Centre and the Clinical Research Facility in Imaging, The Rosetrees Trust (KZ-P). YY acknowledges support by Cancer Research UK (C45982/A21808), Breast Cancer Now (2015NovPR638) and was also supported in part by the Wellcome Trust (105104/Z/14/Z). YJ was a Children with Cancer UK Research Fellow.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01045/full#supplementary-material

REFERENCES

1. Gurcan MN, Boucheron L, Can A, Madabhushi A, Rajpoot N, Yener B. Histopathological image analysis: a review. *IEEE Rev. Biomed. Eng.* (2009) 2:147. doi: 10.1109/RBME.2009.2034865
2. Kothari S, Phan JH, Stokes TH, Wang MD. Pathology imaging informatics for quantitative analysis of whole-slide images. *J Am Med Inform Assoc.* (2013) 20:1099–108. doi: 10.1136/amiajnl-2012-001540
3. Jones TR, Kang IH, Wheeler DB, Lindquist RA, Papallo A, Sabatini DM, et al. CellProfiler analyst: data exploration and analysis software for complex image-based screens. *BMC Bioinf.* 9:482. doi: 10.1186/1471-2105-9-482
4. Yuan Y, Failmezger H, Rueda OM, Ali HR, Gräf S, Chin SF, et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci. Trans. Med*. (2012) 4:157ra143. doi: 10.1126/scitranslmed.3004330
5. Chen CL, Mahjoubfar A, Tai L-C, Blaby IK, Huang A, Niazi KR, et al. Deep learning in label-free cell classification. *Sci Rep.* (2016) 6:21471. doi: 10.1038/srep21471
6. Sirinukunwattana K, Raza SEA, Tsang Y-W, Snead DR, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans Med Imag.* (2016) 35:1196–206. doi: 10.1109/TMI.2016.2525803
7. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep.* 7:16878. doi: 10.1038/s41598-017-17204-5
8. Khoshdeli M, Cong R, Parvin B. Detection of nuclei in H&E stained sections using convolutional neural networks. In: *Biomedical and Health Informatics (BHI), 2017 IEEE EMBS International Conference on*. Orlando, FL: IEEE (2017). p. 105–8.
9. Piccinini F, Balassa T, Szkalisity A, Molnar C, Paavolainen L, Kujala K, et al. Advanced cell classifier: user-friendly machine-learning-based software for discovering phenotypes in high-content imaging data. *Cell Syst*. (2017) 4:651–5.e655. doi: 10.1016/j.cels.2017.05.012
10. Held M, Schmitz MH, Fischer B, Walter T, Neumann B, Olma MH, et al. CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. *Nat Method*. (2010) 7:747–54. doi: 10.1038/nmeth.1486
11. Zhong Q, Busetto AG, Fededa JP, Buhmann JM, Gerlich DW. Unsupervised modeling of cell morphology dynamics for time-lapse microscopy. *Nat Methods.* (2012) 9:711–3. doi: 10.1038/nmeth.2046
12. Failmezger H, Frohlich H, Tresch A. Unsupervised automated high throughput phenotyping of RNAi time-lapse movies. *BMC Bioinf.* (2013) 14:292. doi: 10.1186/1471-2105-14-292
13. Failmezger H, Praveen P, Tresch A, Frohlich H. Learning gene network structure from time laps cell imaging in RNAi Knock downs. *Bioinformatics*. (2013) 29:1534–40. doi: 10.1093/bioinformatics/btt179
14. Niederberger T, Failmezger H, Uskat D, Poron D, Glauche I, Scherf N, et al. Factor graph analysis of live cell-imaging data reveals mechanisms of cell fate decisions. *Bioinformatics.* (2015) 31:1816–23. doi: 10.1093/bioinformatics/btv040
15. Failmezger H, Dursun E, Dumcke S, Endele M, Poron D, Schroeder T, et al. Clustering of samples with a tree-shaped dependence structure, with an application to microscopic time lapse imaging. *Bioinformatics.* (2018). doi: 10.1093/bioinformatics/bty939
16. Karimaghaloo Z, Arnold DL, Arbel T. Adaptive multi-level conditional random fields for detection and segmentation of small enhanced pathology in medical images. *Med Image Anal.* (2016) 27:17–30. doi: 10.1016/j.media.2015.06.004
17. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans Pattern Anal Machine Intell.* (2018) 40:834–48. doi: 10.1109/TPAMI.2017.2699184
18. Li Y, Ping W. Cancer metastasis detection with neural conditional random field. In: *1st Conference on Medical Imaging With Deep Learning (MIDL).* Amsterdam (2018).
19. Zanjani FG, Zinger S. Cancer detection in histopathology whole-slide images using conditional random fields on deep embedded spaces. In: *Medical Imaging 2018: Digital Pathology*. Houston, TX: International Society for Optics and Photonics (2018). p. 105810I.
20. Rajapakse JC, Liu S. Staging tissues with conditional random fields. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Boston, MA: IEEE (2011). p. 5128–31. doi: 10.1109/IEMBS.2011.6091270
21. Li C, Chen H, Xue D, Hu Z, Zhang L, He L, et al. Weakly supervised cervical histopathological image classification using multilayer hidden conditional random fields. In: *International Conference on Information Technologies in Biomedicine*. Kamień Śląski: Springer (2019). p. 209–21. doi: 10.1007/978-3-030-23762-2_19
22. Paramanandam M, O'Byrne M, Ghosh B, Mammen JJ, Manipadam MT, Thamburaj R, et al. Automated segmentation of nuclei in breast cancer histopathology images. *PLoS ONE.* 11:e0162053. doi: 10.1371/journal.pone.0162053
23. Zormpas-Petridis K, Failmezger H, Roxanis I, Blackledge M, Jamin Y, Yuan Y. Capturing global spatial context for accurate cell classification in skin cancer histology. *Comput Pathol Ophth Med Image Anal*. (2018) 11039:52–60. doi: 10.1007/978-3-030-00949-6_7
24. Reinhard E, Adhikhmin M, Gooch B, Shirley P. Color transfer between images. *IEEE Comp Grap Appl.* (2001) 21:34–41. doi: 10.1109/38.946629
25. Khan AM, Rajpoot N, Treanor D, Magee D. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Trans Biomed Eng.* (2014) 61:1729–38. doi: 10.1109/TBME.2014.2303294
26. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell.* (2012) 34:2274–82. doi: 10.1109/TPAMI.2012.120
27. Haralick RM, Shanmugam K. Textural features for image classification. *IEEE Trans Syst Man Cyber.* (1973) 3:610–21. doi: 10.1109/TSMC.1973.4309314
28. Imran M, Hashim R, Khalid NEA. Segmentation-based fractal texture analysis and color layout descriptor for content based image retrieval. In: *Intelligent Systems Design and Applications (ISDA), 2014 14th International Conference on IEEE.* Okinawa (2014). p. 30–3.
29. Lafferty J, McCallum A, Pereira FC. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001).* Williamstown, MA: Morgan Kaufmann Publishers (2001). p. 282–9.
30. Schmidt M. *UGM: A Matlab Toolbox for Probabilistic Undirected Graphical Models.* (2007). Available online at: http://www.cs.ubc.ca/~schmidtm/Software/UGM.html
31. de Kruijf EM, van Nes JG, van de Velde CJ, Putter H, Smit VT, Liefers GJ, et al. Tumor-stroma ratio in the primary tumor is a prognostic factor in

early breast cancer patients, especially in triple-negative carcinoma patients. *Breast Cancer Res Treat.* (2011) 125:687–96. doi: 10.1007/s10549-010-0855-6
32. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013
33. Wu J, Liang C, Chen M, Su W. Association between tumor-stroma ratio and prognosis in solid tumor patients: a systematic review and meta-analysis. *Oncotarget.* (2016) 7:68954–65. doi: 10.18632/oncotarget.12135
34. Scheer R, Baidoshvili A, Zoidze S, Elferink MAG, Berkel AEM, Klaase JM, et al. Tumor-stroma ratio as prognostic factor for survival in rectal adenocarcinoma: a retrospective cohort study. *World J Gastrointest Oncol.* (2017) 9:466–74. doi: 10.4251/wjgo.v9.i12.466
35. Barnes TA, Amir E. HYPE or HOPE: the prognostic value of infiltrating immune cells in cancer. *Br J Cancer.* (2017) 117:451–60. doi: 10.1038/bjc.2017.220
36. Kim JY, Gatenby RA. Quantitative clinical imaging methods for monitoring intratumoral evolution. *Methods Mol Biol.* (2017) 1513:61–81. doi: 10.1007/978-1-4939-6539-7_6
37. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Semantic image segmentation with deep convolutional nets and fully connected CRFS. In: *3rd International Conference on Learning Representations, ICLR 2015.* San Diego, CA (2014).
38. Kokkinos I. Ubernet: training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Honolulu, HI (2017). p. 6129–38.
39. Arnab A, Zheng S, Jayasumana S, Romera-Paredes B, Larsson M, Kirillov A, et al. Conditional random fields meet deep neural networks for semantic segmentation. *IEEE Signal Process Mag.* 35:37–52. doi: 10.1109/MSP.2017.2762355
40. Qin Y, Kamnitsas K, Ancha S, Nanavati J, Cottrell G, Criminisi A, et al. Autofocus layer for semantic segmentation. *arXiv:1805.08403* (2018). doi: 10.1007/978-3-030-00931-1_69
41. Roy AG, Navab N, Wachinger C. Concurrent spatial and channel 'Squeeze & Excitation' in fully convolutional networks. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018. Lecture Notes in Computer Science, Vol 11070.* Cham: Springer (2018). doi: 10.1007/978-3-030-00928-1_48
42. Raza SEA, Cheung L, Shaban M, Graham S, Epstein D, Pelengaris S, et al. Micro-Net: a unified model for segmentation of various objects in microscopy images. *Med. Image Anal.* (2019) 52:160–73. doi: 10.1016/j.media.2018.12.003
43. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2016). p. 2818–26.
44. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In*: Thirty-First AAAI Conference on Artificial Intelligence.* San Francisco, CA (2017).
45. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *3rd International Conference on Learning Representations, ICLR 2015.* San Diego, CA (2014).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 15 October 2019
doi: 10.3389/fonc.2019.01064

Radiomic Nomogram Improves Preoperative T Category Accuracy in Locally Advanced Laryngeal Carcinoma

Fei Wang[1†], Bin Zhang[1,2†], Xiangjun Wu[3,4†], Lizhi Liu[5†], Jin Fang[1], Qiuying Chen[1,2], Minmin Li[1,2], Zhuozhi Chen[1,2], Yueyue Li[1], Di Dong[3,4], Jie Tian[3,4,6*] and Shuixing Zhang[1*]*

[1] Department of Radiology, The First Affiliated Hospital, Jinan University, Guangzhou, China, [2] First Clinical Medical College, Jinan University, Guangzhou, China, [3] CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China, [4] College of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China, [5] State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Sun Yat-sen University Cancer Center, Guangzhou, China, [6] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing, China

Surgical decision-making on advanced laryngeal carcinoma is heavily depended on the identification of preoperative T category (T3 vs. T4), which is challenging for surgeons. A T category prediction radiomics (TCPR) model would be helpful for subsequent surgery. A total of 211 patients with locally advanced laryngeal cancer who had undergone total laryngectomy were randomly classified into the training cohort (n = 150) and the validation cohort (n = 61). We extracted 1,390 radiomic features from the contrast-enhanced computed tomography images. Interclass correlation coefficient and the least absolute shrinkage and selection operator (LASSO) analyses were performed to select features associated with pathology-confirmed T category. Eight radiomic features were found associated with preoperative T category. The radiomic signature was constructed by Support Vector Machine algorithm with the radiomic features. We developed a nomogram incorporating radiomic signature and T category reported by experienced radiologists. The performance of the model was evaluated by the area under the curve (AUC). The T category reported by radiologists achieved an AUC of 0.775 (95% CI: 0.667–0.883); while the radiomic signature yielded a significantly higher AUC of 0.862 (95% CI: 0.772–0.952). The predictive performance of the nomogram incorporating radiomic signature and T category reported by radiologists further improved, with an AUC of 0.892 (95% CI: 0.811–0.974). Consequently, for locally advanced laryngeal cancer, the TCPR model incorporating radiomic signature and T category reported by experienced radiologists have great potential to be applied for individual accurate preoperative T category. The TCPR model may benefit decision-making regarding total laryngectomy or larynx-preserving treatment.

Keywords: advanced laryngeal cancer, computed tomography, radiomics, T category, nomogram

OPEN ACCESS

Edited by:
Lei Deng,
Jacobi Medical Center, United States

Reviewed by:
Zaiyi Liu,
Guangdong Provincial People's Hospital, China
Jinhua Yu,
Fudan University, China

***Correspondence:**
Di Dong
di.dong@ia.ac.cn
Jie Tian
jie.tian@ia.ac.cn
Shuixing Zhang
shui7515@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *22 August 2019*
Accepted: *30 September 2019*
Published: *15 October 2019*

Citation:
Wang F, Zhang B, Wu X, Liu L, Fang J, Chen Q, Li M, Chen Z, Li Y, Dong D, Tian J and Zhang S (2019) Radiomic Nomogram Improves Preoperative T Category Accuracy in Locally Advanced Laryngeal Carcinoma. Front. Oncol. 9:1064. doi: 10.3389/fonc.2019.01064

BACKGROUND

Laryngeal cancer is a common malignant tumor in the head and neck and occurs mainly in smoking men (1). A study by the International Agency for Research on Cancer showed that 177,422 new laryngeal cancer cases occurred and resulted in 74,771 cancer-related deaths in 2018 (2).

The cancer control and functional outcomes of laryngeal cancer patients are highly relied on the treatment strategy. However, the management of laryngeal cancer remains controversial to date (3, 4). Currently, total laryngectomy is considered the most appropriate therapy for patients with advanced laryngeal carcinoma because they usually have a poor prognosis. Although total laryngectomy helps disease control, it has obvious adverse effects on patients' quality of life due to the loss of voice, permanent tracheostomy and issues with swallowing. In respect of which, Larynx-preserving surgery was thus performed to preserve laryngeal function (5). Decision-making about surgery are highly relied on tumor T category pursuant to the newest National Comprehensive Cancer Network (NCCN) Guidelines. The guidelines recommends total laryngectomy for all T4 stage and most of T3 stage laryngeal cancers, while some T3 stage diseases can benefit from larynx-preserving surgery instead (6).

Usually, the distinction between T3 and T4 categories is mainly based on the destruction degree of the extra-laryngeal spread and/or outer cortex of thyroid cartilage (7). However, accurate preoperative T category is clinically challenging. Currently, the most commonly used imaging techniques for T category (T3 vs. T4) are conventional imaging techniques including CT and MRI. CT generally demonstrates higher specificity but lower sensitivity as compared with MRI when identifying thyroid cartilage invasion (8). Although CT is useful in assessing the extent of extra-laryngeal spread or thyroid cartilage penetration of tumor, it has obvious limitations. Beitler et al. showed 74 and 81% positive predictive value of CT for assessing the extent of thyroid cartilage invasion and extra-laryngeal spread, respectively (9). However, Li et al. indicated that CT was less useful for assessing full-thickness cartilage invasion, with 47% of T4 disease being down-staged to T3 disease after pathological review (10). In contrast, MRI is more sensitive than CT in detecting cartilage invasion, yet peritumoral inflammation, edema and fibrosis may demonstrate similar features with cartilage invasion (11). These findings indicated the difficulty of accurate T category before surgery. Therefore, to develop new non-invasive methods for preoperative evaluation are needed for the purpose of determining the extent of extra-laryngeal spread and thyroid cartilage penetration, which are the most important considerations for selecting total laryngectomy or larynx conservation.

In recent years, the proposed "radiomics" is developed rapidly and has attracted great attention. It aims to extract huge amounts of objective features from medical images and find out the significant features which have great potential to expose disease characteristics that failed to be discovered by naked eyes (12–15). Previous studies showed that radiomic signatures as biomarkers have close correlations with clinical stages, lymph node metastasis, and survival outcomes (16–19). As there is no study explored whether radiomics would enhance the accuracy of preoperative T category for patients with advanced laryngeal cancer, we tried to explore CT-based TCPR as a novel approach for individual accurate preoperative T category for those patients, which would benefit clinical decision-making (total laryngectomy or larynx conservation) before surgery.

Abbreviations: TCPR, T category prediction radiomics; LASSO, least absolute shrinkage and selection operator; AUC, area under the curve; ICC, interclass correlation coefficient; ROC, receiver operating characteristic; SVM, support vector machine.

MATERIALS AND METHODS

Patient Population

This retrospective study was approved by the Institutional Review Board and the informed consent requirement was waived. The whole cohort of this study was acquired from the medical records of the Institutional database from April 2007 to March 2015. Patients with histologically confirmed laryngeal cancer who had received total laryngectomy were included. Contrast-enhanced CT examinations of the neck had been performed on all patients before surgery. The inclusion criteria were as follows: (1) newly diagnosed patients underwent contrast-enhanced CT scans of neck before any treatment; (2) patients received total laryngectomy 15 days after initial CT acquisition; and (3) patients had pathologically confirmed T3 or T4 stage laryngeal cancer after operation. The exclusion criteria were as follows: (1) poor quality of CT images due to patients' movement or artifacts, etc.; (2) the slice thickness of CT scan >2.5 mm; and (3) patients received treatment.

A total of 211 patients met these criteria. Among which, 150 patients constituted the training cohort, including 146 males and four females with mean age of 61.38 ± 8.54 ranging from 39 to 85. A total of 61 patients (59 males, two females) with mean age of 60.23 ± 6.65 ranging from 30 to 78 were allocated to the validation cohort.

Clinicopathologic data was collected from the medical records and the data of baseline CT scans, including age, gender, preoperative T category reported by head and neck radiologists, and pathologically confirmed T category. T classification was conducted pursuant to the 8th Edition of AJCC TNM Staging System Guidelines (20), and then reassessed by a head and neck radiologist with 20 years of experience who was blinded to the pathology results. **Figure 1** showed the workflow of radiomic analysis in the current study.

CT Image Acquisition and Tumor Segmentation

Two CT systems were adopted for CT image acquisition: United Imaging uCT780 and Siemens SOMATOM Force CT. The parameters for CT image acquisition were as follows: 110–120 kV; 116–168 mAs; detector collimation: 192×0.6 mm or 160×0.25 mm; rotation time: 1.0 s; slice thickness: 1–2.5 mm; field of view: 250×250 mm; matrix: 512×512.

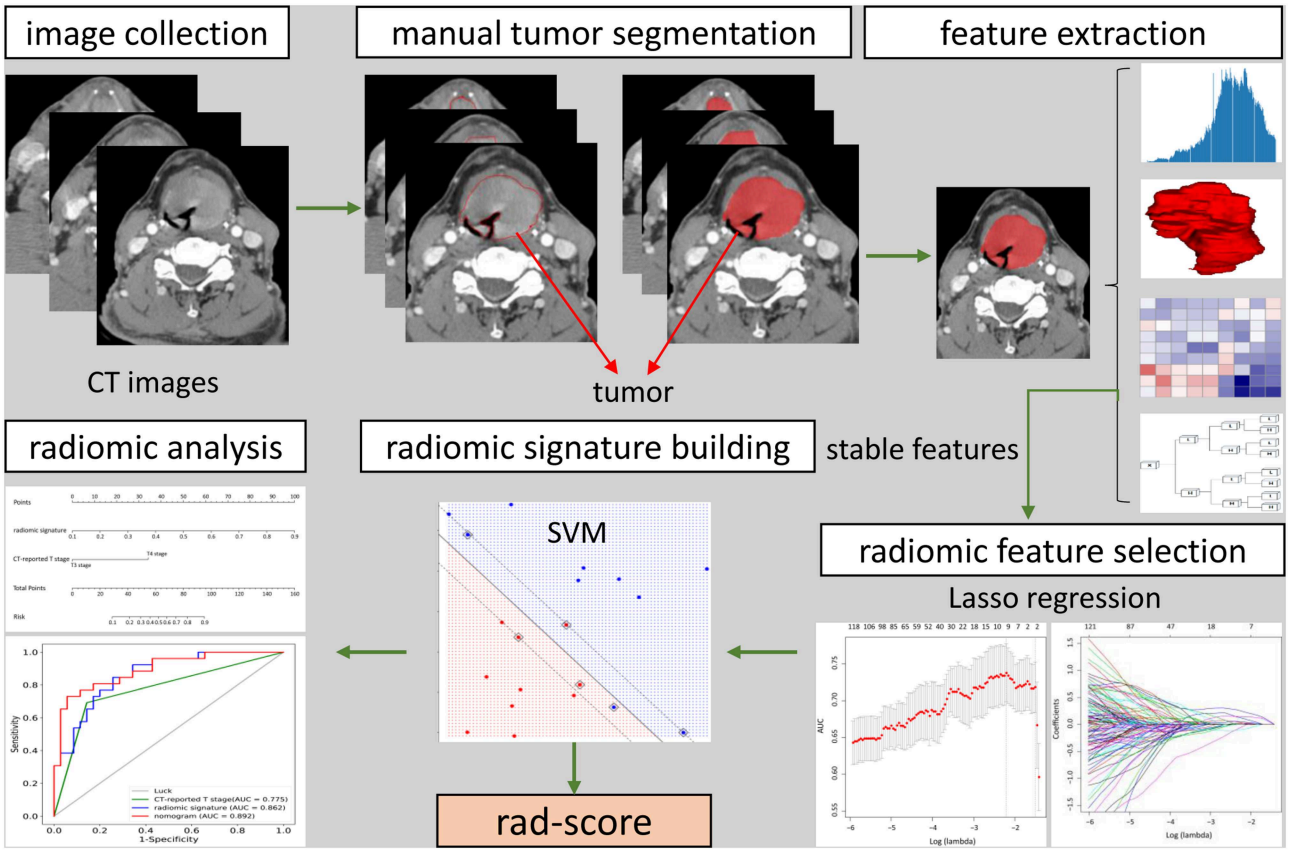


FIGURE 1 | The workflow of radiomic analysis in the current study. After feature extraction, stable features were selected by LASSO for further analysis. SVM model was used to build radiomic signature. The predictive nomogram was constructed based on the radiomic signature and other predictors.

Axial venous phase CT images (DICOM format) were prepared for tumor segmentation. An open-source software ITK-SNAP (www.itk-snap.org) was applied for the three dimensional manual segmentation. Tumor region in each layer was outlined by a radiologist with 12 years of experience in head and neck cancer, and then validated by a senior radiologist with 20 years of experience in head and neck cancer. The regions of interest covering the entire tumor were used for subsequent feature extraction.

Radiomic Features Extraction and Radiomic Signature Construction

Radiomic features were extracted by Pyradiomics (version 2.1.2), an open-source python platform (http://www.radiomics.io/pyradiomics.html) (21). Pyradiomics provides a stably operated open-source platform for easy and reproducible radiomic features extraction that can be compared across different institutions. Features of high throughput were extracted from CT images by matrix operation, wavelet transform and other mathematical methods, whose purpose was to find out the association between radiomic features and pathologically confirmed T category. The extracted radiomic features were classified into 4 categories: first-order features ($n = 126$), textural features ($n = 515$), shape-based features ($n = 13$), wavelet features ($n = 736$). In order to identify the most significant features, we used the interclass correlation coefficient and least absolute shrinkage and selection operator (ICC-LASSO) to remove abundant high dimensional features. Only features with an ICC > 0.75 were retained for further LASSO, while the remaining radiomic features were excluded to ensure the stability and reproducibility. After that, the most significant features were used to build the support vector machine (SVM) machine learning prediction model. Grid search and cross validation were conducted to select model parameters, which optimize the performance of the model. Then, radiomic signature was obtained from the trained SVM model.

Diagnostic Validation of Radiomic Signature

We used AUC, sensitivity, specificity, and accuracy to verify the association between radiomic signature and pathologically confirmed T category in order to determine the overall performance of the model. The performance of radiomic signature was established in the training cohort and internally validated in the validation cohort.

Development of an Individualized T Category Prediction Nomogram

Univariate analysis was performed on clinical features, such as age, gender, tumor location, and T category reported by an experienced radiologist. The most significant clinical features and radiomic signature were combined to establish a multivariable logistics model so as to develop a radiomic nomogram.

Validation of the Radiomic Nomogram

The utility of the radiomic nomogram in the training and validation cohorts was assessed by the receiver operating characteristic (ROC) curves. The ROC curve was plotted basing on the predictors of multi-logistics model including AUC, sensitivity, specificity, and accuracy. In addition, we plotted the calibration curves and conducted the Hosmer-Lemeshow test to demonstrate the calibration of the radiomic nomogram.

Statistical Analysis

Continuous data were presented as mean ± standard deviation (SD), while categorical data were presented as counts and percentages. Continuous and categorical data were compared by independent t (or Mann-Whitney U) test and Chi-square (or Fisher's exact) statistics, respectively. Patients were randomly divided into the training and validation cohorts at a ratio of ∼2.5:1. The average performance of the model was obtained by bootstrapping for 2,000 times. All statistical analyses were conducted by R software (version 3.5.1) and Python (version 3.6). The R software was used for features selection and building nomogram with packages of "psych," "glmnet," and "rms," while the Python was used to build SVM model with "sklearn" package.

RESULTS

Clinical Characteristics

Table 1 summarizes the patient characteristics of the training and validation cohorts. Only T category reported by radiologists showed significant difference ($P < 0.001$). After pathological review, 20.5% (17/83) of patients down-staged from T4 to T3, and 28.9% (37/128) of patients over-staged from T3 to T4.

Radiomic Features Extraction and Radiomic Signature Construction

We extracted 1,390 features in total from CT images, among which, 565 had ICC > 0.75, which indicted a good inter-measurer agreement. LASSO was then used to remove the redundancy of high dimensional features, and eight significant radiomic features were selected at last (**Figure 2**), including two first order features (gradient_first order_Skewness, lbp.2D_first order_Mean), two shape features (original_shape_LeastAxis, original_shape_Sphericity), and four wavelet features (wavelet-LLH_first order_Kurtosis, wavelet-LLH_glcm_Idn, wavelet-LLH_first order_Median, wavelet-LLL_glcm_Imc1). A SVM-based radiomic signature was constructed based on the eight features.

Diagnostic Validation of Radiomic Signature

The AUCs of radiomic signature were 0.850 (95% CI: 0.788–0.912) and 0.862 (95% CI: 0.772–0.952) in the training and validation cohorts, respectively (**Table 2**). Correspondingly, the specificity were 0.792 (95% CI: 0.698–0.885) and 0.743 (0.598–0.888); the sensitivity were 0.782 (95% CI: 0.690–0.874) and 0.808 (95% CI: 0.656–0.959); and the accuracy were 0.787 (95% CI: 0.784–0.789) and 0.770 (95% CI: 0.765–0.776) (**Table 2**).

Development of an Individualized T Category Prediction Nomogram

Logistic regression analysis of clinical features demonstrated that only T category reported by experienced radiologist was significantly correlated with pathologically confirmed T category ($p < 0.001$). Radiomic nomogram was established by combining radiomic signature and T category reported by radiologists (**Figure 3A**). The calibration curves of nomogram showed a good agreement between prediction and observation in both of the training and validation cohorts (**Figures 3B,C**).

Validation of the Radiomic Nomogram

In the training cohort, the AUC of T category reported by radiologists was 0.751 (95% CI: 0.684–0.818), with specificity of 0.861 (95% CI: 0.781–0.941), sensitivity of 0.641 (95% CI: 0.535–0.747), and accuracy of 0.747 (95% CI: 0.744–0.749) (**Table 2**). The AUC of the combined nomogram incorporating radiomic signature and T category reported by radiologists was 0.899 (95% CI: 0.850–0.947), with sensitivity of 0.782 (95% CI: 0.690–0.874), specificity of 0.889 (95% CI: 0.816–0.961), and accuracy of 0.833 (95% CI: 0.832–0.835) (**Table 2**).

In the validation cohort, the AUC of T category reported by radiologists was 0.775 (95% CI: 0.667–0.883) with specificity of 0.857 (95% CI: 0.741–0.973), sensitivity of 0.692 (95% CI: 0.515–0.870), and accuracy of 0.787 (95% CI: 0.781–0.792) (**Table 2**, **Figure 4**). The AUC of the nomogram incorporating radiomic signature and T category reported by radiologists was 0.892 (95% CI: 0.811–0.974), with sensitivity of 0.808 (95% CI: 0.656–0.959), specificity of 0.771 (95% CI: 0.632–0.911), and accuracy of 0.787 (95% CI: 0.781–0.792) (**Table 2**, **Figure 4**).

DISCUSSION

This study developed and validated a radiomic nomogram for the accurate prediction of T category (T3 vs. T4) before surgery for patients with locally advanced laryngeal cancer. The combined nomogram incorporated the CT-reported T stage and the radiomic signature. By only CT, radiologists couldn't satisfactorily stratified patients into T3 and T4 categories (AUC = 0.775). However, the combination of the radiomic signature and the T category reported by radiologists could significantly improve the predictive performance, achieving an AUC of 0.892 in the validation cohort.

Locally advanced laryngeal cancer includes those classified as T3 or T4 category (22). For locally advanced laryngeal cancer, the treatment option of total laryngectomy or organ preservation

TABLE 1 | Patient characteristics in the training and validation cohorts.

Characteristics	Training cohort ($n = 150$)			Validation cohort ($n = 61$)		
	T3 category	T4 category	*P*	T3 category	T4 category	*P*
Gender, No (%)			0.343			0.501
Male	70 (95.9%)	76 (98.7%)		33 (94.3%)	26 (100%)	
Female	3 (4.1%)	1 (1.3%)		2 (5.7%)	0	
Age, mean ± SD, years	61.38 ± 8.54	63.72 ± 8.97	0.157	60.23 ± 6.65	60.31 ± 10.91	0.737
Location, No (%)			0.022			0.579
Supra-glottis	31 (42.5%)	21 (27.3%)		11 (31.4%)	10 (38.5%)	
Glottis	40 (54.8%)	56 (72.7%)		24 (68.6%)	16 (61.5%)	
Sub-glottis	2(2.7%)	0		0	0	
T category reported by radiologist, No (%)			<0.001			<0.001
T3 category	61 (83.6%)	29 (37.7%)		30 (85.7%)	8 (30.8%)	
T4 category	12 (16.4%)	48 (62.3%)		5 (14.3%)	18 (69.2%)	

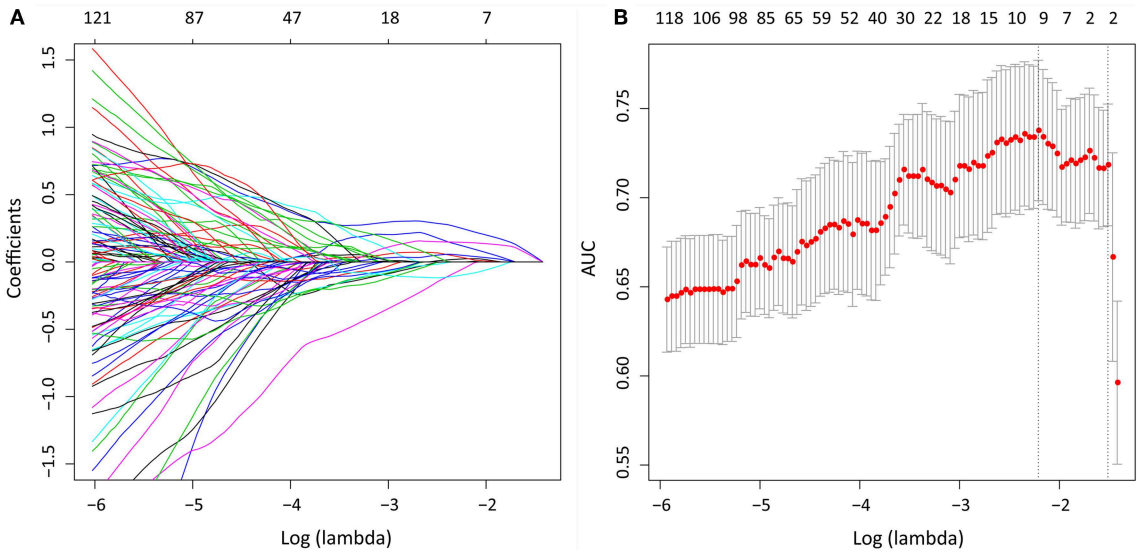


FIGURE 2 | After initial screening by ICC analysis, feature selection was performed using the LASSO method with a logistic regression model. **(A)** The model coefficient trendlines of the 1,390 radiomics features. The profile graph was plotted by coefficients against the L1 norm (inverse proportional to log $\lambda = -2.184$). **(B)** Tuning parameter λ in the LASSO model. The parameter $\lambda = 0.220$ were selected under the minimum criteria. The vertical line was drawn at the value selected by 10-fold cross-validation, including optimized eight non-zero coefficients.

remains to be a hot-debated topic. The goal of larynx preservation is to achieve good function without altering patients' survival. When determining larynx preservation or total laryngectomy for patients, some issues must be considered, such as T category of tumor, patients' will, and prospects for a good functional outcome (23). Therefore, preoperative T category is particularly important. If we could distinguish T3 from T4 patients with laryngeal cancer, they can receive appropriate treatment and benefit a lot. This study is focused on patients with local advanced laryngeal cancer and tried to find out a new method to distinguish T3 from T4 disease accurately.

Previous studies demonstrated that CT, MRI, PET-CT images can reflect the tumor morphology (24–26). Clinicians rely on medical imaging to determine whether patients suffered from T3 or T4 disease. Reliable imaging tools are indispensable. CT is the preferred imaging method for laryngeal cancer staging (11), which is much faster than MRI. MRI has better discrimination of soft tissue changes and cartilage abnormalities, however, it requires longer image acquisition time, thus challenging patients' cooperation and hampering its utilization (27). Still, it is very important for imaging techniques being able to differentiate inner cortical invasion (T3) from destruction of the outer cortex and extra-laryngeal spread. The evaluation on thyroid cartilage invasion and extra-laryngeal spread is important and sometimes difficult, and the positive predictive value of CT-reported T category is 71.1%, similar with Li et al. (10). MRI seems to be more sensitive than CT in detecting cartilage invasion, however, the MRI findings are not specific, and the positive predictive value of MRI was unsatisfactory (9). This is because that peritumoral inflammation, edema and fibrosis may demonstrate

TABLE 2 | Diagnostic performance of models in the training and validation cohorts.

Models	Training cohort ($n = 150$)				Validation cohort ($n = 61$)			
	AUC (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Accuracy (95% CI)	AUC (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Accuracy (95% CI)
T category reported by radiologist	0.751 (0.684–0.818)	0.861 (0.781–0.941)	0.641 (0.535–0.747)	0.747 (0.744–0.749)	0.775 (0.667–0.883)	0.857 (0.741–0.973)	0.692 (0.515–0.870)	0.787 (0.781–0.792)
Radiomic signature	0.850 (0.788–0.912)	0.792 (0.698–0.885)	0.782 (0.690–0.874)	0.787 (0.784–0.789)	0.862 (0.772–0.952)	0.743 (0.598–0.888)	0.808 (0.656–0.959)	0.770 (0.765–0.776)
Combined nomogram	0.899 (0.850–0.947)	0.889 (0.816–0.961)	0.782 (0.690–0.874)	0.833 (0.832–0.835)	0.892 (0.811–0.974)	0.771 (0.632–0.911)	0.808 (0.656–0.959)	0.787 (0.781–0.792)

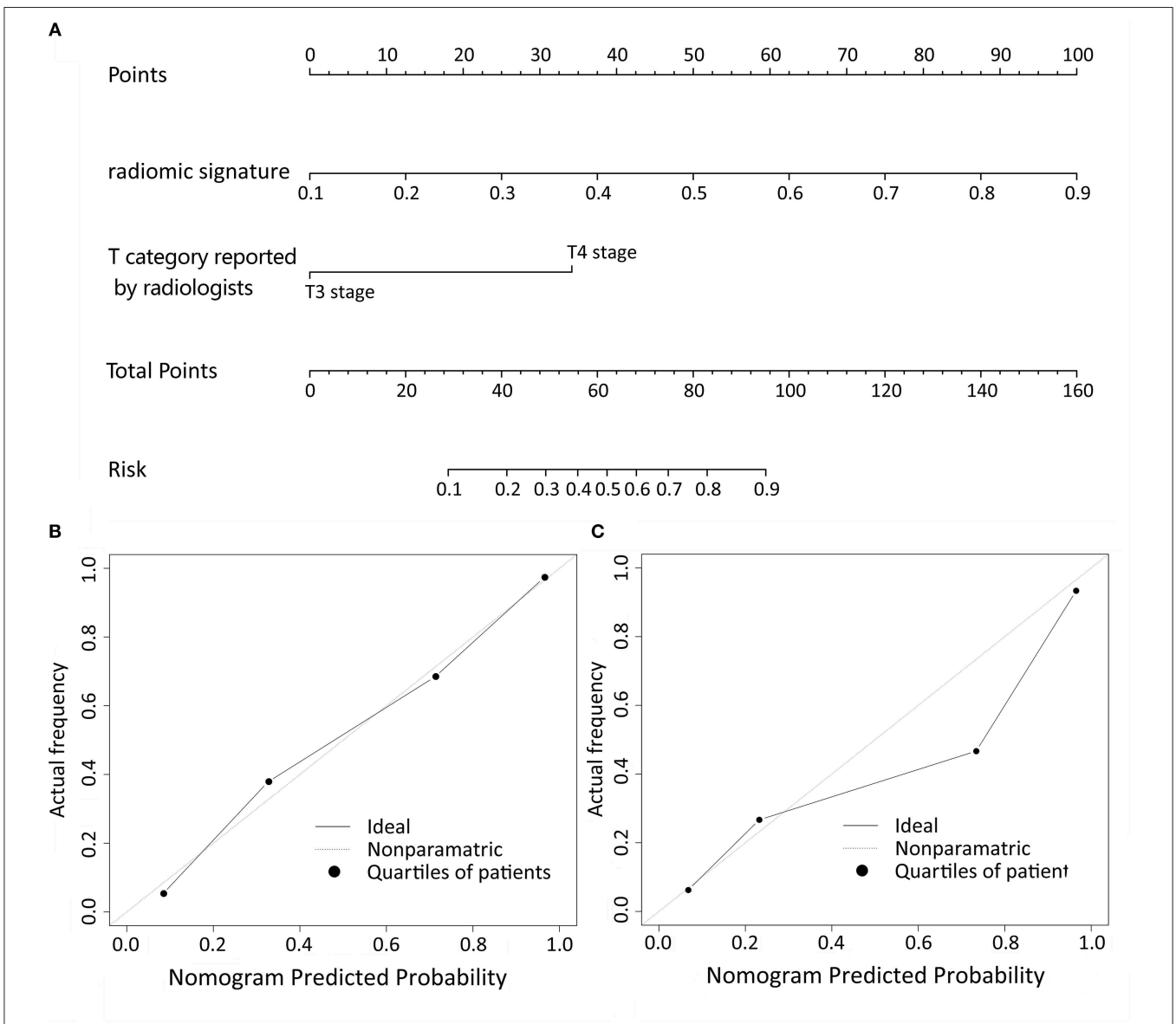


FIGURE 3 | The nomogram of T category diagnostic model. Our radiomics based nomogram was constructed in the training cohort. The radiomic signature, T category reported by radiologist were incorporated as factors **(A)**. The calibration curves showed good agreement between the nomogram-predicted T category and actual T category in the training cohort **(B)** and validation cohort **(C)**.

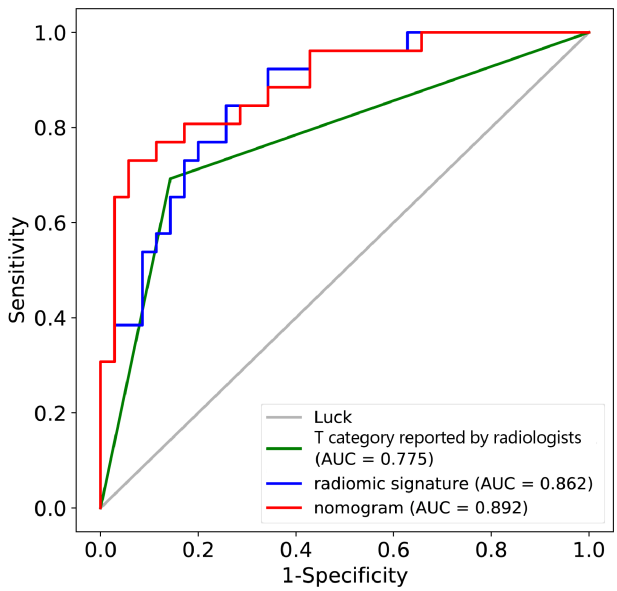


FIGURE 4 | ROC curves for the nomogram, radiomic signature, and T category reported by radiologist in the training and validation datasets.

similar features with cartilage invasion (11). Currently, the guidelines recommends total laryngectomy for patients with T4 stage diseases, while for T3 stage diseases, organ preservation or total laryngectomy are all listed options (6). It is indicated that some patients who were treated by total laryngectomy could have been offered laryngeal preservation or who received laryngeal preservation actually need total laryngectomy to extend the survival time if more accurately staging was performed at initial diagnosis (10, 28).

Radiomics is a new concept in recent years, and it is gaining importance in cancer research for improving diagnostic, prognostic, and predictive accuracy (29). Zhang et al. established and internally validated MRI-derived radiomics as a new approach to evaluate progression-free survival in patients with stage III–IVb nasopharyngeal carcinoma before treatment (30). Liang et al. demonstrated that a combined nomogram model could preoperatively predict histologic grade in patients with pancreatic neuroendocrine tumors (18). For patients with locally advanced laryngeal cancer, we identified a radiomic nomogram to perform preoperative predicting of tumor T category (T3 vs. T4). To construct a radiomic signature, 1,382 (99.4%) radiomic features were filtered, and only eight features were saved by ICC and LASSO analysis. For huge amounts of radiomic features extracted from a relatively small sample, LASSO can avoid model overfitting (31). In addition, the features selected by LASSO are generally accurate and can be easily interpreted because the vast majority of irrelevant features' coefficients are shrunk toward zero during model fitting. The radiomic signature was constructed through LASSO-SVM by combining radiomic features with other clinical features, such as age, gender, tumor location, T category reported by radiologists. Our study showed that preoperative T category reported by radiologists was easily obtained and significantly correlated with actual T category, and it might significantly influence the accuracy of the prediction of T category. Therefore, the radiomic nomogram incorporates both the T category reported by radiologists and the radiomic signature to ensure accuracy. The T category reported by radiologists can stratify patients into T3 and T4 groups with an AUC of 0.751. However, the combined nomogram model can further improve the predictive performance, achieving an AUC of 0.899. This prediction model was also tested by the validation cohort (AUC = 0.892), verifying the reliability and reproducibility.

The main limitation of this study derived from its retrospective nature. To keep the consistency of data, the training and validation cohorts were from a single institution. When determining the most suitable treatment strategy for advanced laryngeal cancer, preoperative T category is not the only factor under consideration, other conditions should also be considered, such as tumor volume, lymphatic metastasis, pre-treatment voice quality, laryngoscopy findings, patient comorbidities, and preferences.

In conclusion, this study established a TCPR model as a novel approach for the accurate preoperative T category for patients with locally advanced laryngeal cancer. As a non-invasive, preoperative and precise T category evaluation tool, the model could assist head and neck surgeons to make an appropriate surgical decision, which will benefit patients in the future.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The First Affiliated Hospital, Jinan University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

FW, BZ, and SZ: conception and design. FW, LL, JF, ML, ZC, YL, and QC: acquisition of data. XW, DD, and JT: analysis and interpretation of data. FW, BZ, DD, JT, and SZ: drafting or revising the article.

FUNDING

This work was supported by a grant of the National Natural Science Foundation of China (81571664, 81871323, 81801665, 81971776, 81771924); the National Natural Science Foundation of Guangdong Province (2018B030311024); the Scientific Research General Project of Guangzhou Science Technology and Innovation Commission (201707010328); and the China Postdoctoral Science Foundation (2016M600145); Beijing Natural Science Foundation (L182061).

REFERENCES

1. Steuer CE, El-Deiry M, Parks JR, Higgins KA, Saba NF. An update on larynx cancer. *CA Cancer J Clin.* (2017) 67:31–50. doi: 10.3322/caac.21386
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
3. Calkovsky V, Wallenfels P, Calkovska A, Hajtman A. Laryngeal cancer: 12-year experience of a single center. *Adv Exp Med Biol.* (2016) 911:9–16. doi: 10.1007/5584_2015_201
4. Rosenthal DI, Mohamed AS, Weber RS, Garden AS, Sevak PR, Kies MS, et al. Long-term outcomes after surgical or nonsurgical initial therapy for patients with T4 squamous cell carcinoma of the larynx: a 3-decade survey. *Cancer.* (2015) 121:1608–19. doi: 10.1002/cncr.29241
5. Lefebvre JL, Ang KK, Larynx Preservation Consensus Panel. Larynx preservation clinical trial design: key issues and recommendations-a consensus panel summary. *Int J Radiat Oncol Biol Phys.* (2009) 73:1293–303. doi: 10.1016/j.ijrobp.2008.10.047
6. Colevas AD, Yom SS, Pfister DG, Spencer S, Adelstein D, Adkins D, et al. NCCN guidelines insights: head and neck cancers, version 1.2018. *J Natl Compr Canc Netw.* (2018) 16:479–90. doi: 10.6004/jnccn.2018.0026
7. Timmermans AJ, Lange CA, de Bois JA, van Werkhoven E, Hamming-Vrieze O, Hilgers FJ, et al. Tumor volume as a prognostic factor for local control and overall survival in advanced larynx cancer. *Laryngoscope.* (2016) 126:E60–7. doi: 10.1002/lary.25567
8. Chu EA, Kim YJ. Laryngeal cancer: diagnosis and preoperative work-up. *Otolaryngol Clin North Am.* (2008) 41:673–95 doi: 10.1016/j.otc.2008.01.016
9. Beitler JJ, Muller S, Grist WJ, Corey A, Klein AM, Johns MM, et al. Prognostic accuracy of computed tomography findings for patients with laryngeal cancer undergoing laryngectomy. *J Clin Oncol.* (2010) 28:2318–22. doi: 10.1200/JCO.2009.24.7544
10. Li B, Bobinski M, Gandour-Edwards R, Farwell DG, Chen AM. Overstaging of cartilage invasion by multidetector CT scans for laryngeal cancer and its potential effect on the use of organ preservation. *Br J Radiol.* (2011) 84:64–9. doi: 10.1259/bjr/66700901
11. Kuno H, Onaya H, Fujii S, Ojiri H, Otani K, Satake M. Primary staging of laryngeal and hypopharyngeal cancer: CT, MR imaging and dual-energy CT. *Eur J Radiol.* (2014) 83:e23–35. doi: 10.1016/j.ejrad.2013.10.022
12. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging.* (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
13. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
15. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. *Expert Rev Precis Med Drug Dev.* (2016) 1:207–26. doi: 10.1080/23808993.2016.1164013
16. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a radiomic nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol.* (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
17. Jiang Y, Chen C, Xie J, Wang W, Zha X, Lv W, et al. Radiomic signature of computed tomography imaging for prediction of survival and chemotherapeutic benefits in gastric cancer. *EBioMedicine.* (2018) 36:171–82. doi: 10.1016/j.ebiom.2018.09.007
18. Liang W, Yang P, Huang R, Xu L, Wang J, Liu W, et al. A combined nomogram model to preoperatively predict histologic grade in pancreatic neuroendocrine tumors. *Clin Cancer Res.* (2019) 25:584–94. doi: 10.1158/1078-0432.CCR-18-1305
19. Shen C, Liu Z, Wang Z, Guo J, Zhang H, Wang Y, et al. Building CT radiomics based nomogram for preoperative esophageal cancer patients lymph node metastasis prediction. *Transl Oncol.* (2018) 11:815–24. doi: 10.1016/j.tranon.2018.04.005
20. Amin MB, Edge SB, Greene FL, Byrd F, Brookland DR, Washington RK, et al. *AJCC Cancer Staging Manual*. 8th ed. New York, NY: Springer (2017).
21. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
22. Eskander A, Blakaj DM, Dziegielewski PT. Decision making in advanced larynx cancer: an evidenced based review. *Oral Oncol.* (2018) 86:195–9. doi: 10.1016/j.oraloncology.2018.09.019
23. Corry J, Peters L, Kleid S, Rischin D. Larynx preservation for patients with locally advanced laryngeal cancer. *J Clin Oncol.* (2013) 31:840–4. doi: 10.1200/JCO.2012.46.9197
24. Blitz AM, Aygun N. Radiologic evaluation of larynx cancer. *Otolaryngol Clin North Am.* (2008) 41:697–713, vi. doi: 10.1016/j.otc.2008.01.015
25. Gilbert K, Dalley RW, Maronian N, Anzai Y. Staging of laryngeal cancer using 64-channel multidetector row CT: comparison of standard neck CT with dedicated breath-maneuver laryngeal CT. *AJNR Am J Neuroradiol.* (2010) 31:251–6. doi: 10.3174/ajnr.A1796
26. Ryu IS, Lee JH, Roh JL, Choi SH, Nam SY, Kim SY, et al. Clinical implication of computed tomography findings in patients with locally advanced squamous cell carcinoma of the larynx and hypopharynx. *Eur Arch Otorhinolaryngol.* (2015) 272:2939–45. doi: 10.1007/s00405-014-3249-2
27. Maroldi R, Ravanelli M, Farina D. Magnetic resonance for laryngeal cancer. *Curr Opin Otolaryngol Head Neck Surg.* (2014) 22:131–9. doi: 10.1097/MOO.0000000000000036
28. Laccourreye O, Malinvaud D, Ménard M, Consoli S, Giraud P, Bonfils P. Total laryngectomy or laryngeal preservation for advanced laryngeal cancer. Impact of the functional risk upon the patient's preferences. *Eur Ann Otorhinolaryngol Head Neck Dis.* (2014) 131:93–7. doi: 10.1016/j.anorl.2013.06.001
29. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
30. Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin Cancer Res.* (2017) 23:4259–69. doi: 10.1158/1078-0432.CCR-16-2910
31. Hepp T, Schmid M, Gefeller O, Waldmann E, Mayr A. Approaches to regularized regression–a comparison between gradient boosting and the Lasso. *Methods Inf Med.* (2016) 55:422–30. doi: 10.3414/ME16-01-0033

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 15 October 2019
doi: 10.3389/fonc.2019.01021

The Utility of Texture Analysis Based on Breast Magnetic Resonance Imaging in Differentiating Phyllodes Tumors From Fibroadenomas

Hui Mai[1,2†], Yifei Mao[3†], Tianfa Dong[1,2], Yu Tan[4], Xiaowei Huang[5], Songxin Wu[4], Shuting Huang[4], Xi Zhong[1], Yingwei Qiu[2], Liangping Luo[1] and Kuiming Jiang[4*]*

[1] Department of Medical Imaging, The First Affiliated Hospital of Jinan University, Guangzhou, China, [2] Department of Radiology, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [3] Department of Radiology, Shenzhen Maternity and Child Healthcare Hospital, Shenzhen, China, [4] Department of Radiology, Guangdong Women and Children Hospital, Guangzhou, China, [5] Paul C. Lauterbur Research Center for Biomedical Imaging, Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

OPEN ACCESS

Edited by:
Xuelei Ma,
Sichuan University, China

Reviewed by:
Laurence Gluch,
The Strathfield Breast Centre, Australia
Johanna Patricia Adevoso Canal,
University of the Philippines Manila, Philippines

***Correspondence:**
Liangping Luo
tluolp@jnu.edu.cn
Kuiming Jiang
kmjiang64@sina.com

[†] These authors have contributed equally to this work as co-first authors

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 15 June 2019
Accepted: 23 September 2019
Published: 15 October 2019

Citation:
Mai H, Mao Y, Dong T, Tan Y, Huang X, Wu S, Huang S, Zhong X, Qiu Y, Luo L and Jiang K (2019) The Utility of Texture Analysis Based on Breast Magnetic Resonance Imaging in Differentiating Phyllodes Tumors From Fibroadenomas. Front. Oncol. 9:1021. doi: 10.3389/fonc.2019.01021

Background: The preoperative diagnosis of phyllodes tumors (PTs) of the breast is critical to appropriate surgical treatment. However, reliable differentiation between PT and fibroadenoma (FA) remains difficult in daily clinical practice. The purpose of this study was to investigate the utility of breast MRI texture analysis for differentiating PTs from FAs.

Materials and Methods: Forty-two PTs and 42 FAs were enrolled in this retrospective study. Clinical and conventional MRI features (CCMF) and MRI texture analysis were used to distinguish between PT and FA. Texture features were extracted from the axial short TI inversion recovery T2-weighted (T2W-STIR), T1-weighted pre-contrast, and two contrast-enhanced series (first contrast and third contrast). The Mann–Whitney *U* test was used to select statistically significant features of texture analysis and CCMF. Using a linear discriminant analysis, the most discriminative features were determined from statistically significant features. The K-nearest neighbor classifier and ROC curve were applied to evaluate the diagnostic performance.

Results: With a higher classification accuracy (89.3%) and an AUC of 0.89, the texture features on T2W-STIR outperformed the texture features on other MRI sequences and CCMF. The AUC of the combination of CCMF with texture features on T2W-STIR was significantly higher than that of CCMF or texture features on T2W-STIR alone ($p < 0.05$). Based on the result of the classification accuracy (95.2%) and AUC (0.95), the diagnostic performance of the combination strategy performed better than texture features on T2W-STIR or CCMF separately.

Conclusions: Texture features on T2W-STIR showed better diagnostic performance compared to CCMF for the distinction between PTs and FAs. After further validation of multi-institutional large datasets, MRI-based texture features may become a potential biomarker and be a useful medical decision tool in clinical trials having patients with breast fibroepithelial neoplasms.

Keywords: texture analysis, breast, magnetic resonance imaging, phyllodes tumor, fibroadenoma

INTRODUCTION

Phyllodes tumor (PT) is a rare tumor accounting for 0.3–1.0% of all mammary tumors and comprises 2–3% of all fibroepithelial mammary neoplasms (1, 2). The histological classification is subdivided into benign, borderline, or malignant (3); however, histological type is found to poorly correlate with clinical behavior (4, 5). Incidence of local relapse is high regardless of the histological grading, and distant metastasis may occur in approximately 25% of malignant PTs (6, 7). With similar clinical features and histopathological appearance, PT may mimic a fibroadenoma (FA), which is the most common benign tumor of the breast. Sometimes, even preoperative invasive procedures such as fine-needle aspiration cytology and core needle biopsy may fail to correctly differentiate these two entities, primarily owing to lack of adequate and representative samples (8, 9). Given the different prognosis, a surgical excision is essential with a wide margin of at least 1 cm for all grades of PT to avoid local relapse and subsequent surgery (10, 11); on the other hand, a FA can usually be safely followed-up or managed by a simple enucleation (12). Therefore, accurate preoperative diagnosis is crucial to offer an appropriate clinical strategy, thus avoiding operative complications resulting from inadequate excision or surgical overtreatment.

Clinically, in contrast to FA, PT can grow rapidly to huge bulky ones with a high reported incidence of local relapse (13). In addition, PT was generally thought to develop later in life than FA (6, 14).

According to previous reports, MRI features have been valuable in the differentiation between PTs and FAs. Kamitani et al. (15) described the MRI features of PTs and noticed a pattern of heterogeneous enhancement, internal cystic components, and increased lobulations in PTs. Although certain clinical and MRI features may raise the index of suspicion, it is challenging to make a reliable differentiation between PT and FA. In daily clinical practice, a benign, small-sized borderline or malignant PT can be easily mistaken for a FA, whereas giant FAs may show overlapping MRI features of PTs.

Radiomics has drawn increasing attention in recent years. It is based on a hypothesis that medical imaging information can be converted into quantitative and mineable features via automatically high-throughput extraction of data characterization algorithms that in turn provide valuable diagnostic, prognostic, or predictive assessment (16–18). Several radiomics studies have shown that some quantitative imaging signatures, such as texture features derived from MRI, can provide an opportunity to facilitate better clinical decision-making in oncology at low cost and non-invasively. For example, texture analysis has been used to predict sentinel lymph node metastasis in breast cancer (19), differentiate estrogen receptor-positive breast cancer molecular subtypes (20), and identify healthy breast tissue and breast cancer lesions (21).

Abbreviations: PT, Phyllodes tumor; FA, Fibroadenoma; T2W-STIR, Short TI inversion recovery T2-weighted; CCMF, Clinical and conventional MRI features; TIC, time–intensity curve.

Thus, in the present study, we hypothesized that texture features on routine, enhancement, and non-enhanced T1- and T2-weighted MR images, could help to improve the differentiation between PTs and FAs.

MATERIALS AND METHODS

Patients

The retrospective study protocol was approved by our institutional review board. In this study, 53 female patients with histologically confirmed PT between June 2012 and June 2018 were enrolled and 78 female patients with histologically confirmed FA were randomly selected. The inclusion criteria were as follows: (1) female patients were histologically diagnosed with PT or FA by two experienced pathologists based on findings in the specimens obtained at surgical resection, (2) those who underwent breast MRI prior to surgical resection, and (3) those with lesions measuring >1 cm in diameter avoiding the possible unfavorable effects on textural features extracted from image data. The exclusion criteria were as follows: (1) a previous history of breast cancer and radiotherapy, and (2) poor image quality. Finally, 41 female patients with 42 PTs and 37 female patients with 42 FAs were eligible in this study.

MRI Acquisition

All patients were scanned using a 1.5T dedicated breast MRI system (Aurora Dedicated Breast MRI Systems) with a single channel breast coil. For dynamic imaging, gadolinium-diethylenetriamine pentaacetic acid (Gd-DTPA, Magnevist) was intravenously injected as a bolus of 0.2 ml per kg of body weight at a rate of 2 mL/s followed by a 20-mL normal saline flush. A dynamic series of transverse T1-weighted fat-suppression images were acquired at pre-contrast and post-contrast at 90, 270, 450, and 630 s by using the following imaging parameters: TR = 29 ms, TE = 4.8 ms, slice thickness = 1.1 mm, matrix = 360 × 360 × 128, and FOV = 36 cm. In addition, axial short TI inversion recovery T2-weighted (T2W-STIR) images were performed under the following conditions: TR = 6,680 ms, TE = 68 ms, slice thickness = 3.0 mm, matrix = 320 × 192, FOV = 36 cm. Fat suppression was applied using a short TI-inversion recovery technique.

Clinical and Conventional MRI Features Assessment

Clinical and conventional MRI features (CCMF) was used to differentiate PTs from FAs. The clinical variables assessed included age, whether the lesions showed rapid enlargement, and whether the lesions were primary or recurrent. The conventional MRI features for each patient were independently reviewed by two radiologists with 12 and 5 years of experience, respectively, blinded to the histopathological diagnoses. For the cases with discrepancies in the CCMF assessment between the two radiologists, these were jointly reviewed by the two radiologists to reach a consensus for further analysis. Interpretation of some conventional MRI features was based on three following characteristics as per the American College of Radiology Breast Imaging Reporting and Data System MR imaging criteria

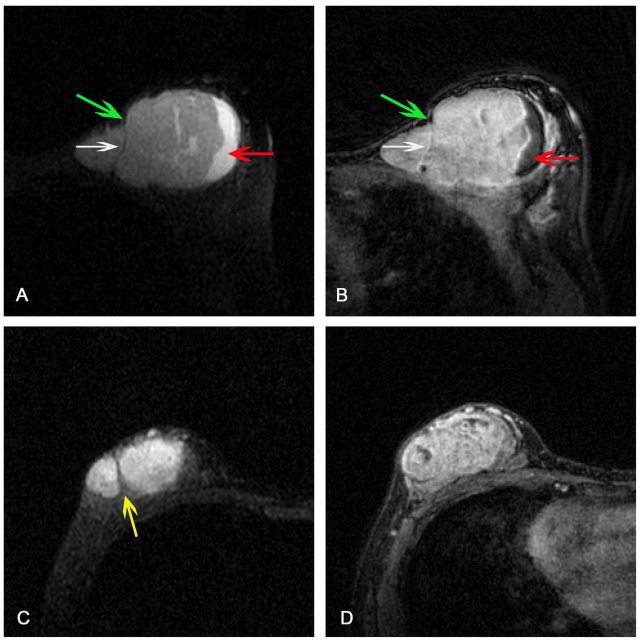


FIGURE 1 | (A) Axial short TI inversion recovery T2-weighted (T2W-STIR) and **(B)** third post-contrast image showing a mass with cystic component (red arrow), weak lobulation with obtuse angle (green arrows), septation (white arrow), and heterogeneous enhancement. Strong lobulation with acute angle (yellow arrow) was detected on **(C)** T2W-STIR and heterogeneous enhancement was detected on **(D)** third post-contrast image.

(version 5) (22, 23): the margin of masses (circumscribed vs. non-circumscribed); initial signal intensity enhancement (slow, medium, or fast); and time–intensity curve (TIC) pattern on dynamic contrast-enhanced images (the persistent, plateau, or washout pattern). The presence or absence of a cystic component and internal septation were determined, and the extent of lobulation was divided into strong (with an acute angle) or weak (obtuse angle). In addition, we analyzed the tumor size (the greatest lesion diameter); tumor signal intensity on T2W-STIR (homogeneous vs. heterogeneous); and signal intensity enhancement of third sequence of post-contrast (homogeneous vs. heterogeneous). For the measurements of enhancement features including initial signal intensity enhancement and TIC, the region of interest (ROI) was placed onto the area of the lesion where the enhancement was strongest in the first sequence of the post-contrast image. Examples of these MRI features were shown in **Figure 1**. For recurrence patients, only clinical and MRI data at the time of recurrence was included and evaluated in this study.

Texture Analysis

The T2W-STIR, T1-weighted pre-contrast, and two contrast-enhanced series were chosen for texture analysis. Image slices were selected on the basis of presentation of the largest lesion diameter. The ROI, containing the entire visible tumor and excluding equivocal normal breast tissue, was manually drawn for each image. Texture analysis was performed by software package MaZda 4.60 (The Technical University of Lodz, Institute of Electronics) (24, 25).

TABLE 1 | Texture features used summary.

Algorithm	Texture features
Histogram	Mean, variance, skewness, kurtosis, percentiles 1, 10, 50, 90, and 99%
Absolute gradient (GrM)	Mean, variance, skewness, kurtosis, and percentage of pixels with non-zero gradient
Co-occurrence matrix (COM)	Angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance and difference entropy; parameters computed for 4 directions: (a, 0), (0, a), (a, a), (a, –a) and 5 distances: a = 1, 2, 3, 4, 5, between image pixels
Run-length matrix (RLM)	Run-length non-uniformity, gray-level non-uniformity, long-run emphasis, short run emphasis, and fraction of image in runs; parameters computed for horizontal, 45°, vertical, and 135° directions
Autoregressive model (ARM)	Model parameter vector includes 4 parameters; Sigma: standard deviation of the driving noise
Wavelet	Energy of the wavelet coefficients in sub-bands

MaZda allows the quantitative analysis of approximately 300 texture features based on the following algorithms: histogram, absolute gradient, run length matrix, co-occurrence matrix, autoregressive model, and wavelet transform (24, 25), as shown in **Table 1**. All these texture features were calculated for each ROI. The co-occurrence matrix parameters were calculated in four directions ($\theta = 0, 45, 90$, and $135°$) with interpixel distances of $n = 1, 2, 3, 4$, and 5. The gray-level normalization, which is known to minimize the effect of contrast variation and brightness, was carried out using a method that normalizes image intensities within $\mu \pm 3\sigma$ (μ, gray-level mean; and σ, gray-level standard deviation).

In MaZda, a combination of feature selection algorithms including mutual information, classification error probability combined with average correlation coefficients, and Fisher coefficient were applied to determine 30 texture parameters with the highest discriminative power for classification on each MRI pulse sequence. These features were then exported for further processing and classification to a statistical program B11 (24).

Feature Selection and Classification

Statistically significant features were selected among the raw texture features on each MRI sequence and CCMF. A linear discriminant analysis was performed for statistically significant features using MaZda to obtain the most discriminative features (26). Then, the K-nearest neighbor classifier ($K = 3$) was employed to distinguish between PT and FA based on the most discriminative features using software routines written in MATLAB 7 (Mathworks). For training the classifier, 28 PTs and 28 FAs were used, whereas for testing the classifier, the remaining 14 PTs and 14 FAs were used.

A workflow chart of the distinction between PT and FA based on texture features and CCMF are illustrated in **Figure 2**.

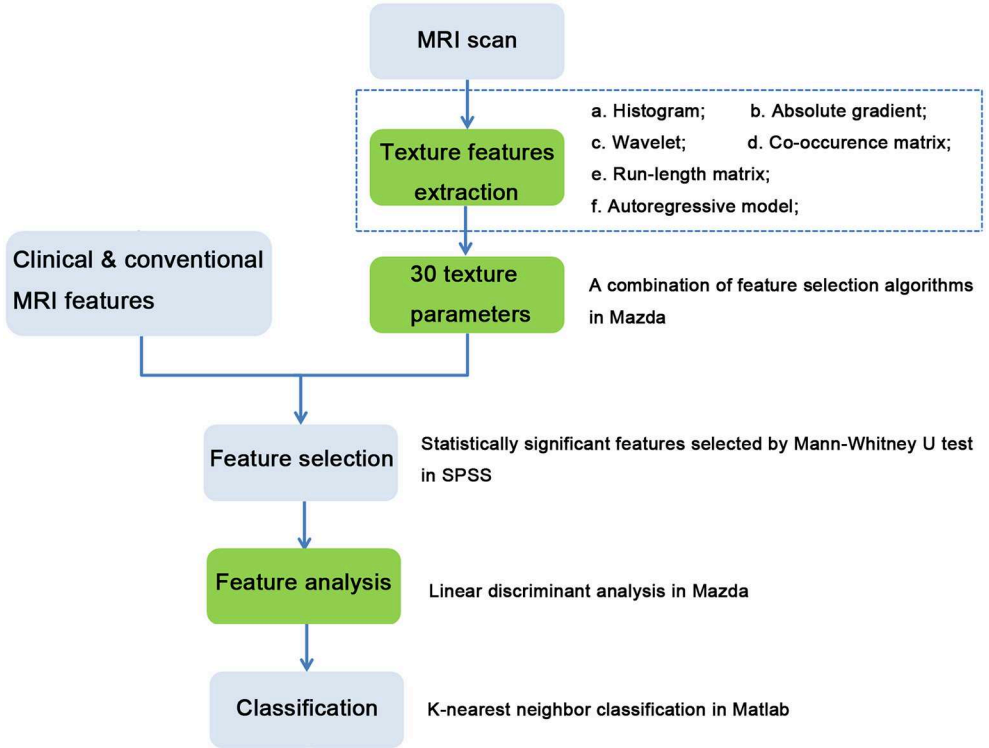


FIGURE 2 | Workflow chart of distinction between phyllodes tumors and fibroadenomas based on clinical and conventional MRI features, and texture features. Processes in green boxes were performed in MaZda.

Statistical Analysis

We compared the raw texture parameters on each sequence and CCMF between PTs and FAs using the Mann–Whitney *U* test. For evaluating the diagnostic efficiency of each approach, we employed receiver operating characteristic (ROC) analysis. These analyses were performed using package SPSS 22.0 for Windows. For each ROC curve, comparisons of the area under the curve (AUC) were performed with methods described by Hanley (27). $P < 0.05$ was considered to indicate statistical significance.

RESULTS

Clinical and Conventional MRI Features

The clinical characteristics and conventional MRI findings of 42 PTs and 42 FAs are summarized in **Table 2**. There were 25 benign, 14 borderline, and 3 malignant PTs based on the histological findings. Patients with PTs were significantly older than those with FAs ($p < 0.001$). The mean maximal diameter (standard deviation) was 4.70 ± 3.45 cm for the PT group and 3.48 ± 2.36 cm for the FA group. The PTs tended to be larger than the FAs, although this difference was not statistically significant ($p > 0.05$). The local recurrence rates of PTs and FAs were 26.2 and 0%, respectively. Overall, 61.9% PTs (26/42) showed strong lobulation, whereas only 13 (31.0%) FAs among 42 expressed strong lobulation. The PTs showed strong lobulation pattern more frequently than FAs ($p = 0.004$). Cystic components were seen in 20 (47.6%) PTs but only in 6 (14.3%) FAs ($p = 0.001$). The PTs had a significantly higher frequency of internal septum than the FAs ($p = 0.009$). The FAs tended to be homogeneous more frequently seen on T2W-STIR than the PTs ($p = 0.001$). There were no significant differences between the PTs and FAs in rapid growth, margin, septation enhancement, tumor signal intensity on the third post-contrast images, initial signal intensity enhancement, and TIC curve pattern.

For clinical and conventional MRI features (CCMF), the classification accuracy of K-nearest neighbor classifier was 76.2%. For ROC analysis, the AUC was 0.76 (95% CI: 0.66, 0.87), and the sensitivity and specificity were both 76.2%.

Texture Features

PTs and FAs presented a differential textural pattern. Certain texture features extracted using MaZda were significantly different, as shown in **Table 3** and **Supplementary Information**. The number of statistically significant texture features on T2W-STIR was greater than other MRI sequences. For texture features on MRI, the classification accuracies were 89.3, 69.1, 71.4, and 67.9%, for T2W-STIR, T1-weighted pre-contrast, and two contrast-enhanced series (first and third post-contrast sequence), respectively. For ROC analysis, the AUCs were 0.89 (95% CI: 0.82, 0.97); 0.69 (95% CI: 0.58, 0.81); 0.71 (95% CI: 0.60, 0.83); and 0.68 (95% CI: 0.56, 0.80) for T2W-STIR, T1-weighted pre-contrast, and the first and third post-contrast sequences, respectively. The most discriminative features on T2W-STIR had higher classification accuracy (89.3%); AUC (0.89, 95% CI: 0.82, 0.97);

TABLE 2 | Clinical and conventional MRI features of phyllodes tumors and fibroadenomas.

	PT	FA	*P*-value
Mean age (SD)	44.38 ± 6.72	35.07 ± 12.90	< 0.001
Rapid enlargement			
Absent	30 (71.4%)	33 (78.6%)	0.614
Present	12 (28.6%)	9 (21.4%)	
Primary/recurrence			
Primary	31 (73.8%)	42 (100%)	< 0.001
Recurrence	11 (26.2%)	0 (0)	
Diameter	4.70 ± 3.45	3.48 ± 2.36	0.07
Margin			
Circumscribed	32 (76.2%)	34 (81.0%)	0.79
Not circumscribed	10 (23.8%)	8 (19.0%)	
Strong lobulation			
Absent	16 (38.1%)	29 (69.0%)	0.004
Present	26 (61.9%)	13 (31.0%)	
Septation			
Absent	15 (35.7%)	27 (64.3%)	0.009
Present	27 (64.3%)	15 (35.7%)	
Enhancement	7 (16.7%)	3 (7.1%)	0.312
No enhancement	35 (83.3%)	39 (92.9%)	
Cystic component			
Absent	22 (52.4%)	36 (85.7%)	0.001
Present	20 (47.6%)	6 (14.3%)	
T2W-STIR			
Homogeneous	22 (52.4%)	36 (85.7%)	0.001
Heterogeneous	20 (47.6%)	6 (14.3%)	
Initial enhancement			
Slow	4 (9.5%)	6 (14.3%)	0.636
Medium	15 (35.7%)	17 (40.5%)	
Fast	23 (54.8%)	19 (45.2%)	
Contrast third			
Homogeneous	16 (38.1%)	23 (54.8%)	0.126
Heterogeneous	26 (61.9%)	19 (45.2%)	
TIC pattern			
Persistent pattern	17 (40.5%)	22 (52.4%)	0.367
Plateau pattern	17 (40.5%)	16 (38.1%)	
Washout pattern	8 (19.0%)	4 (9.5%)	

TABLE 3 | Statistically significant texture features on axial short TI inversion recovery T2-weighted images.

Texture feature	*P*	Z
WavEnHH_s-3	< 0.001	−3.757
WavEnHH_s-1	< 0.001	−4.258
WavEnHL_s-1	0.002	−3.042
GrKurtosis	< 0.001	−4.634
GrSkewness	< 0.001	−5.573
GrMean	< 0.001	−3.569
45dgr_Fraction	< 0.001	−4.258
45dgr_ShrtREmp	< 0.001	−4.169
45dgr_LngREmph	< 0.001	−4.258
S(5,5)SumAverg	0.002	−3.051
S(0,5)SumAverg	0.021	−2.308
S(0,5)InvDfMom	0.003	−2.934
S(4,4)SumAverg	0.003	−2.952
S(4,4)InvDfMom	< 0.001	−3.918
S(3,0)Contrast	0.014	−2.460
S(2,2)InvDfMom	< 0.001	−3.811
S(2,0)DifVarnc	0.011	−2.541
S(1, −1)DifEntrp	0.004	−2.845
S(1,1)DifEntrp	< 0.001	−3.695
S(1,1)InvDfMom	< 0.001	−4.053
S(1,1)Correlat	0.002	−3.131
S(1,1)Contrast	0.001	−3.382
S(1,0)DifEntrp	< 0.001	−3.543
S(1,0)Correlat	0.001	−3.185
S(1,0)Contrast	0.001	−3.319
Variance	< 0.001	−4.348

sensitivity (88.1%); and specificity (90.5%) than those on other MRI sequences. The result of K-nearest neighbor classifier and ROC analysis are listed in **Table 4**.

Combination

For the combination of CCMF with texture features on T2W-STIR, the classification accuracy was 95.2%. The AUC was 0.95 (95% CI: 0.90, 1.00), with a specificity of 95.2% and sensitivity of 95.2%.

Comparison of Diagnostic Performance

Figure 3 shows the ROC curves for the K-nearest neighbor classifier when the classifier was trained with most discriminative features of CCMF, texture features on each MRI sequence, and the combination strategy. The texture features on T2W-STIR, with higher classification accuracy (89.3 vs. 76.2%) and AUC (0.89 vs. 0.76), outperformed CCMF. In addition, CCMF was less sensitive than texture features on T2W-STIR (76.2 vs. 88.1%) resulting in a few false negative results (example shown in **Figure 4**), and exhibited lower specificity (76.2 vs. 90.5%) resulting in a few false positive results (example shown in **Figure 5**). The AUC of the combination was significantly higher than that of CCMF or texture features on T2W-STIR alone ($p < 0.05$). According to the result of K-nearest neighbor classification and AUC, the diagnostic performance of the combination performed better than texture features on T2W-STIR or CCMF alone.

DISCUSSION

In the current study, texture analysis based on MRI was applied to evaluate the differential diagnosis between breast PTs and FAs. Texture features on T2W-STIR with higher classification accuracy and AUC performed better than clinical and conventional MRI features (CCMF). Texture features on T2W-STIR were more sensitive than CCMF which exhibited higher specificity. In our study, PT could be mistaken for FA using CCMF but was correctly identified using texture features on T2W-STIR, regardless of whether the lesion was benign or malignant. In addition, we found that the diagnostic performance using the combination of CCMF with texture features based on

TABLE 4 | Features classification and receiver operating characteristic analysis of phyllodes tumors and fibroadenomas.

	Classification accuracy	AUC (95% CI)	Sensitivity	Specificity
T2W-STIR	89.3%	0.89 (0.82, 0.97)	88.1% (37/42)	90.5% (38/42)
Pre-contrast	69.1%	0.69 (0.58, 0.81)	73.8% (31/42)	64.3% (27/42)
First post-contrast	71.4%	0.71 (0.60, 0.83)	71.4% (30/42)	71.4% (30/42)
Third post-contrast	67.9%	0.68 (0.56, 0.80)	66.7% (28/42)	69.0% (29/42)
CCMF	76.2%	0.76 (0.66, 0.87)	76.2% (32/42)	76.2% (32/42)
Combination	95.2%	0.95 (0.90, 1.00)	95.2% (40/42)	95.2% (40/42)

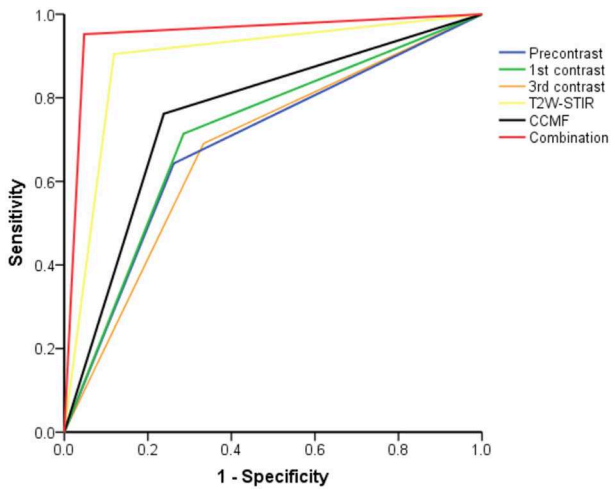


FIGURE 3 | The receiver operating characteristic curves from each approach for differentiation between phyllodes tumors and fibroadenomas.

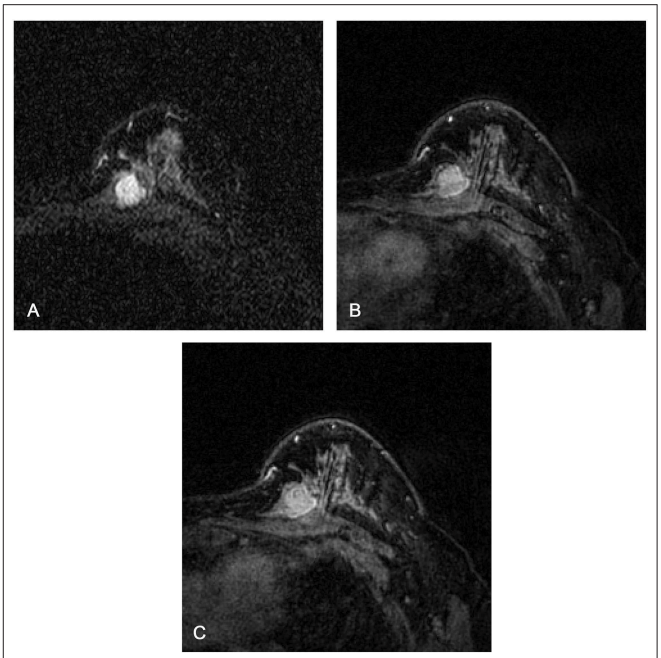


FIGURE 4 | Magnetic resonance images of a 37-year-old female patient with a borderline phyllodes tumor: **(A)** axial short TI inversion recovery T2-weighted (T2W-STIR) **(B)** first post-contrast **(C)** third post-contrast. The texture features on T2W-STIR correctly identified a phyllodes tumor which was falsely interpreted as a fibroadenoma on clinical and conventional MRI features, possibly owing to the weak lobulation, homogeneous signal on T2W-STIR, and absence of cystic component and septation.

T2W-STIR was better than CCMF or texture features on T2W-STIR alone. The classification accuracy reached 95.2%, when the most discriminative features of combination strategy were used to train the classifier. By using a combination strategy, the AUC, specificity, and sensitivity were 0.95, 95.2%, and 95.2%, respectively.

Prior studies (15, 28–30) have indicated differences in the clinical and conventional MRI characteristics for differentiating between PTs and FAs, which was also validated in this study. Our study showed that higher age, recurrence, strong lobulation, and internal cystic components were detected significantly more frequently in PTs than in FAs, which were in line with prior studies (15, 28–30). Some groups report that hypointense internal septation was more likely to be presented in PTs than in FAs, but this difference was not statistically significant (15). Even though PTs showed significantly more frequent hypointense internal septations than FAs on MRI in this study, septation enhancement was not statistically significant between PTs and FAs. As reported in some articles (8, 15), PTs were frequently larger than FAs, but there was no significant difference in size between them in our study, likely because the selected tumors were of a relatively large size (>1 cm in diameter). Kamitani et al. (15) found that FAs tended to be homogeneous more frequently than PTs on T1-weighted post-contrast images, but this difference was not significant; there was no significant difference in the TIC curve pattern between the two groups; both of which were consistent with our results. In addition, we found a significantly higher frequency of heterogeneous signals in PTs than in FAs on T2W-STIR, which had been rarely mentioned in previous literatures (15).

Both PTs and FAs are breast fibroepithelial neoplasms. Histologically, they share a dimorphic pattern with both epithelial and stromal components. However, PT can usually be differentiated from FA by its exaggerated intracanalicular growth pattern with increased and heterogeneous stromal cellularity (9). Internal cystic components, septation, and heterogeneous signal on T2W-STIR may be caused by the histopathologically heterogeneous nature of PT, and the stronger lobulation might be related to the rapid growth.

Texture analysis was utilized to evaluate the ability to differentiate PTs from FAs. The number of statistically significant texture features on T2W-STIR was larger than those on T1-weighted pre-contrast and two contrast-enhanced series. Furthermore, the diagnostic performance of these statistically significant texture features on T2W-STIR outperformed that on other MRI sequences, with an AUC of 0.89 and a classification accuracy of 89.3%. The result of texture analysis was in line

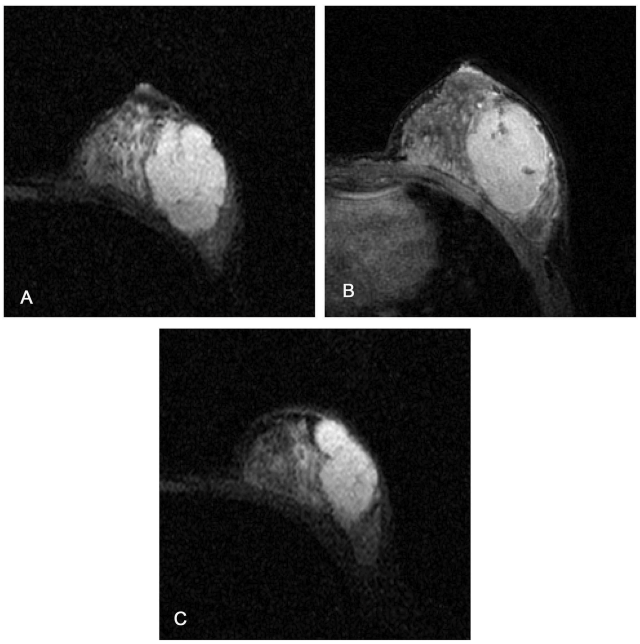


FIGURE 5 | Magnetic resonance images of a 26-year-old female patient with fibroadenomas: **(A,C)** axial short TI inversion recovery T2-weighted (T2W-STIR), **(B)** third post-contrast. The texture features on T2W-STIR correctly identified the fibroadenoma which was falsely interpreted as a phyllodes tumor on clinical and conventional MRI features, possibly due to the cystic component, strong lobulation, and septation.

with that of conventional MRI characteristics that the signal intensity on T2W-STIR was significantly different, but there were no significant differences in features after enhancement between PTs and FAs, such as tumor signal intensity on the third sequence of post-contrast images, initial signal intensity enhancement, and the TIC curve pattern. Previous breast MRI studies mainly focused on dynamic enhancement sequence probably because of its detailed morphological and hemodynamic information; however, dynamic contrast-enhanced MRI was less significant than T2W-STIR to distinguish between PTs and FAs in our study. The echo time of T2W-STIR is relatively long, which offers a higher signal-to-noise ratio, spatial resolution, and soft tissue contrast of breast PTs and FAs. Hence, we hypothesized that texture analysis based on these T2W-STIR images might reveal more subtle alterations in the tumor microenvironment. Textural features extracted from T2W-STIR reflect more differences between PTs and FAs, by capturing the intra-tumoral heterogeneity.

In our study, mean lesion sizes were >3 cm for both PT and FA. With lesion sizes >3 cm, it would be advisable that all these lesions should be resected anyway (8), but they require different surgical procedures. FAs need only enucleation, whereas both benign and malignant PTs require wide local excision with a margin of at least 1 cm (10, 11) because the high recurrence rate in patients with resection margins of <1 cm around the primary tumor (10, 13). A combination of CCMF with texture features on T2W-STIR can provide accurate preoperative diagnosis for these cases with mean sizes >3 cm, which allows appropriate clinical strategy and avoidance of operative complications resulting from inadequate excision or surgical overtreatment.

There were several limitations in our study. First, some of the MRI images were collected after fine needle aspiration of the primary tumor, and thus the hemorrhage or edema caused by the biopsy could have potentially affected feature calculation. Second, we did not explore differences among PTs of all the histologic grading due to the lack of a sufficient number of borderline and malignant PTs. Third, little pathophysiological semantics of the textural features are currently known. Additional work is necessary to understand the underlying biology of these tumors evaluated by texture analysis. Last, as a retrospective study with a small sample size of 84 cases, inherent variations and biases may have influenced the results. Further validation with a larger dataset from different centers and scanners should be strongly considered.

In conclusion, textural features extracted from T2W-STIR showed better diagnostic performance than CCMF. In addition, a combination of CCMF with texture features on T2W-STIR can reflect better diagnostic performance than CCMF or texture features on T2W-STIR alone. Texture analysis provided a novel approach to non-invasively and accurately distinguish PTs from FAs. With ongoing validation, MRI-based texture features may become a potential biomarker and provide a useful medical decision tool in clinical trials in patients with breast fibroepithelial neoplasms.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/**Supplementary Files**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board Guangdong Women and Children Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HM, YM, KJ, and LL: conception and design. HM, YM, and KJ: manuscript writing. HM, YM, and TD: provision of study materials or patients. YT, SW, and SH: collection and assembly of data. HM, YM, TD, and YQ: MRI analysis and interpretation. HM, YM, XZ, and XH: statistical analysis. HM, YM, KJ, and LL: final approval of manuscript.

FUNDING

This work has received funding by The Third Affiliated Hospital of Guangzhou Medical University Youth Research Project (Grant Number 2017Q07), Guangzhou Key Laboratory of Molecular and Functional Imaging for Clinical Translation (Grant Number 201905010003), Key Program of Natural Science

Foundation of Guangdong Province of China (Grant Number 2018B0303110011), and National Natural Science Foundation of China (Grant Number 21317241, 81971672).

ACKNOWLEDGMENTS

We gratefully acknowledge Engineering Research Center of Medical Imaging Artificial Intelligence for Precision Diagnosis and Treatment of Guangdong Province for their technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01021/full#supplementary-material

REFERENCES

1. Rowell MD, Perry RR, Hsiu JG, Barranco SC. Phyllodes tumors. *Am J Surg.* (1993) 165:376–9. doi: 10.1016/s0002-9610(05)80849-9
2. Liberman L, Bonaccio E, Hamele-Bena D, Abramson AF, Cohen MA, Dershaw DD. Benign and malignant phyllodes tumors: mammographic and sonographic findings. *Radiology.* (1996) 198:121–4. doi: 10.1148/radiology.198.1.8539362
3. Hoda SA, Kaplan RE. World Health Organization (WHO) Classification of Breast Tumours, 4th ed. *Am J Surg Pathol.* (2013) 37:309–10.
4. Karim RZ, Gerega SK, Yang YH, Spillane A, Carmalt H, Scolyer RA, et al. Phyllodes tumours of the breast: a clinicopathological analysis of 65 cases from a single institution. *Breast.* (2009) 18:165–70. doi: 10.1016/j.breast.2009.03.001
5. Tan PH, Thike AA, Tan WJ, Thu MM, Busmanis I, Li H, et al. Predicting clinical behaviour of breast phyllodes tumours: a nomogram based on histological criteria and surgical margins. *J Clin Pathol.* (2012) 65:69–76. doi: 10.1136/jclinpath-2011-200368
6. Reifuss M. The treatment of and prognosis of patients with phyllodes tumor of the breast: an analysis of 170 cases. *Cancer.* (1996) 77:910–6.
7. Bhargav PR, Mishra A, Agarwal G, Agarwal A, Verma AK, Mishra SK. Phyllodes tumour of the breast: clinicopathological analysis of recurrent vs. non-recurrent cases. *Asian J Surg.* (2009) 32:224–8. doi: 10.1016/S1015-9584(09)60398-5
8. Foxcroft LM, Evans EB, Porter AJ. Difficulties in the pre-operative diagnosis of phyllodes tumours of the breast: a study of 84 cases. *Breast.* (2007) 16:27–37. doi: 10.1016/j.breast.2006.05.004
9. Jacklin RK, Ridgway PF, Ziprin P, Healy V, Hadjiminas D, Darzi A. Optimising preoperative diagnosis in phyllodes tumour of the breast. *J Clin Pathol.* (2006) 59:454–9. doi: 10.1136/jcp.2005.025866
10. Mangi AA, Smith BL, Gadd MA, Tanabe KK, Ott MJ, Souba WW. Surgical management of phyllodes tumors. *Arch Surg.* (1999) 134:492–3. doi: 10.1001/archsurg.134.5.487
11. Chaney AW, Pollack A, Mcneese, MD, Zagars, GK, Pisters PW, Pollock RE, et al. Primary treatment of cystosarcoma phyllodes of the breast. *Cancer.* (2000) 89:1502–11. doi: 10.1002/1097-0142(20001001)89:7<1502::aid-cncr13>3.0.co;2-p
12. Yabuuchi H, Soeda H, Matsuo Y, Okafuji T, Eguchi T, Sakai S, et al. Phyllodes tumor of the breast: correlation between MR findings and histologic grade. *Radiology.* (2006) 241:702. doi: 10.1148/radiol.2413051470
13. Chen WH, Cheng SP, Tzen CY, Yang TL, Jeng KS, Liu CL, et al. Surgical treatment of phyllodes tumors of the breast: retrospective review of 172 cases. *J Surg Oncol.* (2005) 91:185–94. doi: 10.1002/jso.20334
14. Cohn-Cedermark G, Rutqvist LE, Rosendahl I, Silfversward C. Prognostic factors in cystosarcoma phyllodes. A clinicopathologic study of 77 patients. *Cancer.* (1991) 68:2017. doi: 10.1002/1097-0142(19911101)68:9<2017::AID-CNCR2820680929>3.0.CO;2-V
15. Kamitani T, Matsuo Y, Yabuuchi H, Fujita N, Nagao M, Kawanami S, et al. Differentiation between benign phyllodes tumors and fibroadenomas of the breast on MR imaging. *Eur J Radiol.* (2014) 83:1344–9. doi: 10.1016/j.ejrad.2014.04.031
16. Lambin P, Leijenaar RT, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* (2017) 14:749. doi: 10.1038/nrclinonc.2017.141
17. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
18. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Resonance Imag.* (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
19. Dong Y, Feng Q, Yang W, Lu Z, Deng C, Zhang L, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI. *Eur Radiol.* (2018) 28:582–91. doi: 10.1007/s00330-017-5005-7
20. Holli-Helenius K, Salminen A, Rinta-Kiikka I, Koskivuo I, Brück N, Boström P, et al. MRI texture analysis in differentiating luminal A and luminal B breast cancer molecular subtypes - a feasibility study. *BMC Med Imag.* (2017) 17:69. doi: 10.1186/s12880-017-0239-z
21. Holli K, Lääperi AL, Harrison L, Luukkaala T, Toivonen T, Ryymin P, et al. Characterization of breast cancer types by texture analysis of magnetic resonance images. *Acad Radiol.* (2010) 17:135–41. doi: 10.1016/j.acra.2009.08.012
22. Radiology ACo. *ACR BI-RADS-Magnetic Resonance Imaging. ACR Breast Imaging and Data System, Breast Imaging Atlas* (2013).
23. Rao AA, Feneis J, Lalonde C, Ojeda-Fournier H. A Pictorial Review of Changes in the BI-RADS Fifth Edition. *Radiographics.* (2016) 36:623–39. doi: 10.1148/rg.2016150178
24. Szczypinski PM, Strzelecki M, Materka A, Klepaczko A. MaZda—a software package for image texture analysis. *Computer Methods Programs Biomed.* (2009) 94:66–76. doi: 10.1016/j.cmpb.2008.08.005
25. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clin Radiol.* (2004) 59:1061–9. doi: 10.1016/j.crad.2004.07.008
26. Huang X, Zhang Y, Qian M, Meng L, Xiao Y, Niu L, et al. Classification of carotid plaque echogenicity by combining texture features and morphologic characteristics. *J Ultrasound Med.* (2016) 35:2253. doi: 10.7863/ultra.15.09002
27. Hanley JA, Mcneil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* (1983) 148:839–43. doi: 10.1148/radiology.148.3.6878708
28. Plaza MJ, Swintelski C, Yaziji H, Torres-Salichs M, Esserman LE. Phyllodes tumor: review of key imaging characteristics. *Breast Dis.* (2015) 35:79–86. doi: 10.3233/BD-150399
29. Brinck U, Fischer U, Korabiowska M, Jutrowski M, Schauer A, Grabbe E. The variability of fibroadenoma in contrast-enhanced dynamic MR mammography. *Am J Roentgenol.* (1997) 168:1331–4. doi: 10.2214/ajr.168.5.9129437
30. Susanne W, Herzog AB, Fischer DR, Marx C, Raabe G, Schneider A, et al. Differentiation of phyllodes breast tumors from fibroadenomas on MRI. *Am J Roentgenol.* (2005) 185:1317–21. doi: 10.2214/AJR.04.1620

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 05 November 2019
doi: 10.3389/fonc.2019.01164

Radiomics-Based Machine Learning Technology Enables Better Differentiation Between Glioblastoma and Anaplastic Oligodendroglioma

Yimeng Fan[1,2†], Chaoyue Chen[3†], Fumin Zhao[4†], Zerong Tian[3], Jian Wang[5], Xuelei Ma[6,7†] and Jianguo Xu[3*]*

[1] Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, China, [2] State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [3] Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu, China, [4] Department of Radiology, West China Second University Hospital, Sichuan University, Chengdu, China, [5] School of Computer Science, Nanjing University of Science and Technology, Nanjing, China, [6] Department of Biotherapy, Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [7] State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, and Collaborative Innovation Center for Biotherapy, Chengdu, China

OPEN ACCESS

Edited by:
Yin Li,
University of Wisconsin-Madison, United States

Reviewed by:
Andre Bongers,
University of New South Wales, Australia
Yanwei Miao,
Dalian Medical University, China

***Correspondence:**
Xuelei Ma
drmaxuelei@gmail.com
Jianguo Xu
drjianguoxu@gmail.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *14 June 2019*
Accepted: *17 October 2019*
Published: *05 November 2019*

Citation:
Fan Y, Chen C, Zhao F, Tian Z, Wang J, Ma X and Xu J (2019) Radiomics-Based Machine Learning Technology Enables Better Differentiation Between Glioblastoma and Anaplastic Oligodendroglioma. Front. Oncol. 9:1164. doi: 10.3389/fonc.2019.01164

Purpose: The aim of this study was to test whether radiomics-based machine learning can enable the better differentiation between glioblastoma (GBM) and anaplastic oligodendroglioma (AO).

Methods: This retrospective study involved 126 patients histologically diagnosed as GBM ($n = 76$) or AO ($n = 50$) in our institution from January 2015 to December 2018. A total number of 40 three-dimensional texture features were extracted from contrast-enhanced T1-weighted images using LIFEx package. Six diagnostic models were established with selection methods and classifiers. The optimal radiomics features were separately selected into three datasets with three feature selection methods [distance correlation, least absolute shrinkage and selection operator (LASSO), and gradient boosting decision tree (GBDT)]. Then datasets were separately adopted into linear discriminant analysis (LDA) and support vector machine (SVM) classifiers. Specificity, sensitivity, accuracy, and area under curve (AUC) of each model were calculated to evaluate their diagnostic performances.

Results: The diagnostic performance of machine learning models was superior to human readers. Both classifiers showed promising ability in discrimination with AUC more than 0.900 when combined with suitable feature selection method. For LDA-based models, the AUC of models were 0.986, 0.994, and 0.970 in the testing group, respectively. For the SVM-based models, the AUC of models were 0.923, 0.817, and 0.500 in the testing group, respectively. The over-fitting model was GBDT + SVM, suggesting that this model was too volatile that unsuitable for classification.

Conclusion: This study indicates radiomics-based machine learning has the potential to be utilized in clinically discriminating GBM from AO.

Keywords: machine learning, magnetic resonance imaging, glioblastoma, anaplastic oligodendroglioma, texture analysis

INTRODUCTION

High-grade gliomas, the most common malignant solidary brain tumors in adults, are traditionally classified into anaplastic oligodendroglioma (AO), anaplastic astrocytoma (AA), and Glioblastoma (GBM) (1, 2). According to the WHO classification, AO was ascribed in Grade III while GBM was ascribed in Grade IV based on their histology characteristics. The early diagnosis of GBM and AO is clinically challenging but necessary due to their different treatment choice as well as the therapeutic responsiveness and patient survival (3). As for GBM extended resection is recommended to increase patient survival, whereas for AO this strategy lacks solid evidence (4–6). The treatment after surgery is also different as well. For GBM, standardized therapy after surgery recommended by NCCN guidelines is standard brain radiation therapy (RT) + concurrent temozolomide (TMZ) followed by adjuvant chemotherapy (7). While for AO, it is recommended to use fractionated external beam RT together with neoadjuvant or adjuvant PCV (procarbazine, lomustine, and vincristine) regarding the specific condition of each patient (7, 8).

A glioma-specific blood biomarker for glioma has not been identified yet. Therefore, the radiology examination is critical for tumor detection and lesion localization. Brain magnetic resonance imaging (MRI) plays a key role in the preoperative diagnostic of gliomas with high image resolution on tumor tissue. However, in some cases, MRI may be unable to provide enough information for differentiation between GBM and AO. The MRI characteristics of two tumors are pretty similar when GBM is characterized by perilesional vasogenic edema and ring-like enhancement (9–11); while AO also shows peritumoral edema and heterogeneous enhancement (11, 12). In this regard, the urgency of new radiological method has been highlighted.

Given that texture analysis on images provides a more objective information beyond naked eye assessment, quantitative descriptions of tumor characteristics could be an option for clinical diagnosis (13–16). Moreover, with digital parameters, new technology, such as machine learning, can be introduced for further statistical analysis. Machine learning, a hotspot in the field of artificial intelligence, enables the extraction of meaningful patterns from massive datasets and thereby achieving precise predictions with the model built (17). Machine learning has demonstrated outstanding performance in previous research including segmentation of the tumor, classification of certain types of tumor, and prediction of survival or genotype (18–23). Although the differentiation between GBM and AO is of high clinical relevance, the machine learning approach has never been explored yet. In this study, we investigated the feasibility of radiomics-based machine learning to differentiate GBM and AO.

Abbreviations: AO, anaplastic oligodendroglioma; AUC, area under curve; GBDT, gradient boosting decision tree; GBM, glioblastoma; LASSO, least absolute shrinkage and selection operator; LDA, linear discriminant analysis; MRI, magnetic resonance imaging; SVM, support vector machine; T1C image, contrast-enhanced T1-weighted image; VOI, volume of interest.

MATERIALS AND METHODS

Study Patients

In this retrospective single-center research, we viewed medical records in neurosurgery department to initially search for patients histologically diagnosed with GBM or AO from January 2015 to December 2018. The medical records were reviewed by two researchers to enroll the potentially qualified patients and to collect relevant clinical information for our research. The inclusion criteria for patients were: (1) with pathological diagnosis of GBM or AO in intraoperative freezing biopsy, and (2) with available high-quality pre-treatment MR scan performed at our institution before surgical resection. Then the pre-surgical MRI images of patients were exported from radiological department though Picture Archiving and Communication Systems (PACS) with uniform standard.

For patients before 2016, we made correction on their pathological diagnoses based on the new World Health Organization 2016 classification of gliomas by a senior neuropathologist with working experience of 10 years (24). The new standards required the presence of both IDH-mt and 1p19q co-deletion for the diagnosis of AO, otherwise it could only be regarded as NOS (Not Otherwise Specified) (24). Therefore, we excluded patients based on new classification who were with incomplete gene reports or with absent presence of both gene expression.

Seventy-nine consecutive patients with GBM and 56 consecutive patients with AO fulfilled inclusion criteria in the initial selection. Three patients with GBM and six patients with AO were excluded in the following evaluation according to the exclusion criteria, which were: (1) presence of motion artifacts on MRI, (2) previous history of brain surgery or biopsy, (3) previous history of intracranial diseases, such as subarachnoid hemorrhage, cerebral infarction, etc. Based on this strategy, a study cohort was built consisting of 76 GBM patients (mean age: 46.5 years) and 50 AO patients (mean age: 47.1 years).

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The institutional review board approved this retrospective study. The written informed consent was obtained from participants enrolled in this study. The written informed consent was necessary before radiological examination (written informed consent for patients <16 years old was signed by parents or guardians) for each patient. The patients agreed to undertake examination when needed and were informed that the statistics (including MR image), which could be used for academic purpose in the future, would be stored in our institutional database. The Ethics Committee of Sichuan University and radiology department of our institution have approved for statistics export and utilization for this study.

MR Image Acquisition

The current study focused on the conventional MR sequences. The suitable sequence should be chosen first for two reasons, that the descriptions on features boundary were vague in some

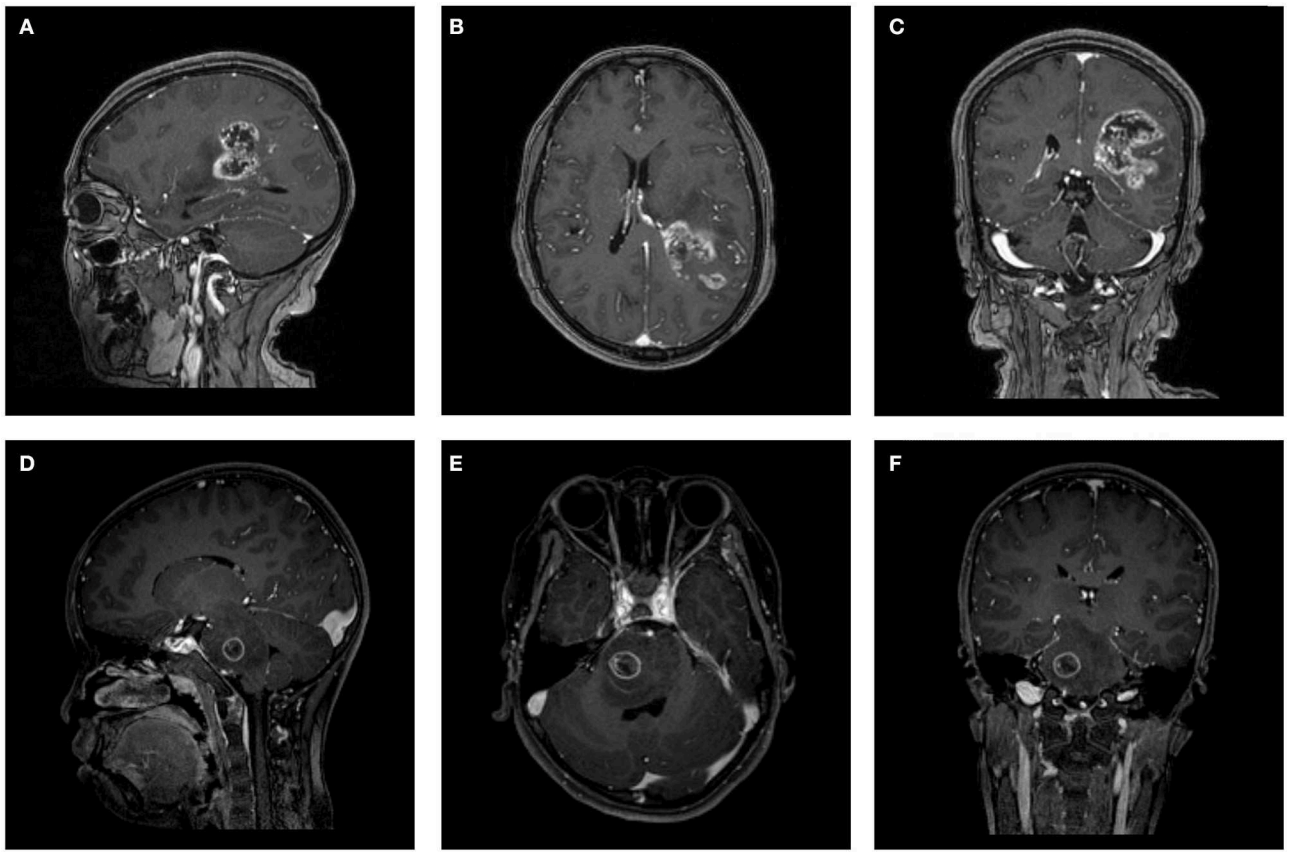


FIGURE 1 | Two examples of contrast-enhanced MRI images. **(A–C)** Patient with GBM in **(A)** parasagittal, **(B)** axial, and **(C)** coronal view. **(D–F)** Patient with AO in **(D)** parasagittal, **(E)** axial, and **(F)** coronal view. GBM, glioblastoma; AO, anaplastic oligodendroglioma.

sequences, and that features from all sequences would bring too much burden on classifiers. After initial evaluation on images and consultation with senior radiologists, the contrast-enhanced MRI sequence was the only one used to perform texture analysis in this study.

The MR scans were performed in the radiology department of institution. The contrast-enhanced MRI sequences were obtained with a 3.0T Siemens Trio Scanner using a MPRAGE sequence with the following imaging parameters: TR/TE/TI = 1900/2.26/900 ms, Flip angle = 9°, slice thickness = 1 mm, axial FOV = 25.6 × 25.6 cm^2 and data matrix = 256 × 256. Intravenous injection of gadopentetate dimeglumine (0.1 mmol/Kg) was taken as contrast agent for patients. Multi-directional data for contrast-enhanced MRI were collected during the interval time of 90-250s. **Figure 1** shows two examples of contrast-enhanced MRI images.

Human Readers Assessment

To test whether machine learning could outperform human readers, the diagnostic performance of them was compared. A senior neurosurgeon and a senior radiologist independently made diagnosis based on contrast-enhanced T1-weighted images, which were presented randomly, regarding classification as GBM or AO. Both readers were blinded to patient information and pathology reports. Then, the accuracy, sensitivity, and specificity were calculated for further analysis.

Texture Feature Extraction

The texture features of tumor tissue were extracted by two researchers using the LIFEx package under the supervision of a senior radiologist (25). Disagreements between researchers were recorded and adjudicated by consulting senior radiologists and neurosurgeons. The volume of interest (VOI) was drawn on T1C images by contouring the outer margin of tumor tissues slice by slice. The peritumoral edema band and adjacent structure invasion were separated from the primary tumor with the difference in contrast enhancement. For the lesions with multiple (>2) enhancement foci, ROI was only performed on the biggest one for those with clear boundary, and on tumor-confirmed area for those with vague boundary. After the ROI delineation, texture features were calculated automatically with default setting.

A total of 40 three-dimensional (3D) texture features were calculated from two orders. In the first order, texture features were calculated from shape histogram-based matrix and histogram-based matrix. In the second or higher order, features were calculated from gray-level co-occurrence matrix (GLCM), gray-level zone length matrix (GLZLM), neighborhood gray-level dependence matrix (NGLDM), and gray-level run length matrix

(GLRLM). To avoid the interference of the lower image matrix resolution, texture analysis performed only on the VOIs with more than 64 voxels by default setting. All original data about extracted features were shown in **Supplementary Material 1**.

To ensure the validity and reproducibility of the extraction, the procedure was performed twice, and the difference between two sets was examined with Manny-Whitney *U*-test. We adjusted the $q < 0.01$ as significant (before was $p < 0.05$) to avoid the interference of false-positive errors rising from a large number of texture features. The results suggested that none of the features were significantly different, implying that the results could be considered reliable and reproducible (shown in **Supplementary Material 2**).

Classification Model Establishment

The purpose of machine learning was to train the models to predict whether each tumor was a GBM or AO with radiomics parameters extracted from the tumor tissue image. However, feature selection was necessary to eliminate statistically insignificant features and to avoid overfitting, which contributes to decreased running time and increased accuracy of the resulting models (26–28). In this study, we employed three selection methods with different selection mechanisms: distance correlation as representative of filter models, least absolute shrinkage and selection operator (LASSO) and gradient boosting decision tree (GBDT) as representatives of embedded models. Then, three datasets were generated with three different selection methods, which were each classified separately. The list of features selected with three different methods are shown in **Supplementary Material 3**, and the explanation of the features are summarized in **Supplementary Material 4**.

The next step was to choose suitable classifiers. Since linear classifier and non-linear classifier represent the state-of-the-art in pattern recognition, we adopted linear discriminant analysis (LDA) and support vector machine (SVM) classification algorithms in the current study as representatives of two classifier types (29). This way, overall six diagnostic models were established based on three selection methods and two classifiers.

As for the algorithm deployment, the study cohort was randomly divided into two subsets as training group and validation group on a proportion of 4:1. When the training on classifiers finished, the validation group was fed to evaluate the diagnostic performance of the models. Sensitivity, specificity, accuracy, and area under receiver operating characteristic curve (AUC) were calculated for both the training and validation group. To appraise the robustness of the methods, the procedure was repeated for 100 cycles with different and independent case assignments. The schematic workflow from image processing to machine learning is shown in **Figure 2**.

RESULTS

Patient Characteristics

Among 126 patients were enrolled in the current study, 76 patients were diagnosed with GBM, and 50 patients with AO. The sex ratio, mean age, and time between MR scan and pathological diagnosis were summarized in **Table 1**. As for the human reader assessment, the accuracy for the neurosurgeon was 63.49%, and for the radiologist was 66.77%. Based on the results, a strong tendency on misdiagnosis of AO could be observed.

Diagnostic Performance of Models

The classification models exhibited promising discriminative ability when combined with suitable selection methods. For LDA-based models, all three models presented feasible performance with the AUC in the validation groups of 0.986, 0.994, and 0.970, respectively. For the SVM-based models, the models showed feasible performance with the AUC in the training groups of 0.923, 0.817, and 0.500. Overfitting was observed in one SVM-based model (SVM + GBDT), suggesting this model was volatile in application. The value of average sensitivity, specificity, accuracy, and AUC of training group and testing group are summarized in **Table 2**.

Figure 3 represents the two-dimensional projection of the LDA-based models, illustrating that the GBM and AO formed distinctive clusters in the space defined by discriminant functions 1 and 2 generated by LASSO + LDA. **Figure 4** shows the examples of performance of LDA-based models in terms of the distribution of the canonical functions in the 100 independent training cycles in the MRI analysis. A clear negative-values shift of the LDA function can be observed for AO, and all positive-values shift for GBM. ROCs of all models are shown in **Supplementary Material 5**.

DISCUSSION

For patients with high-grade gliomas, accurate tumor classification is clinically important because of its close relation with treatment strategy as well as therapeutic responsiveness and prognosis (3). In this study, we applied radiomics-based machine learning to pre-surgically differentiate between GBM and AO. Six models based on three selection methods (distance correlation, LASSO, and GBDT) and two classifiers (LDA and SVM) were built and evaluated. Our results demonstrated that machine learning approaches can be utilized and are clearly superior to human reader diagnosis.

Previous studies have explored the possibility of using machine learning for classification of brain tumor types (18, 21, 30). In the setting of gliomas, several studies have proved the utility of machine learning to differentiate between high-grade and low-grade glioma with high accuracy (0.80 and 0.945) (31, 32). In the setting of differentiation among specific histological subtypes of gliomas, a computer-aided diagnosis system was proposed and evaluated in a previous study to distinguish GBM from lower-grade gliomas, with positive results (33). A multicenter investigation also confirmed the feasibility of using 3D texture analysis for pediatric glioma subtype classification (medulloblastoma, pilocytic astrocytoma, and ependymoma) with an overall accuracy of 0.87 (34). The current study investigated a subject that has never been explored before, that the feasibility of radiomics-based machine learning in discriminating GBM from AO. Diagnostic performance of six

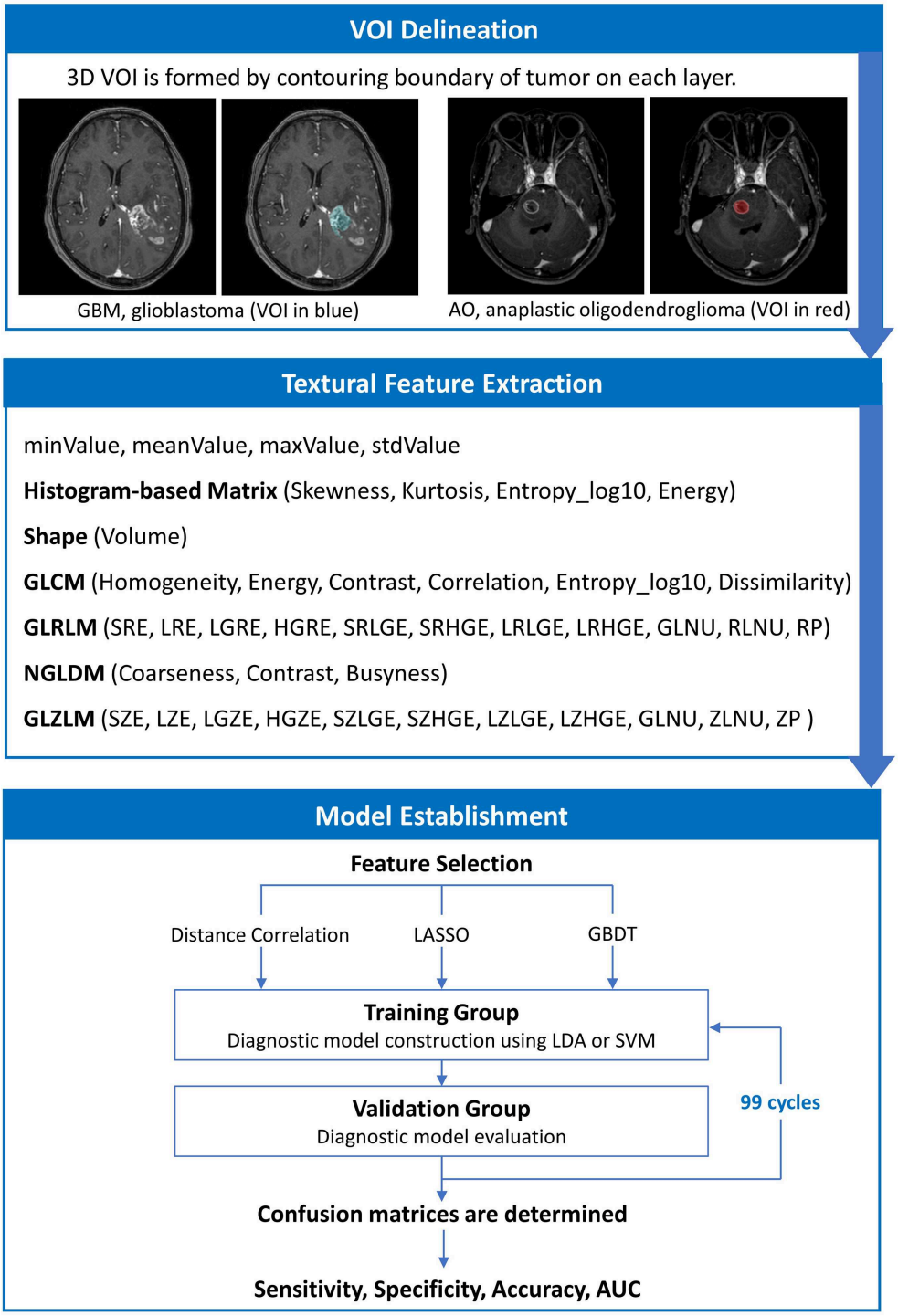


FIGURE 2 | The schematic workflow from image processing to machine learning.

models was assessed in the current study built on three selection methods (distance correlation, LASSO, and GBDT) coupled with two classifiers (LDA and SVM). In general, both classifiers showed high diagnostic performance with AUC more than 0.900 when combined with a suitable selection method. Nevertheless, when comparing between two classifiers, LDA-based models had slightly better diagnostic performance than that of SVM-based models.

The diagnostical models were established based on two types of classifies which differ in computing mechanism considering the performance of a certain classifier may be various in the settings of different tumors. LDA is a representative of the linear

TABLE 1 | Demographics of patients.

	Number	Sex	Age, mean (range), y	Time between MR scan and pathological diagnosis	Human reader 1 accuracy	Human reader 2 accuracy
GBM	76	47% male, 53% female	46.5 (15–80)	6.5 days	81.58%	85.53%
AO	50	50% male, 50% female	47.1 (16–76)	7.9 days	36.00%	38.00%
All patients	126	48% male, 52% female	46.8 (15–80)	7.1 days	63.49%	66.77%

GBM, glioblastoma; AO, anaplastic oligodendroglioma.

TABLE 2 | Results of the discriminative model in distinguishing GBM from AO in the training and validation group.

Classifier	Selection Method	Training group				Validation group			
		AUC	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity
LDA	Distance correlation	0.992	0.994	0.994	0.990	0.986	0.988	0.993	0.982
	LASSO	0.997	0.997	0.993	0.998	0.994	0.992	0.980	0.995
	GBDT	0.969	0.963	0.916	0.994	0.970	0.962	0.907	0.992
SVM	Distance correlation	0.922	0.938	1.000	0.906	0.923	0.938	1.000	0.910
	LASSO	0.831	0.868	0.972	0.826	0.817	0.831	0.935	0.798
	GBDT (over-fitting)	1.000	1.000	1.000	1.000	0.500	0.623	0.935	0.798

GBM, glioblastoma; AO, anaplastic oligodendroglioma; AUC, area under curve; LDA, linear discriminant analysis; LASSO, least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree; SVM, support vector machine.

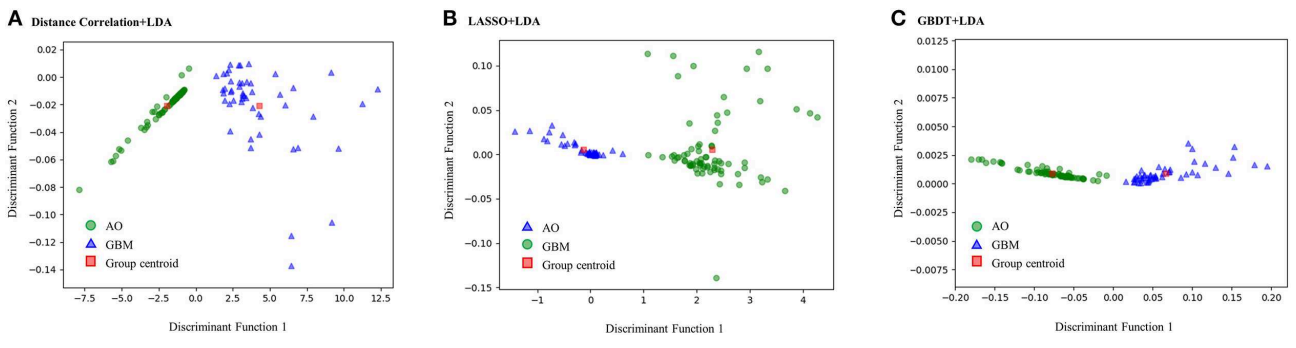


FIGURE 3 | Relationships between the canonical discriminant functions for GBM, AO and the group centroids. The two distinctive clusters formed by GBM and AO suggest three LDA-based models have excellent discriminant ability for GBM and AO. GBM, glioblastoma; AO, anaplastic oligodendroglioma; LDA, linear discriminant analysis; LASSO, least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree; SVM, support vector machine. **(A)** Canonical discriminant functions for Distance Correlation + LDA. **(B)** Canonical discriminant functions for LASSO + LDA. **(C)** Canonical discriminant functions for GBDT + LDA.

classifier which uses a straight line (a vector) to separates two classes (GBM and AO in this case), while SVM, a representative of the non-linear classifier, uses so-called support vectors to define a polynomic hyperplane to separate classes (35). In the settings of differentiating GBM and AO, our results showed LDA-based models had slightly better diagnostic performance than that of SVM-based models. However, the difference between the models was too slight to select the superior one, specifically given that all models investigated seemed to perform quite comparably and variance in AUC might be partially attributed due to the small statistical group. Therefore, limited by the small study cohort and relatively complicated methods, our results could only be regarded as hypothesis generation for future larger studies.

The results also implied that feature selection methods have impacts on diagnostic performance, especially for SVM-based models. Current feature selection methods can be categorized into three types depending on their selection mechanism: (1) Filter models select features by ranking them based on certain general characteristics such as correlation to remove irrelevant features without using any machine-learning algorithms. (2) Wrapper Models utilize a specific classifier to evaluate the quality of selected features, and offer a simple and powerful way to address the problem of feature selection, regardless of the chosen learning machine. (3) Embedded models are similar to wrapper models but embeds feature selection with classifier construction. Such models have the advantages of wrapper models-they include the interaction with the classification model, while embedded models are far less computationally intensive than wrapper models (28). In this study, we employed three selection methods as representatives of different selection mechanisms: distance correlation as representative of filter models, LASSO, and GBDT

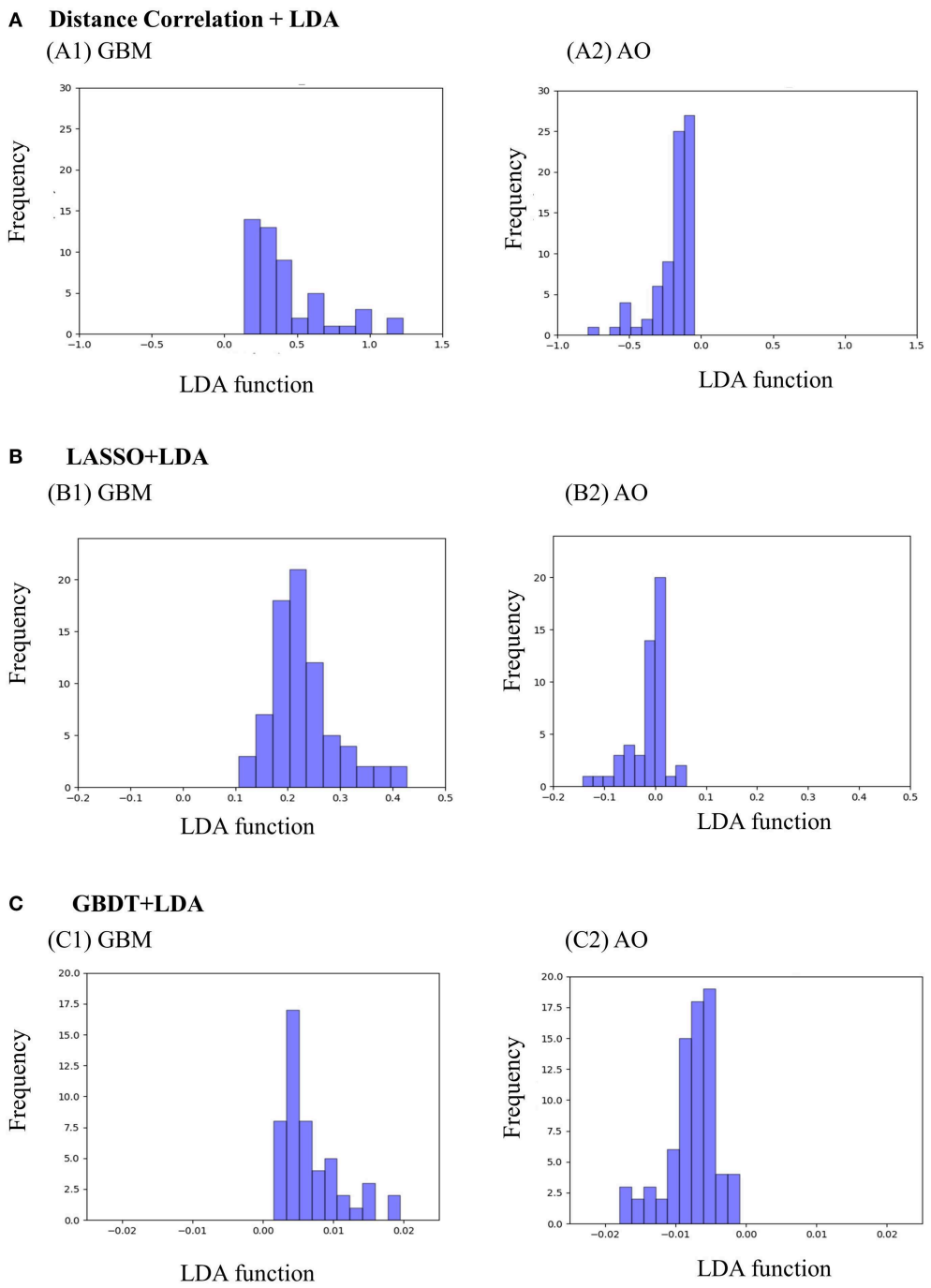


FIGURE 4 | Examples of performance of the LDA-based models in terms of the distribution of the canonical functions determined for the GBM and AO for two of the 100 independent training cycles in the MRI analysis. Minimal overlap is observed, suggesting high differential ability of the models. **(A)** Distance Correlation + LDA, **(B)** LASSO + LDA, **(C)** GBDT + LDA. GBM, glioblastoma; AO, anaplastic oligodendroglioma; AUC, area under curve; LDA, linear discriminant analysis; LASSO, least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree.

as representatives of embedded models. There was a common set of features selected by all three selection methods or two of the methods, which suggested these features might be important for the classification. For other features, it is hard to tell what extent they influenced the algorithms, since the AUCs showed minimal difference. However, even with feature selection, overfitting was still observed in one model (GBDT + SVM). We are unable to provide the exact reason but hypothesis that this model might be overly complex to be used as a discriminative tool to differentiate between GBM and AO.

Besides the comparison between machine learning models, we also performed comparison between machine and human

readers in this study. Two readers unaware of the information on the exact number of patients were asked to make diagnosis on GBM or AO based on MR images. The readers were chosen from neurosurgery department or radiology department to ensure the convincing and reliable conclusions. The results were unexpected, considering there were only two options, that the diagnostical accuracy on AO was even lower than 0.500. Specifically, AO was easily misdiagnosed as GBM in human readers' radiological assessments. The explanations from the readers were the same that GBM and AO usually represented similar patterns on MR images, and they prefer to choose GBM rather than AO in these cases due to the epidemiological reason that the incidence of GBM is much higher than AO. As we mentioned before, the accurate pre-surgical diagnosis for two types of tumor is clinical important given the differences in surgical strategy. Therefore, it is reasonable to draw the conclusion that the patients will benefit from better treatment with machine learning clinical assistances. Machine intelligence will urge the radiological practice to change dramatically. However, we should also realize that the current machine technology is far from replacing human readers, and a combination of radiologist and machine might be the best choice for the foreseeable future. Radiologists still lead the central role in diagnosis while machine only act as assistance. This combination virtually eliminates simple blunders, increases play level, and provides better insight into the decision process (36).

Our study has several limitations. Firstly, it was a retrospective single-center investigation, which may lead to a patient selection bias and limited sample size. However, at present stage, it is still unknowable how much data is required to establish a predictive model, which may be answered through empirical investigation. The number of patients enrolled in previous studies focusing on similar topic ranged from 25 to 534 (31–34, 37–39). Secondly, we did not perform subgroup analysis regarding the IDH mutation status of GBM patients. Recent studies reported machine-learning based MRI texture analysis could be used as a new method for prediction of IDH mutational status, which suggested that IDH mutational status might have bearing on texture features (37, 39, 40). Thirdly, we used conventional contrast-enhanced MRI images only and did not use other sequences or advanced imaging tools such as magnetic resonance spectroscopy (MRS). Contrast-enhanced MRI sequence was chosen in this study for the clear delineation of tumor boundaries. The combined use of other sequences or imaging tools may enable better diagnostic ability. Fourthly, models built in current study were not externally validated. Since medical centers use different MRI scanners, imaging parameters and contrast, radiomic features may change accordingly. Therefore, the efficacy of machine learning-based models in this study cannot be guaranteed for external datasets. Nevertheless, we used the open-source package to perform the image processing and texture analysis, which allows others to reproduce the texture analysis with other datasets.

CONCLUSION

In conclusion, radiomics-based machine learning enables differentiation between glioblastoma and anaplastic oligodendroglioma. Our data indicate that the performance of this approach is superior to a human reader. This method may be a valuable addition to routine clinical practice to improve GBM and AO differentiation. However, multicenter investigations including larger patient cohorts and analysis combined with other MRI sequences or imaging techniques are warranted so that this non-invasive approach can be introduced into routine clinical practice.

DATA AVAILABILITY STATEMENT

We are pleased to share our data to any qualified researchers without undue reservation. Please contact corresponding author if there is anything they need.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Sichuan University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

YF, CC, JX, and XM contributed conception and design of the study. CC and ZT enrolled eligible patients, obtained medical records, and MRI images of each patient. FZ and CC independently made diagnosis based on contrast-enhanced T1-weighted images. YF and ZT did texture analysis. JW established the models and performed other statistical analysis. YF wrote the first draft of the manuscript. CC wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01164/full#supplementary-material

Supplementary Material 1 | Original data of extracted features. LASSO, least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree.

Supplementary Material 2 | Results of Manny-Whitney *U*-test to examine the difference between two sets of data extraction.

Supplementary Material 3 | Texture features selected with three different methods. Features selected by all three methods are labeled red. Features selected by two methods are labeled in yellow. LASSO, least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree.

Supplementary Material 4 | Explanation of each texture features.

Supplementary Material 5 | ROCs of six models to differentiate between GBM and AO. GBM, glioblastoma; AO, anaplastic oligodendroglioma; LASSO, least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree; LDA, linear discriminant analysis; SVM, support vector machine.

REFERENCES

1. Wesseling P, Capper D. WHO 2016 Classification of gliomas. *Neuropathol Appl Neurobiol*. (2018) 44:139–50. doi: 10.1111/nan.12432
2. Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2011-2015. *Neuro Oncol*. (2018) 20 (Suppl. 4):iv1–86. doi: 10.1093/neuonc/noy131
3. Petrella F, Diotti C, Rimessi A, Spaggiari L. Pulmonary metastasectomy: an overview. *J Thorac Dis*. (2017) 9 (Suppl. 12):S1291–8. doi: 10.21037/jtd.2017.03.175
4. Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *J Neurosurg*. (2001) 95:190–8. doi: 10.3171/jns.2001.95.2.0190
5. Sanai N, Polley MY, McDermott MW, Parsa AT, Berger MS. An extent of resection threshold for newly diagnosed glioblastomas. *J Neurosurg*. (2011) 115:3–8. doi: 10.3171/2011.2.JNS10998
6. Simpson JR, Horton J, Scott C, Curran WJ, Rubin P, Fischbach J, et al. Influence of location and extent of surgical resection on survival of patients with glioblastoma multiforme: results of three consecutive Radiation Therapy Oncology Group (RTOG) clinical trials. *Int J Radiat Oncol Biol Phys*. (1993) 26:239–44.
7. Nabors LB, Portnow J, Ammirati M, Baehring J, Brem H, Butowski N, et al. NCCN guidelines insights: central nervous system cancers, version 1.2019. *J Natl Compr Canc Netw*. (2019) 15:1331–45. doi: 10.6004/jnccn.2017.0166
8. Thomas AA, Abrey LE, Terziev R, Raizer J, Martinez NL, Forsyth P, et al. Multicenter phase II study of temozolomide and myeloablative chemotherapy with autologous stem cell transplant for newly diagnosed anaplastic oligodendroglioma. *Neuro Oncol*. (2017) 19:1380–90. doi: 10.1093/neuonc/nox086
9. Jenkinson MD, Du Plessis DG, Walker C, Smith TS. Advanced MRI in the management of adult gliomas. *Br J Neurosurg*. (2007) 21:550–61. doi: 10.1080/02688690701642020
10. Rees J. Advances in magnetic resonance imaging of brain tumours. *Curr Opin Neurol*. (2003) 16:643–50. doi: 10.1097/00019052-200312000-00001
11. Baig MA, Klein JP, Mechtler LL. Imaging of brain tumors. *Continuum*. (2016) 22:1529–52. doi: 10.1212/CON.0000000000000388
12. Engelhard HH, Stelea A, Mundt A. Oligodendroglioma and anaplastic oligodendroglioma: clinical features, treatment, and prognosis. *Surg Neurol*. (2003) 60:443–56. doi: 10.1016/S0090-3019(03)00167-8
13. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging*. (2017) 44:151–65. doi: 10.1007/s00259-016-3427-0
14. Hainc N, Stippich C, Stieltjes B, Leu S, Bink A. Experimental texture analysis in glioblastoma: a methodological study. *Invest Radiol*. (2017) 52:367–73. doi: 10.1097/RLI.0000000000000354
15. Ahn SY, Park CM, Park SJ, Kim HJ, Song C, Lee SM, et al. Prognostic value of computed tomography texture features in non-small cell lung cancers treated with definitive concomitant chemoradiotherapy. *Invest Radiol*. (2015) 50:719–25. doi: 10.1097/RLI.0000000000000174
16. Pickles MD, Lowry M, Gibbs P. Pretreatment prognostic value of dynamic contrast-enhanced magnetic resonance imaging vascular, texture, shape, and size parameters compared with traditional survival indicators obtained from locally advanced breast cancer patients. *Invest Radiol*. (2016) 51:177–85. doi: 10.1097/RLI.0000000000000222
17. Deo RC. Machine learning in medicine. *Circulation*. (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
18. Blumenthal DT, Artzi M, Liberman G, Bokstein F, Aizenstein O, Ben Bashat D. Classification of high-grade glioma into tumor and nontumor components using support vector machine. *AJNR Am J Neuroradiol*. (2017) 38:908–14. doi: 10.3174/ajnr.A5127
19. Citak-Er F, Firat Z, Kovanlikaya I, Ture U, Ozturk-Isik E. Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3T. *Comput Biol Med*. (2018) 99:154–60. doi: 10.1016/j.compbiomed.2018.06.009
20. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with deep neural networks. *Med Image Anal*. (2017) 35:18–31. doi: 10.1016/j.media.2016.05.004
21. Kunimatsu A, Kunimatsu N, Yasaka K, Akai H, Kamiya K, Watadani T, et al. Machine learning-based texture analysis of contrast-enhanced MR imaging to differentiate between glioblastoma and primary central nervous system lymphoma. *Magn Reson Med Sci*. (2019) 18:44–52. doi: 10.2463/mrms.mp.2017-0178
22. Papp L, Pötsch N, Grahovac M, Schmidbauer V, Woehrer A, Preusser M, et al. Glioma survival prediction with combined analysis of *in vivo* (11)C-MET PET features, *ex vivo* features, and patient features by supervised machine learning. *J Nucl Med*. (2018) 59:892–9. doi: 10.2967/jnumed.117.202267
23. Ranjith G, Parvathy R, Vikas V, Chandrasekharan K, Nair S. Machine learning methods for the classification of gliomas: initial results using features extracted from MR spectroscopy. *Neuroradiol J*. (2015) 28:106–11. doi: 10.1177/1971400915576637
24. van den Bent MJ, Smits M, Kros JM, Chang SM. Diffuse infiltrating oligodendroglioma and astrocytoma. *J Clin Oncol*. (2017) 35:2394–401. doi: 10.1200/JCO.2017.72.6737
25. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res*. (2018) 78:4786–9. doi: 10.1158/0008-5472.CAN-18-0125
26. Koller D, Sahami M. Toward optimal feature selection. In: *Proceedings of the 13th International Conference on Machine Learning (ICML-1996)*. Bari (2000). p. 96.
27. Dash M, Liu H. Feature selection for classification. *Intell Data Anal*. (1997) 1:131–56. doi: 10.3233/IDA-1997-1302
28. Tang J, Alelyani S, Liu H. Feature selection for classification: a review. *Data Classification*: Algorithms and Applications, 2014:37–64.
29. Dellacasa Bellingegni A, Gruppioni E, Colazzo G, Davalli A, Sacchetti R, Guglielmelli E, et al. NLR, MLP, SVM, and LDA: a comparative analysis on EMG data from people with trans-radial amputation. *J Neuroeng Rehabil*. (2017) 14:82. doi: 10.1186/s12984-017-0290-6
30. Sakai K, Yamada K. Machine learning studies on major brain diseases: 5-year trends of 2014–2018. *Jpn J Radiol*. (2019) 37:34–72. doi: 10.1007/s11604-018-0794-4
31. Inano R, Oishi N, Kunieda T, Arakawa Y, Yamao Y, Shibata S, et al. Voxel-based clustered imaging by multiparameter diffusion tensor images for glioma grading. *Neuroimage Clin*. (2014) 5:396–407. doi: 10.1016/j.nicl.2014.08.001
32. Zhang X, Yan LF, Hu YC, Li G, Yang Y, Han Y, et al. Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features. *Oncotarget*. (2017) 8:47816–30. doi: 10.18632/oncotarget.18001
33. Li-Chun Hsieh K, Chen CY, Lo CM. Quantitative glioma grading using transformed gray-scale invariant textures of MRI. *Comput Biol Med*. (2017) 83:102–8. doi: 10.1016/j.compbiomed.2017.02.012
34. Fetit AE, Novak J, Rodriguez D, Auer DP, Clark CA, Grundy RG, et al. Radiomics in paediatric neuro-oncology: a multicentre study on MRI texture analysis. *NMR Biomed*. (2018) 31:e3781. doi: 10.1002/nbm.3781
35. Gerhardt N, Schwolow S, Rohn S, Pérez-Cacho PR, Galán-Soldevilla H, Arce L, et al. Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: comparison of different processing approaches by LDA, kNN, and SVM. *Food Chem. (2019)* 278:720–8. doi: 10.1016/j.foodchem.2018.11.095
36. Dreyer KJ, Geis JR. When machines think: radiology's next frontier. *Radiology*. (2017) 285:713–8. doi: 10.1148/radiol.2017171183
37. Bisdas S, Shen H, Thust S, Katsaros V, Stranjalis G, Boskos C, et al. Texture analysis- and support vector machine-assisted diffusional kurtosis imaging may allow in vivo gliomas grading and IDH-mutation status prediction: a preliminary study. *Sci Rep*. (2018) 8:6108. doi: 10.1038/s41598-018-24438-4
38. Hu LS, Ning S, Eschbacher JM, Gaw N, Dueck AC, Smith KA, et al. Multi-parametric MRI and texture analysis to visualize spatial histologic

heterogeneity and tumor extent in glioblastoma. *PLoS ONE*. (2015) 10:e0141506. doi: 10.1371/journal.pone.0141506

39. Jakola AS, Zhang YH, Skjulsvik AJ, Solheim O, Bø HK, Berntsen EM, et al. Quantitative texture analysis in the prediction of IDH status in low-grade gliomas. *Clin Neurol Neurosurg*. (2018) 164:114–20. doi: 10.1016/j.clineuro.2017.12.007
40. Han L, Wang S, Miao Y, Shen H, Guo Y, Xie L, et al. MRI texture analysis based on 3D tumor measurement reflects the IDH1 mutations in gliomas - a preliminary study. *Eur J Radiol*. (2019) 112:169–79. doi: 10.1016/j.ejrad.2019.01.025

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 14 November 2019
doi: 10.3389/fonc.2019.01203

Preoperative Prediction of Microvascular Invasion of Hepatocellular Carcinoma: Radiomics Algorithm Based on Ultrasound Original Radio Frequency Signals

Yi Dong[1†], Qing-Min Wang[2†], Qian Li[3], Le-Yin Li[2], Qi Zhang[1], Zhao Yao[2], Meng Dai[2], Jinhua Yu[2] and Wen-Ping Wang[1*]*

[1] Zhongshan Hospital, Fudan University, Shanghai, China, [2] Department of Electronic Engineering, Fudan University, Shanghai, China, [3] Harvard Medical School, Massachusetts General Hospital, Boston, MA, United States

OPEN ACCESS

Edited by:
Rong Tian,
Sichuan University, China

Reviewed by:
Zhi-Cheng Li,
Shenzhen Institutes of Advanced Technology (CAS), China
Di Dong,
Institute of Automation (CAS), China

****Correspondence:***
Jinhua Yu
jhyu@fudan.edu.cn
Wen-Ping Wang
puguang61@126.com

[†]These authors share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *23 July 2019*
Accepted: *23 October 2019*
Published: *14 November 2019*

Citation:
Dong Y, Wang Q-M, Li Q, Li L-Y, Zhang Q, Yao Z, Dai M, Yu J and Wang W-P (2019) Preoperative Prediction of Microvascular Invasion of Hepatocellular Carcinoma: Radiomics Algorithm Based on Ultrasound Original Radio Frequency Signals. Front. Oncol. 9:1203. doi: 10.3389/fonc.2019.01203

Background: To evaluate the accuracy of radiomics algorithm based on original radio frequency (ORF) signals for prospective prediction of microvascular invasion (MVI) in hepatocellular carcinoma (HCC) lesions.

Methods: In this prospective study, we enrolled 42 inpatients diagnosed with HCC from January 2018 to December 2018. All HCC lesions were proved by surgical resection and histopathology results, including 21 lesions with MVI. Ultrasound ORF data and grayscale ultrasound images of HCC lesions were collected before operation for further radiomics analysis. Three ultrasound feature maps were calculated using signal analysis and processing (SAP) technology in first feature extraction. The diagnostic accuracy of model based on ORF signals was compared with the model based on grayscale ultrasound images.

Results: A total of 1,050 radiomics features were extracted from ORF signals of each HCC lesion. The performance of MVI prediction model based on ORF was better than those based on grayscale ultrasound images. The best area under curve, accuracy, sensitivity, and specificity of ultrasound radiomics in prediction of MVI were 95.01, 92.86, 85.71, and 100%, respectively.

Conclusions: Radiomics algorithm based on ultrasound ORF data combined with SAP technology can effectively predict MVI, which has potential clinical application value for non-invasively preoperative prediction of MVI in HCC patients.

Keywords: hepatocellular carcinoma, microvascular invasion, prediction, radiomics analysis, original radio frequency signals

INTRODUCTION

Hepatocellular carcinoma (HCC) is the third leading cause of cancer-related death worldwide and the first leading cancer in East Asia (1). Resection is the most commonly used treatment for patients with early stage HCC. However, recurrence within 2 years after surgery still occurs in 30–50% of patients, which becomes the major cause of mortality (2). The early recurrence of HCC has been

found to be associated with the microvascular invasion (MVI) of tumor emboli in close proximity to the primary HCC (3). MVI was proved to be an important factor not only for predicting early recurrence but also for assessing long-term patient survival (4). The presence of MVI is a histopathological indication of aggressive behavior of HCC (5), especially in the first 2 years after liver resection and transplantation (3). Both univariable and multivariable analyses revealed that MVI was independently associated with poorer overall survival rate and recurrence-free survival rate after partial hepatectomy for HCC (6). Accurate and successful preoperative assessment of MVI in patients with HCC may be helpful to make appropriate clinical management strategy, and finally, to improve survival rate of HCC patients.

At present, MVI could only be diagnosed by surgical pathology after operations and was reportedly presented in 15.0–57.1% HCC surgical specimens (5, 7). Some studies have made persistent endeavors toward the preoperative prediction of MVI (8–10). Several radiological features on contrast-enhanced magnetic resonance imaging (MRI) and computed tomography (CT) images, such as tumor margin, internal arteries, and hypodense halos, were found to be associated with MVI (11–13). However, MR or CT imaging has limitations for predicting the tumor MVI in HCC (14, 15). The reported sensitivity and specificity of preoperative prediction of MVI in HCC lesions based on contrast-enhanced CT were only 81.7 and 88.1%, respectively (16). The results of MRI showed that the mismatch between diffusion-weighted imaging (DWI) and T2-weighted imaging of regions was an independent predictor of MVI, with higher specificity (95.65%) but less sensitivity (18.18%) (14, 15). In addition, it is difficult to predict MVI in small tumors; the imaging predictors such as internal arteries and hypodense halos were not frequently observed in small tumors (8). Up till now, there is still debate about the best imaging predictive feature of MVI in HCC (11–13).

Recently, radiomics analysis based on ultrasound imaging (RA-USI) technology has achieved some good results in the early diagnosis, prognosis, and prediction of diseases (17–19). The accuracy of grading diagnosis of liver cirrhosis using RA-USI was proved to be more accurate than that of traditional ultrasound elastography technology (20). In our previous study, we also confirmed that the multiparametric ultrasound model based RA-USI achieved a good performance with mean AUC values of 0.78–0.85 (20). However, current radiomics analysis was based on conventional ultrasound images; it faced some limitations, such as influence of standardization of ultrasound images, diversity of electronic characteristics caused by different ultrasound equipment, and speckle noise of different ultrasound equipment (19).

To improve the diagnosis and treatment efficiency, original image with abundant signal information might be necessary. Comparing with conventional ultrasound images, ultrasound original radio frequency (ORF) signal is not affected by postprocess such as brightness compensation, envelope detection, depth compensation, or dynamic range adjustment (21). ORF contains all the acoustic information, including attenuation, scattering, sound speed, phase, and so on, which might provide more abundant tissue information than conventional ultrasound images (22). ORF signal would only be related to the physical transmitting and receiving mechanism of imaging equipment (23). Therefore, ORF signal contains more abundant macro- and microtissue information than conventional ultrasound images (24). It is expected to obtain higher stability and consistency in further radiomics analysis process.

In this study, we aimed to investigate the value of radiomics algorithm based on ultrasound ORF data (RA-ORF) in preoperative detection of MVI in HCC patients.

Abbreviations: ORF, Original radio frequency signals; MVI, Microvascular invasion; HCC, Hepatocellular carcinoma; RA-ORF, Radiomics analysis method based on ultrasound original radio frequency signal; ROI, Region of interest; SR, Sparse representation; SVM, Support vector machine; LOOCV, Leave-one-out cross-validation; DEA, Direct energy attenuation; OND, Omega of Nakagami distribution; SDSD, Standard deviation of spectrum difference; SAP, Signal analysis and processing; DM, Microvascular invasion prediction model based on direct energy attenuation; DOM, Microvascular invasion prediction model based on direct energy attenuation and omega of Nakagami distribution; DOSM, Microvascular invasion prediction model based on direct energy attenuation, omega of Nakagami distribution and standard deviation of spectrum difference; AUC, Area under curve; MRI, Magnetic resonance imaging; CT, Computed tomography; RA-USI, Radiomics analysis based on conventional ultrasound image; ANOVA, Analysis of variance; ROC, Receiver operating characteristic curve; PRC, Precision recall curve; DCA, Decision curve analysis.

MATERIALS AND METHODS

Patients

From January 2018 to December 2018, patients preoperatively diagnosed with HCC in a single institution were enrolled. The inclusion criteria were (1) adult patients suspected to be primary HCC by imaging methods and planned to accept surgery in our hospital; (2) solitary tumor; (3) all patients accepted preoperative grayscale ultrasound examinations within 1 week before surgery; (4) HCC lesions located in the right lobe of liver; and (5) all cases were confirmed by histopathological examination and MVI evaluation.

Exclusion criteria included the following: (1) target HCC lesion not clearly visible on the grayscale ultrasound scan; (2) patients with preoperative biopsy or adjuvant therapy (radio frequency therapy, chemotherapy, targeted therapy, etc.); (3) incomplete clinical or histopathological data; and (4) patients with HCC larger than 5 cm in maximum diameter, since such tumors are known to have a greater risk of MVI.

Final Diagnosis

The final histopathological results including MVI grade were the gold standard for our current study. According to the practice guidelines of Chinese Society of Pathology, MVI was defined based on the number of cells that can be found in the endothelial vascular lumen under microscopy. MVI were divided into three additional subgrades, including M0, no MVI; M1 (the low-risk group), $\leq$5 MVI in adjacent liver tissue $\leq$1 cm away from the tumor; and M2 (the high-risk group), $>$5 MVI or MVI in liver tissue $>$1 cm away from the tumor (25).

Two pathologists with at least 10 years of experience in hepatic pathology reviewed all the specimen slices. Both investigators

were blinded to the clinical and imaging information of the patients. In cases of discordance, a consensus reading was performed.

Ultrasound Imaging Procedure and ORF Data Acquisition

All patients fasted for at least 8 h before ultrasound examinations. The grayscale ultrasound examinations of the hepatic lesions were performed according to the standardized protocol. Ultrasound examinations were performed by a single experienced radiologist (more than 18 years' experience of liver ultrasound scan), who was aware of the patients' clinical history. All ultrasound examinations were performed with an EPIQ-7 ultrasound system certificated with ORF data (Philips Medical Company). A C5-1 curved transducer (1–5 MHz) was used for data acquisition.

First, conventional grayscale ultrasound scan was performed. After a clear ultrasound image of tumor was obtained, the process of ORF data acquisition was started. We clicked the "freeze" button to freeze the grayscale ultrasound images and to save the current ORF data retrospectively. The corresponding conventional grayscale ultrasound images were also captured to build a comparison test for ultrasound ORF signals. Both of them would be further used to establish MVI preoperative prediction radiomics models.

ORF Data Processing and Radiomics Analysis Procedure

Overall Design

RA-ORF method was applied for MVI preoperative prediction. The radiomics analysis process consisted of the following steps: (1) to obtain grayscale image and ORF data of HCC lesions before operation; (2) tumor segmentation on gray scale ultrasound images of ORF data to obtain the ORF data from the region of interest (ROI) in the tumor; (3) first feature extraction to obtain three ultrasound feature maps of ORF data of ROI; (4) second feature extraction to obtain radiomics features from three ultrasound feature maps and related grayscale ultrasound images; (5) feature selection based on sparse representation (SR) algorithm (19); and (6) train support vector machine (SVM) classifier with the features sorted in step (5) to achieve further feature selection and dimension reduction, and predict MVI in patients with HCC (**Figure 1**).

The radiomics analysis based on ultrasound ORF signal (RA-ORF) method will be built on three ultrasound feature parameters, including direct energy attenuation (DEA), omega of Nakagami distribution (OND), and standard deviation of spectrum difference (SDSD). Leave-one-outcross-validation (LOOCV) was employed to validate the trained model.

Conventional grayscale ultrasound images will be used as the control group. The radiomics analysis for conventional ultrasound images processing included tumor segmentation, feature extraction, feature selection, and classification preoperative prediction.

All images and data were processed on MATLAB R2014b (Math Works, Inv., Natick, MA, USA).

Tumor Segmentation

For conventional grayscale ultrasound images obtained from the first step of "data acquisition," the ROIs were marked by an ultrasound doctor as four white forks points; then, the grayscale data of the tumor could obtained from the conventional grayscale ultrasound images by segmenting along those markers (**Figure 2A**).

For ORF data matrix, they were drawn directly in columns called scan-line way (**Figure 2B**). Data were covered with the whole picture. It is different from **Figure 2A**, which had values of 0 outside the sector area. Adding Hilbert transform and logarithmic compression to **Figure 2B**, we could get the grayscale ultrasound images in scan-line way, which clearly showed the location of the tumor. Then, segmentation was processed to obtain the location of ROI and get the ROI's ORF data. The shapes of ROI were stretched laterally at a shallow position. ROI segmented by an ultrasound doctor was used to ensure the accuracy of segmentation (**Figure 2C**).

First Feature Extraction

Feature extraction of multiparameter ultrasound features was the key step of the RA-ORF method. Three kinds of ultrasound feature parameters of ORF included time domain feature, frequency domain feature, and statistical feature and were applied.

In the first feature extraction, ORF data of ROI was used to calculate three ultrasound feature parameters and further form the corresponding three ultrasound feature maps. Three ultrasound feature maps, including DEA feature map (time-domain feature), SDSD feature map (frequency-domain feature), and OND feature map (statistical feature), were established and saved in *.bmp formats (**Figure 3**).

Second Feature Extraction

Second feature extraction were based on ROIs of conventional grayscale ultrasound images and the ROIs of three ultrasound feature maps obtained from ORF data. Each image can get 70 texture features: 16 features of histogram, 23 features based on gray-level co-occurrence matrix (26), 13 features based on gray-level run-length matrix (27), 13 features based on gray-level size-zone matrix (28), and five features based on neighborhood gray-tone difference matrix (29). Summary of the 70 texture features was listed in the feature extraction section of the **Appendix**. Then, the wavelet transformation to strip the image information layer-upon-layer by high- and low-pass filters were performed. Thereafter, four images of different frequency sub-bands and another 280 texture features could be obtained. Finally, we obtained 350 texture features from each grayscale ultrasound image and ultrasound feature maps.

Feature Selection and Dimension Reduction

Iterative SR method were used to select key features for the classifier before classification to improve the stability of final models (30, 31). The SR coefficients of each feature were calculated by selecting part of the 42 samples in each iteration. In the SR method, the threshold Tal is set to 0.004. Then, the

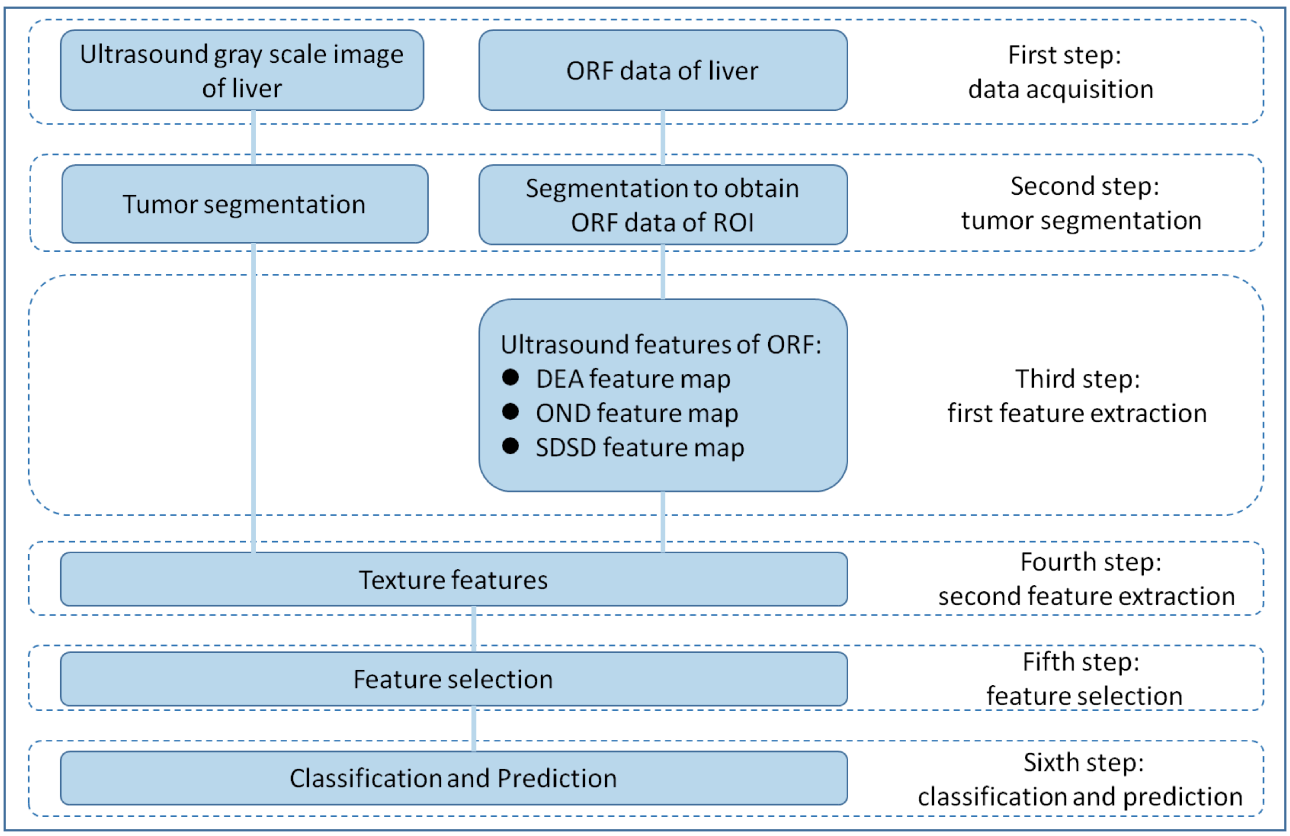


FIGURE 1 | Overall design of radiomics analysis. The radiomics analysis process consisted of the following steps: (1) grayscale images and original radio frequency (ORF) data of HCC lesions obtained; (2) tumor segmentation on grayscale ultrasound images for ORF data; (3) first feature extraction to obtain three ultrasound feature maps of ORF data of region of interest (ROI); (4) second feature extraction to obtain radiomics features from three ultrasound feature maps and related grayscale ultrasound images; (5) feature selection based on sparse representation (SR) algorithm; and (6) support vector machine (SVM) classifier trained with the selected features for MVI prediction.

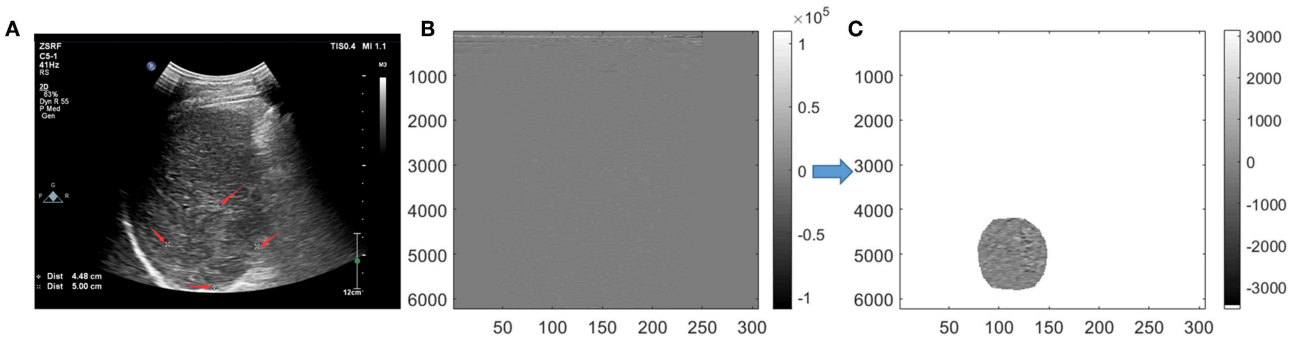


FIGURE 2 | Tumor segmentation. On grayscale ultrasound image, the region of interest (ROI) was manually marked by a doctor with four white forks points **(A)**. Image were segmented in scan-line imaging way for original radio frequency (ORF) data **(B)**. After Hilbert transform and logarithmic compression of ORF signals, the grayscale ultrasound images under the scan-line images could be obtained **(C)**.

average SR coefficients of each feature were taken as the final SR coefficients of each feature. The importance of the features was quantified as SR coefficients. Finally, the features were sorted based on the absolute value of the final SR coefficients, and features that did not meet the threshold Tal condition were remove to achieve feature dimensionality reduction. A detailed description of SR method in feature selection is included in the feature selection section of the **Supplementary Appendix**.

Classification and Prediction

SVM classifier was used in this section. Starting from number 1, the different numbers of features ranked by SR method

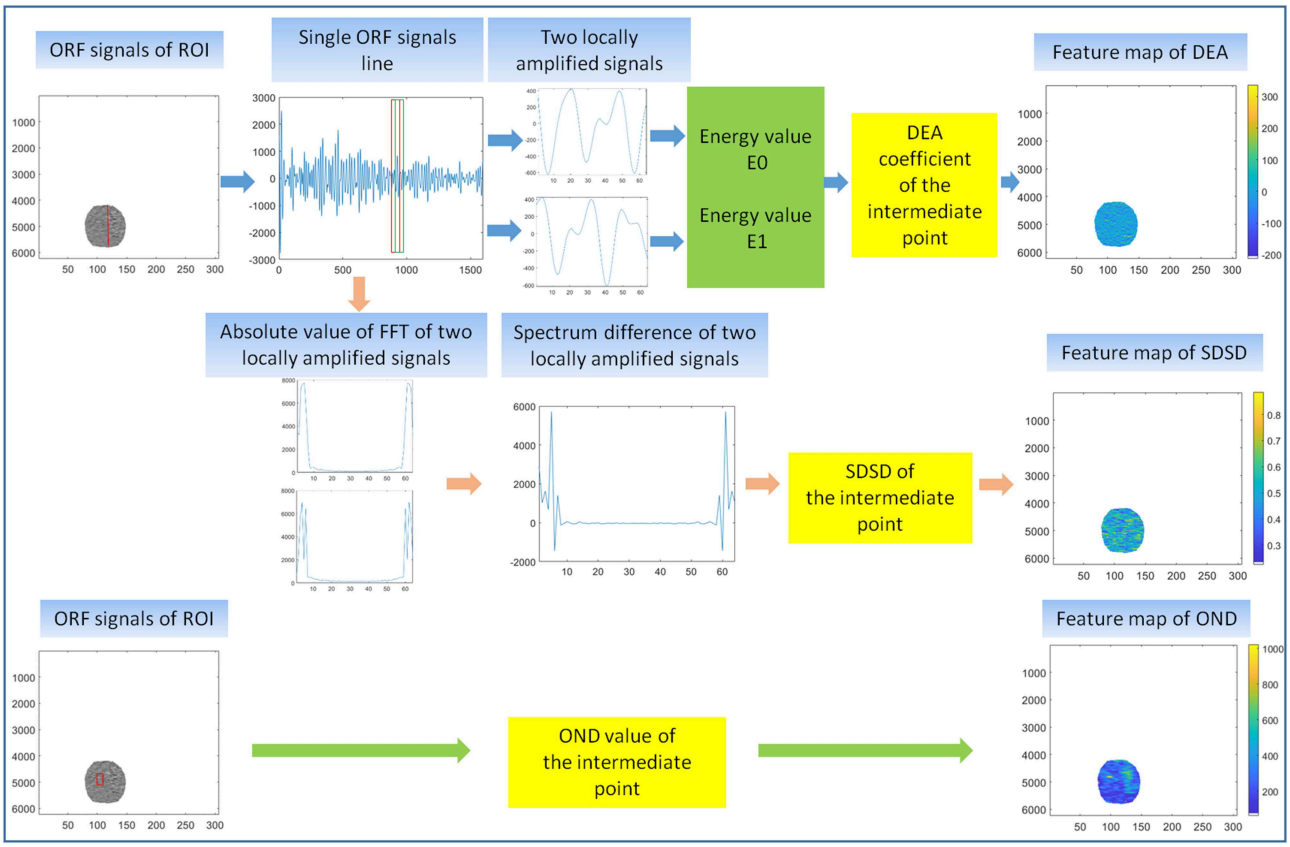


FIGURE 3 | First feature extraction. The calculation principle of original radio frequency signal (ORF) signals from region of interest (ROI) in three ultrasound feature maps, including direct energy attenuation (DEA) feature map, standard deviation of spectrum difference (SDSD) feature map, and omega of Nakagami distribution (OND) feature map.

in *Feature Selection and Dimension Reduction* were put into the SVM classifier to calculate AUC, accuracy, sensitivity, and specificity of MVI prediction in patients with HCC. We evaluated the MVI prediction models through the above parameters. The final feature dimensions of the MVI prediction models were the number of features put into the SVM classifier with the best performance in MVI prediction. This process effectively realized dimension reduction of features. Feature selection is mainly based on sparse representation, but the dimensions of features are still high after sparse representation. When implementing the classifier, the SVM uses the kernel function mapping technique to obtain the same classification result as the high-dimensional space in the low-dimensional space. In this sense, the SVM implements the further selection of features.

Statistical Analysis

Descriptive statistics are summarized as the mean ± SD. LOOCV statistical analysis method was used to evaluate the MVI prediction models. A Tukey test, in conjunction with analysis of variance (ANOVA), was used to test the signification between any two pairs of the three ultrasound features. Receiver operating characteristic curve (ROC), precision–recall curve (PRC), and model decision curve analysis (DCA) were employed to show the overall performance of the models. Other assessment indicator included area under the ROC (AUC), accuracy, sensitivity, and specificity.

RESULTS

Final Diagnosis of Patients

A total of 42 HCC patients (34 men and 8 women; age range, 23–80 years; mean, 58.5 ± 11.9 years) were finally included in our study. The surgical procedures comprised segmentectomy (n = 12), right anterior sectionectomy (n = 19), and right posterior sectionectomy (n = 11). The mean time between ultrasound scan and surgery was 6 days (range, 3–7 days).

Pathology data revealed the presence of MVI in 21 HCC patients as grade 1 (M1), and 21 patients were diagnosed without MVI as grade 0 (M0).

Multiparameter Ultrasound Feature Extraction Results of ORF Signals

Multiple ultrasound parameters were extracted from ORF signals, including DEA, OND, and SDSD. They played various degrees of positive role in the MVI preoperative prediction.

Compared with the M0 group, the M1 group showed larger absolute value of DEA and more serious attenuation. ANOVA analysis showed significant difference in DEA, OND, and SDSD between patient with and without MVI ($P < 0.05$).

Second Feature Extraction and Feature Selection Results

Four pictures were included in our second feature extraction results, including grayscale ultrasound image, DEA feature map, OND feature map, and SDSD feature map. The MVI prediction model based on ultrasound grayscale image was referred to as GM. The MVI prediction model based on DEA feature map was referred to as DM. The MVI prediction model based on DEA feature map and OND feature map was referred to as DOM. The MVI prediction model based on DEA feature map, OND feature map, and SDSD feature map was referred to as DOSM.

In this texture feature extraction, we extracted 350 texture features from MVI prediction model of GM, 350 texture features from DM, 700 texture features from DOM, and 1,050 texture features from DOSM. The number of selected features of GM, DM, DOM, and DOSM MVI prediction model based SR method were 214, 253, 427, and 536, respectively.

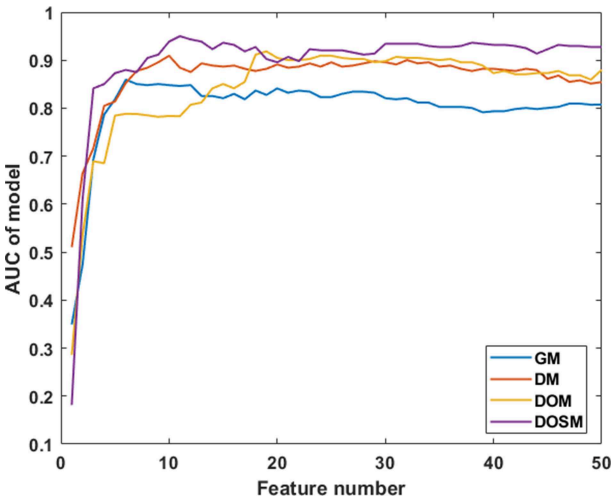


FIGURE 4 | Diagnostic performances of MVI prediction models with different number of features. After feature selection, the performance of DOSM, DOM, DM, and GM models were increased gradually and maintained at a relative stable level. The changes in AUC with the increase in feature numbers were helpful to find the optimal feature dimensions of each model. The final feature dimensions of MVI prediction models of GM, DM, DOM, and DOSM were 6, 10, 19, and 11, respectively.

Diagnostic Performances of Different MVI Prediction Models

In SVM classifier to construct MVI prediction model, the training process of the above-mentioned model achieved further feature dimensionality reduction. **Figure 4** used top 50 features after feature selection to show the performance of models utilizing different number of features. According to **Figure 4**, the final feature dimensions of MVI prediction models of GM, DM, DOM, and DOSM were 6, 10, 19, and 11, respectively. The maximum accuracy of the corresponding above four models by dimension reduction were 83.33, 85.71, 88.1, and 92.86%.

Table 1 shows the performance parameters of the GM model based on the conventional grayscale ultrasound images and the other three models based on the ORF signals. GM based on grayscale ultrasound image was used as a comparison test to the three MVI prediction models based on ORF signals. The AUC, accuracy, sensitivity, and specificity of GM were the lowest among the four MVI prediction models of GM, DM, DOM, and DOSM, respectively. Among the three ORF-based prediction models, the accuracy, AUC, sensitivity, and specificity of the DOSM were the highest. In the 11 selected features of DOSM, 6 features were obtained from the DEA ultrasound feature map, three features from the OND ultrasound feature map, and two features from the SDSD ultrasound feature map.

The AUC of DOSM (95.01%, 0.835–0.993) was the highest one among the four prediction models. The AUC of GM (85.94%, 0.717–0.947) was the lowest (**Figure 5**).

Precision recall curves (PRC) of DOSM, DOM, DM, and GM are shown in **Figure 6**. The results showed that DOSM based on three ultrasound feature maps selected from ORF signals had more advantage compared with the other three models in predicting the MVI classification of HCC.

DISCUSSION

Previously, several studies proved that radiomics analysis algorithm based on ultrasound images could be helpful to extract massive features and to assist clinical decision-making. The reported ultrasound radiomics analysis algorithm

TABLE 1 | Diagnostic performance of DOSM, DM, DOM, and GM for MVI classification.

Model type	AUC (%, 95% CI)	Accuracy (%)	Sensitivity (%)	Specificity (%)
DOSM	95.01 (0.835–0.993)	92.86	85.71	100
DOM	91.84 (0.792–0.980)	88.1	80.95	95.24
DM	90.93 (0.780–0.976)	85.71	80.95	90.48
GM	85.94 (0.717–0.947)	83.33	80.95	85.71

AUC, area under the receiver operating characteristic curve; DOSM, MVI prediction model based on DEA feature map, OND feature map and SDSD feature map of ORF signals; DM, MVI prediction model based on DEA feature map; DOM, MVI prediction model based on DEA feature map and OND feature map; GM, MVI prediction model based on gray-scale ultrasound image.

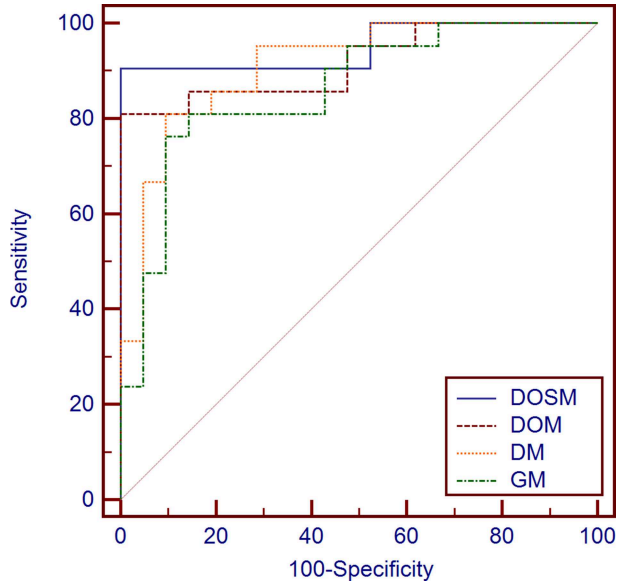


FIGURE 5 | Diagnostic performances of different MVI prediction models. While comparing the AUC curves between DOSM, DOM, DM, and GM models. The AUC of DOSM (95.01%, 0.835–0.993) was the highest one among the DOSM, DOM, DM, and GM models. The AUC of GM (85.94%, 0.717–0.947) was the lowest. The AUC of DOSM is 0.95 ± 0.04, which is the highest one among the four MVI prediction models.

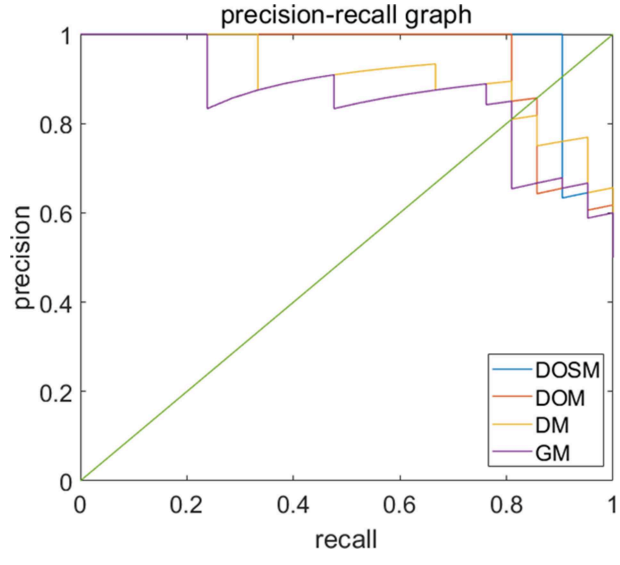


FIGURE 6 | Precision-recall curves of models for prediction of MVI. Precision–recall curves (PRC) of GM, DM, DOM, and DOSM models for prediction of microvascular invasion (MVI). The DOSM based on three ultrasound feature maps selected from radio frequency signals (ORF) signals showed the best performance among GM, DM, DOM, and DOSM models in predicting the MVI classification of HCC.

based on grayscale ultrasound images, ultrasound elastography images, and contrast enhanced ultrasound images (19, 32–34). With the development of radiomics analysis, a large number of valuable features could be extracted from conventional ultrasound images, including texture features, morphological features, and some other specific features (35, 36). However, conventional ultrasound images might be affected by post-processing procedure; as a result, they will lose a lot of useful information compared with ORF signals (21–24). The radiomics analysis technology based on ORF data was applied in our present study. We extracted three ultrasound feature maps of ORF signal of HCC lesions, combining with the iterative SR method and SVM classifiers to reduce the feature dimensions and build MVI prediction model. In our results, 11 highly correlated radiomics features were finally obtained to establish an effective MVI prediction model of DOSM. DOSM prediction model based on RA-ORF showed superior performance for MVI prediction, which make full use of the advantages of signal processing technology. It could extract more useful radiomics features and improve the accuracy of MVI classification.

Previously, several studies tried to classify diseases by ORF signal combined with radiomics analysis (34, 37) to prove that time-domain features (38), statistical distribution features, and frequency-domain features (39) of ultrasound ORF signals be helpful in disease recognition (40). In signal processing, the ultrasound feature parameters of DEA, SDSD, and OND, which were obtained from ORF signals in time, frequency, and statistics domains, always have clear and valuable physical significance. DEA of time-domain characteristics of ORF signals represents the direct energy attenuation in ROI. When the normal tissue changes, its microstructure will change accordingly, which leads to the change in attenuation. SDSD of frequency-domain characteristics of ORF signals represents standard deviation of spectrum difference, which is a common parameter to reflect spectrum differences between tissues in spectrum analysis. OND of statistical characteristics of ORF signals represents omega of Nakagami distribution of ROI. The parameter values of Nakagami distribution for the second harmonic envelope signals from different degrees of non-linearity in tissue are significantly different. According to this, we can quantitatively analyze the difference in non-linear characteristics between normal and diseased biological tissue (41). At present, advanced radiomics method makes it possible to extract huge amounts of features and to select valuable features from multiclass ultrasound feature maps consisting of DEA, SDSD, and OND. In our results, ROC and PRC curves both validated the reliability of DOSM model in MVI prediction of HCC lesions. Our RA-ORF method combined ORF-based signal processing technology with radiomics analysis, which showed a good classification performance on MVI prediction. Among the three ORF-based prediction models, the accuracy, AUC, sensitivity, and specificity were gradually improved. Some valuable radiomics features were further extracted for MVI prediction. Meanwhile, the performance of MVI prediction models in HCC lesions was improved accordingly. The radiomics algorithm based on ORF signal was superior to that based on conventional grayscale ultrasound images.

Pathologically, MVI is defined as the presence of micrometastatic HCC emboli within the vessels of the liver (9). Relevant studies have shown that there is a correlation

between tissue microstructures and spectrum feature (42). Spectrum analysis based on ORF signals can obtain abundant microstructural information, which might be completely lost in conventional grayscale ultrasound images (21–24, 42). Therefore, by extracting frequency-domain features and combining radiomics analysis, different pathological tissues could be analyzed. The presence of MVI in HCC lesions may cause changes in tissue attenuation coefficient accordingly. It is possible for us to use the time-domain features of DEA calculated from radiomics analysis of ORF signals to predict MVI in HCC lesions. The DOSM prediction model based on RA-ORF in our study reached sensitivity of 85.71%, specificity of 100%, and AUC of 95.01%. It was proved to be superior to DOM, DM, and GM models. Our initial results showed that the AUC of the DM model based on RA-ORF, which uses time-domain features of DEA, was better than the GM model based on RA-USI with conventional grayscale ultrasound images.

Our study has several limitations: the patient number is relatively limited; only three ultrasound parameters of DEA, OND, and SDSD based on ORF signals were included. The stability evaluation of RA-ORF based radiomic analysis would be further improved by multicenter studies in the future.

CONCLUSION

In conclusion, radiomics algorithm based on RA-ORF and SAP technology might provide useful information for preoperative MVI prediction in HCC lesions. Depending on the unique advantages of ultrasound imaging such as real-time imaging, low cost, and no radiation exposure risk, it might be a promising method in future clinical application.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Zhongshan Hospital, Fudan University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Each author had participated sufficiently in the paper and approved the manuscript for submission. W-PW and JY contributed to this paper with conception and design of the study. YD and QL performed literature review and analysis. YD and QZ performed the clinical ultrasound scan. LY-L, ZY, and MD conducted the image analysis. YD and Q-MW contributed to drafting and critical revision and editing, and all authors gave final approval of the final version.

FUNDING

This study was Supported by National Natural Science Foundation of China (Grant Nos. 81571676, 81501471); Supported by Shanghai Municipal Science and Technology Medical Guidance Project (Grant No. 18411967200); Supported by Shanghai Municipal Science and Technology Innovation Action Plan Clinical Medicine Project (Grant No. 17411954200); Supported by Shanghai Municipal Health and Family Planning Commission Research Project (Grant No. 201840215); Supported by Shanghai Municipal Science and Technology Major Project (Grant No. 2017SHZDZX01). All the funding supported equally in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01203/full#supplementary-material

REFERENCES

1. Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*. (2018) 67:358–80. doi: 10.1002/hep.29086
2. Chan AWH, Zhong J, Berhane S, Toyoda H, Cucchetti A, Shi K, et al. Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. *J Hepatol.* (2018) 69:1284–93. doi: 10.1016/j.jhep.2018.08.027
3. Renzulli M, Buonfiglioli F, Conti F, Brocchi S, Serio I, Foschi FG, et al. Imaging features of microvascular invasion in hepatocellular carcinoma developed after direct-acting antiviral therapy in HCV-related cirrhosis. *European radiology*. (2018) 28:506–13. doi: 10.1007/s00330-017-5033-3
4. Sumie S, Kuromatsu R, Okuda K, Ando E, Takata A, Fukushima N, et al. Microvascular invasion in patients with hepatocellular carcinoma and its predictable clinicopathological factors. *Ann Surg Oncol.* (2008) 15:1375–82. doi: 10.1245/s10434-008-9846-9
5. Zhang X, Li J, Shen F, Lau WY. Significance of presence of microvascular invasion in specimens obtained after surgical treatment of hepatocellular carcinoma. *J Gastroenterol Hepatol.* (2018) 33:347–54. doi: 10.1111/jgh.13843
6. Liu J, Zhu Q, Li Y, Qiao GL, Xu C, Guo DL, et al. Microvascular invasion and positive HB e antigen are associated with poorer survival after hepatectomy of early hepatocellular carcinoma: A retrospective cohort study. *Clin Res Hepatol Gastroenterol.* (2018) 42:330–8. doi: 10.1016/j.clinre.2018.02.003
7. Lei Z, Li J, Wu D, Xia Y, Wang Q, Si A, et al. Nomogram for preoperative estimation of microvascular invasion risk in hepatitis B virus-related hepatocellular carcinoma within the milan criteria. *JAMA Surg.* (2016) 151:356–63. doi: 10.1001/jamasurg.2015.4257
8. Kim MJ, Lee M, Choi JY, Park YN. Imaging features of small hepatocellular carcinomas with microvascular invasion on gadoxetic acid-enhanced MR imaging. *Eur J Radiol.* (2012) 81:2507–12. doi: 10.1016/j.ejrad.2011.11.014
9. Rodriguez-Peralvarez M, Luong TV, Andreana L, Meyer T, Dhillon AP, Burroughs AK. A systematic review of microvascular invasion in hepatocellular carcinoma: diagnostic and prognostic variability. *Ann Surg Oncol.* (2013) 20:325–39. doi: 10.1245/s10434-012-2513-1

10. Witjes CD, Willemssen FE, Verheij J, van der Veer SJ, Hansen BE, Verhoef C, et al. Histological differentiation grade and microvascular invasion of hepatocellular carcinoma predicted by dynamic contrast-enhanced MRI. *J Magnet Resonan Imaging.* (2012) 36:641–7. doi: 10.1002/jmri.23681
11. Banerjee S, Wang DS, Kim HJ, Sirlin CB, Chan MG, Korn RL, et al. A computed tomography radiogenomic biomarker predicts microvascular invasion and clinical outcomes in hepatocellular carcinoma. *Hepatology.* (2015) 62:792–800. doi: 10.1002/hep.27877
12. Lee S, Kim SH, Lee JE, Sinn DH, Park CK. Preoperative gadoxetic acid-enhanced MRI for predicting microvascular invasion in patients with single hepatocellular carcinoma. *J Hepatol.* (2017) 67:526–34. doi: 10.1016/j.jhep.2017.04.024
13. Renzulli M, Brocchi S, Cucchetti A, Mazzotti F, Mosconi C, Sportoletti C, et al. Can current preoperative imaging be used to detect microvascular invasion of hepatocellular carcinoma? *Radiology.* (2016) 279:432–42. doi: 10.1148/radiol.2015150998
14. Yang C, Wang H, Sheng R, Ji Y, Rao S, Zeng M. Microvascular invasion in hepatocellular carcinoma: is it predictable with a new, preoperative application of diffusion-weighted imaging? *Clin Imaging.* (2017) 41:101–5. doi: 10.1016/j.clinimag.2016.10.004
15. Ahn SY, Lee JM, Joo I, Lee ES, Lee SJ, Cheon GJ, et al. Prediction of microvascular invasion of hepatocellular carcinoma using gadoxetic acid-enhanced MR and (18)F-FDG PET/CT. *Abdom Imaging.* (2015) 40:843–51. doi: 10.1007/s00261-014-0256-0
16. Ma X, Wei J, Gu D, Zhu Y, Feng B, Liang M, et al. Preoperative radiomics nomogram for microvascular invasion prediction in hepatocellular carcinoma using contrast-enhanced CT. *Eur Radiol.* (2019) 29:3595–605. doi: 10.1007/s00330-018-5985-y
17. Qiao M, Hu Y, Guo Y, Wang Y, Yu J. Breast tumor classification based on a computerized breast imaging reporting and data system feature system. *J Ultrasound Med.* (2018) 37:403–15. doi: 10.1002/jum.14350
18. Zhang Q, Xiao Y, Suo J, Shi J, Yu J, Guo Y, et al. Sonoelastomics for breast tumor classification: a radiomics approach with clustering-based feature selection on sonoelastography. *Ultrasound Med Biol.* (2017) 43:1058–69. doi: 10.1016/j.ultrasmedbio.2016.12.016
19. Yao Z, Dong Y, Wu G, Zhang Q, Yang D, Yu JH, et al. Preoperative diagnosis and prediction of hepatocellular carcinoma: radiomics analysis based on multi-modal ultrasound images. *BMC Cancer.* (2018) 18:1089. doi: 10.1186/s12885-018-5003-4
20. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut.* (2019) 68:729–41. doi: 10.1136/gutjnl-2018-316204
21. Xia C, Zhao A, Liu DC, editors. Optimized GPU framework for ultrasound B-mode imaging. In: *International Conference on Bioinformatics & Biomedical Engineering.* Chengdu (2010).
22. Abbey CK, Nguyen NQ, Insana MF. Effects of frequency and bandwidth on diagnostic information transfer in ultrasonic B-Mode imaging. *Ultrason Ferroelectr Freq Control.* (2012) 59:1115–26. doi: 10.1109/TUFFC.2012.2302
23. Snare SR, Torp H, editors. Estimating frequency dependent attenuation to improve autmatic time gain compensation in B-mode imaging. In: *Ultrasonics Symposium.* Beijing (2008). doi: 10.1109/ULTSYM.2008.0320
24. Dydenko I, Friboulet D, Gorce JM, D'Hooge J, Bijnens B, Magnin IE. Towards ultrasound cardiac image segmentation based on the radiofrequency signal. *Med Image Analy.* (2003) 7:353–67. doi: 10.1016/S1361-8415(03)00010-0
25. Cong WM, Bu H, Chen J, Dong H, Zhu YY, Feng LH, et al. Practice guidelines for the pathological diagnosis of primary liver cancer: 2015 update. *World J Gastroenterol.* (2016) 22:9279–87. doi: 10.3748/wjg.v22.i42.9279
26. Jing Z, Li GL, He SW, editors. Texture-based image retrieval by edge detection matching GLCM. In: *IEEE International Conference on High Performance Computing & Communications.* Dalian (2008).
27. Sohail ASM, Bhattacharya P, Mudur SP, Krishnamurthy S, editors. Local relative GLRLM-based texture feature extraction for classifying ultrasound medical images. In: *Electrical & Computer Engineering* (2011). doi: 10.1109/CCECE.2011.6030630
28. Gumaste PP, Jadhav DV, editors. MR image feature extraction using advanced statistical matrices. In: *International Conference on Signal Processing.* Paralakhemundi (2017). doi: 10.1109/SCOPES.2016.7955759
29. Huan Y, Curtis C, Katherine M, Daniel M. Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. *IEEE Trans Med Imaging.* (2009) 28:374–83. doi: 10.1109/TMI.2008.2004425
30. Lin D, Cao H, Calhoun VD, Wang YP. Sparse models for correlative and integrative analysis of imaging and genetic data. *J Neurosci Methods.* (2014) 237:69–78. doi: 10.1016/j.jneumeth.2014.09.001
31. Cao H, Duan J, Lin D, Yin YS, Calhoun V, Wang YP. Sparse representation based biomarker selection for schizophrenia with integrated analysis of fMRI and SNPs. *Neuroimage.* (2014) 102:220–8. doi: 10.1016/j.neuroimage.2014.01.021
32. Suo J, Zhang Q, Chang W, Shi J, Yan Z, Chen M. [Evaluation of axillary lymph node metastasis by using radiomics of dual-modal ultrasound composed of elastography and b-mode]. *Zhongguo yi liao qi xie za zhi.* (2017) 41:313–6. doi: 10.3969/j.issn.1671-7104.2017.05.001
33. Liu T, Ge X, Yu J, Guo Y, Wang Y, Wang W, et al. Comparison of the application of B-mode and strain elastography ultrasound in the estimation of lymph node metastasis of papillary thyroid carcinoma based on a radiomics approach. *Int J Comp Assisted Radiol Surg.* (2018) 13:1617–27. doi: 10.1007/s11548-018-1796-5
34. Hu HT, Wang Z, Huang XW, Chen SL, Zheng X, Ruan SM, et al. Ultrasound-based radiomics score: a potential biomarker for the prediction of microvascular invasion in hepatocellular carcinoma. *European radiology.* (2019) 29:2890–901. doi: 10.1007/s00330-018-5797-0
35. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedus L, et al. American association of clinical endocrinologists, american college of endocrinology, and associazione medici endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules—2016 update. *Endocr Pract.* (2016) 22:622–39. doi: 10.4158/EP161208.GL
36. Frates MC, Benson CB, Charboneau JW, Cibas ES, Clark OH, Coleman BG, et al. Management of thyroid nodules detected at US: society of radiologists in ultrasound consensus conference statement. *Radiology.* (2005) 237:794–800. doi: 10.1148/radiol.2373050220
37. Tan X, Zelan M, Lifen Y, Weitao Y, Zaiyi L, Changhong L. Radiomics nomogram outperforms size criteria in discriminating lymph node metastasis in resectable esophageal squamous cell carcinoma. *Eur Radiol.* (2019) 29:392–400. doi: 10.1007/s00330-018-5581-1
38. Sun Z, Wang L, Zhou Y. [Automated tissue characterization of intravascular ultrasound gray-scale images]. *Sheng wu yi xue gong cheng xue za zhi.* (2016) 33:287–302.
39. Lin CY, Yi T, Gao YZ, Zhou JH, Huang QH. Early detection and assessment of liver fibrosis by using ultrasound RF time series. *J Med Biol Eng.* (2017) 37:1–13. doi: 10.1007/s40846-017-0261-1
40. Bouhlel N, Sevestre-Ghalila S. Nakagami Markov random field as texture model for ultrasound RF envelope image. *Comp Biol Med.* (2009) 39:535–44. doi: 10.1016/j.compbiomed.2009.03.010
41. Ke-Yan WU, Yu-Feng Z, Zheng-Peng Z, Lian G, Ke-Xin Z, Jun-Hua Z, et al. Characterization for ultrasonic harmonic of tissue based on nakagami distribution. *Acta Electron Sinica.* (2018) 46:1639–43. doi: 10.3969/j.issn.0372-2112.2018.07.014
42. Liu T, Lizzi FL, Silverman RH, Kutcher GJ. Ultrasonic tissue characterization using 2-D spectrum analysis and its application in ocular tumor diagnosis. *Med Phys.* (2004) 31:1032–9. doi: 10.1118/1.1690196

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 19 November 2019
doi: 10.3389/fonc.2019.01241

CT Morphological Features Integrated With Whole-Lesion Histogram Parameters to Predict Lung Metastasis for Colorectal Cancer Patients With Pulmonary Nodules

TingDan Hu[1,2†], ShengPing Wang[1,2†], Xiangyu E[3], Ye Yuan[3], Lv Huang[3], JiaZhou Wang[3], DeBing Shi[4], Yuan Li[5], WeiJun Peng[1,2] and Tong Tong[1,2]*

[1] Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai, China, [2] Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China, [3] Department of Radiotherapy, Fudan University Shanghai Cancer Center, Fudan University, Shanghai, China, [4] Department of Colorectal Surgery, Fudan University Shanghai Cancer Center, Fudan University, Shanghai, China, [5] Department of Pathology, Fudan University Shanghai Cancer Center, Fudan University, Shanghai, China

Purpose: To retrospectively identify the relationships between both CT morphological features and histogram parameters with pulmonary metastasis in patients with colorectal cancer (CRC) and compare the efficacy of single-slice and whole-lesion histogram analysis.

Methods: Our study enrolled 196 CRC patients with pulmonary nodules (136 in the training dataset and 60 in the validation dataset). Twenty morphological features of contrast-enhanced chest CT were evaluated. The regions of interests were delineated in single-slice and whole-tumor lesions, and 22 histogram parameters were extracted. Stepwise logistic regression analyses were applied to choose the independent factors of lung metastasis in the morphological features model, the single-slice histogram model and whole-lesion histogram model. The areas under the curve (AUC) was applied to quantify the predictive accuracy of each model. Finally, we built a morphological-histogram nomogram for pulmonary metastasis prediction.

Results: The whole-lesion histogram analysis (AUC of 0.888 and 0.865 in the training and validation datasets, respectively) outperformed the single-slice histogram analysis (AUC of 0.872 and 0.819 in the training and validation datasets, respectively) and the CT morphological features model (AUC of 0.869 and 0.845 in the training and validation datasets, respectively). The morphological-histogram model, developed with significant morphological features and whole-lesion histogram parameters, achieved favorable discrimination in both the training dataset (AUC = 0.919) and validation dataset (AUC = 0.895), and good calibration.

OPEN ACCESS

Edited by:
Chunxiao Guo, University of Texas MD Anderson Cancer Center, United States

Reviewed by:
DingWei Ye, Fudan University Shanghai Cancer Center, China
Emily Ann Thompson, University of Texas MD Anderson Cancer Center, United States

***Correspondence:**
Tong Tong
t983352@126.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *07 July 2019*
Accepted: *29 October 2019*
Published: *19 November 2019*

Citation:
Hu T, Wang S, E X, Yuan Y, Huang L, Wang J, Shi D, Li Y, Peng W and Tong T (2019) CT Morphological Features Integrated With Whole-Lesion Histogram Parameters to Predict Lung Metastasis for Colorectal Cancer Patients With Pulmonary Nodules. Front. Oncol. 9:1241. doi: 10.3389/fonc.2019.01241

Conclusions: CT morphological features in combination with whole-lesion histogram parameters can be used to prognosticate pulmonary metastasis for patients with colorectal cancer.

Keywords: colorectal cancer, pulmonary metastases, histogram, morphological, morphological features, nomogram

INTRODUCTION

Colorectal cancer (CRC) is the third common cause of morbidity and mortality worldwide (1, 2). Pulmonary is the most common extra-abdominal site of metastasis for those with CRC, with 5–10% of CRC patients developing pulmonary metastasis (PM) (3, 4). The 5-year survival rates after initial colorectal surgery in patients with and without resection for pulmonary metastasis are 68 and 13%, respectively (3). The strong survival benefits of pulmonary metastasectomy make this treatment the generally accepted treatment for patients to achieve long-term survival when there is a definite and clear diagnosis (5, 6). Furthermore, if pulmonary metastasis is diagnosed early and resected aggressively, the survival rate is further improved (7).

However, with chest CT applied as part of preoperative routine examination, an increasing number of CRC patients are being diagnosed with indeterminate pulmonary nodules (IPNs) of unknown nature (8). The reported incidence of IPNs in CRC patients is 25–45.5% (8–10). Further diagnostic tests can also be problematic as nodules <10 mm in diameter may fall below the threshold of detection for positron emission tomography (PET) (11), and fine-needle aspiration cytology may not be feasible for thoracoscopic localization (12). Therefore, in CRC patients with IPNs, the accurate diagnosis of metastatic disease at an early and surgically treatable stage remains a challenge.

Though early-stage metastatic nodules and benign lesions have similar appearance in images, the importance of morphology should not be underestimated (13). CT imaging allows detailed observation of the morphological features of nodules and lesions, such as their internal density, shape, margin, and other typical characteristics. In recent years, texture analysis has emerged as a valuable methodology for facilitating diagnosis through the deep mining of information from medical images (14, 15). It has achieved great utility in evaluating many kinds of pulmonary diseases, including pulmonary embolisms (16), interstitial lung disease (17), and pulmonary nodules (18, 19). By extracting features of subtle pixel distributions and spatial variations of the gray levels of lesions that are imperceptible to the naked eye, texture analysis provides a complementary method for evaluating subjective and megascopic morphological features.

To date, studies concentrating on the morphological and textural features of IPNs 5–20 mm in diameter on contrast-enhanced CT in CRC patients remain limited. This study sought to determine the morphological characteristics and histogram parameters derived from texture analysis for CRC patients with IPNs and to construct a risk model with a combination of independent predictors to facilitate the accurate diagnosis of pulmonary metastasis.

MATERIALS AND METHODS

Patients

This retrospective analysis had obtained the ethical approval, and the informed consent requirement was waived. Our study enrolled 196 consecutive colorectal cancer patients (88F/108M; age range, 32–80 years; mean age, 58.49 ± 10.80 years) with lung nodules admitted in our institution between January 2010 and December 2017. The inclusion criteria were as follows: (i) colorectal cancer was histopathologically confirmed; (ii) at least one lung nodule measuring 5–20 mm detected by contrast-enhanced chest CT examination; (iii) available pathology reports with diagnosis of pulmonary metastasis or primary lung cancer for the malignant nodules and at least 2 years follow-up for the benign nodules; and (iv) complete medical history. The exclusion protocol were as follows: (i) with pretreatment 6 months before initial CT examination (including chemotherapy or pneumonectomy); (ii) obsolete nodules detected 6 months before colorectal cancer was detected; (iii) obvious benign nodules with typical imaging characteristics (such as cysts, tuberculosis, or inflammatory nodules); and (iv) adjuvant therapy (including radiation therapy or chemotherapy) applied for no-progress lesions in the process of follow-up. When there are multiple nodules, we choose the largest nodule for morphological and radiomics analysis. Of the 196 people included in the study, 194 of them have been published in our previous research (20).

Nodules were divided into two groups: (i) a pathologically confirmed lung metastasis group (95 PMs; 42F/53M; mean age, 57.46 ± 10.58 years), and (ii) a non-metastasis (NM) group (101 NMs; 46F/55M; mean age, 59.47 ± 10.91 years), including benign nodules (90 cases) with at least 2 years follow-up (88 cases) and pathology confirmation (2 cases) or primary lung cancer confirmed by pathology (11 cases). We used a computer algorithm to randomly divide the patients into a training dataset and a validation dataset at the ratio of 7:3. **Figure 1** shows the process of patients' recruitment.

CT Scanning Protocol

Chest CT examinations were performed at our institution with the Sensation 64 scanner (Siemens Healthcare) or the Somatom Definition AS scanner (Siemens Healthcare). The Contrast-enhanced CT scan parameters were as follows: contrast medium, inhexol; tube voltage, 120 kVp; tube current, 250–350 mA; slice thickness, 1.5 mm; slice interval, 1.5 mm; matrix, 512 × 512; field of view (FOV), 35–50 cm; pitch, 1.078; reconstruction algorithm, standard. The arterial phase of the target nodule which was pathologically confirmed or under follow-up was selected for reconstruction.

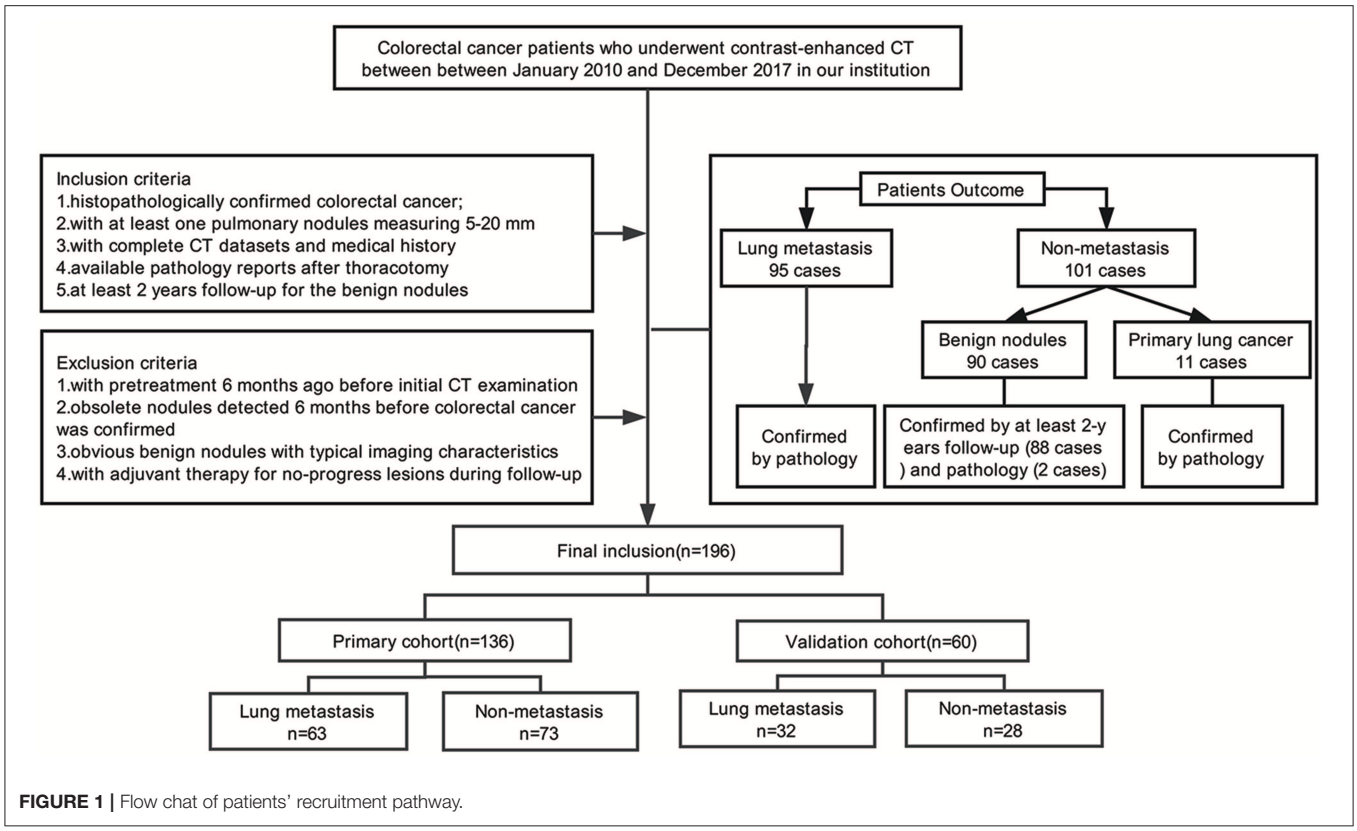


FIGURE 1 | Flow chat of patients' recruitment pathway.

CT Image Interpretation

The interpretations of CT features are listed in **Supplementary Table 1**. The CT morphological features were independently evaluated by two operators (SW and TH, with 20 and 3 years of experience in chest CT, respectively). In cases of disagreement, a third radiologist (TT, with 20 years of experience in CT imaging) was consulted, and the majority value was used. Mean values were calculated for continuous variables. The CT images were read with both mediastinal and lung window settings. All of the operators were blinded to the clinical and histologic findings.

Histogram Analysis

Reconstructed images were transferred to the MIM software (v6.6.3; MIM Software Inc.) for histogram analysis. For each patient, regions of interest (ROIs) were first semi-automatically contoured in the largest-cross sectional area of the tumor outline and then manually delineated by an operator and verified by an expert radiologist. Each ROI was propagated to include the entire tumor volume in each consecutive slice using the same contouring method. In the process of delineation, we excluded the border of the lesion and any other irrelevant tissues or regions, such as pleura, normal tissue, air, peripheral vessels, and surrounding organs. **Supplementary Figure 1** shows an example of ROI delineation.

The histogram parameters were automatically measured by the software using a volumetric approach on the ROI of the nodule. Single-slice and whole-lesion histogram parameters were extracted and analyzed. From each segmented tumor, we extracted 11 single-slice histogram parameters and 11 whole-lesion histogram features. More information about the methodology used to extract histogram features can be found in **Supplementary Material**.

Statistical Analysis

R software (version 3.3) was applied for statistical analysis. To measure the agreement of CT morphological features between two readers, intraclass correlation coefficients (ICCs) were calculated (poor: 0.00–0.20; fair: 0.21–0.40; moderate: 0.41–0.60; good: 0.61–0.80; excellent: 0.81–1.00). To compare the proportional differences between the training dataset and the validation dataset, chi-square tests were applied for the categorical variables, and two-sample *t*-tests were used for the continuous variables. To compare the differences between the PM and NM group, chi-square and two-sample *t*-tests were applied as appropriate for both the training and validation datasets. Two-sided $p < 0.05$ was considered significant.

Model Selection

The significant factors were introduced into the stepwise logistic regression to select the independent features for the CT morphological model, the single-slice histogram model and the whole-lesion histogram model. The Akaike information criterion (AIC) was employed as the stopping rule. The validation dataset was used to test the diagnostic performance of the models by applying the multivariable regression formula derived from the training dataset to the patients of the validation dataset, and the

TABLE 1 | Comparison of morphological features of lung metastasis (PM) and non- metastasis (NM) in the training and validation datasets.

	Training dataset			Validation dataset		
Characteristics	**NM**	**PM**	***p***	**NM**	**PM**	***p***
Age	59.82 ± 10.530	58.17 ± 11.360	0.382	58.54 ± 12.188	56.06 ± 9.055	0.372
Gender			0.676			0.316
1	42 (57.5%)	34 (54.0%)		13 (46.4%)	19 (59.4%)	
2	31 (42.5%)	29 (46.0%)		15 (53.6%)	13 (40.6%)	
Lobe location			0.025*			0.507
1	12 (16.4%)	19 (30.2%)		7 (25.0%)	6 (18.8%)	
2	12 (16.4%)	10 (15.9%)		4 (14.3%)	10 (31.3%)	
3	23 (31.5%)	9 (14.3%)		8 (28.6%)	5 (15.6%)	
4	12 (16.4%)	5 (7.9%)		1 (3.6%)	1 (3.1%)	
5	14 (19.2%)	20 (31.7%)		8 (28.6%)	10 (31.3%)	
Size category			<0.001*	0.00%	0.00%	0.095
1	56 (76.7%)	22 (34.9%)		20 (71.4%)	14 (43.8%)	
2	10 (13.7%)	24 (38.1%)		4 (14.3%)	10 (31.3%)	
3	7 (9.6%)	17 (27.0%)		4 (14.3%)	8 (25.0%)	
Long-axis diameter	8.988 ± 3.493	12.662 ± 4.200	<0.001*	9.221 ± 4.422	11.819 ± 4.180	0.023*
Short-axis diameter	5.825 ± 2.144	9.611 ± 9.253	0.001*	5.682 ± 2.031	8.184 ± 3.071	<0.001*
Density			<0.001*			0.002*
1	12 (16.4%)	0 (0.0%)		6 (21.4%)	0 (0.0%)	
2	15 (20.5%)	2 (3.2%)		6 (21.4%)	2 (6.3%)	
3	46 (63.0%)	61 (96.8%)		16 (57.1%)	30 (93.8%)	
Contour			0043*			0.011*
1	2 (2.7%)	10 (15.9%)		1 (3.6%)	1 (3.1%)	
2	21 (28.8%)	20 (31.7%)		2 (7.1%)	8 (25.0%)	
3	29 (39.7%)	19 (30.2%)		11 (39.3%)	19 (59.4%)	
4	21 (28.8%)	14 (22.2%)		14 (50.0%)	4 (12.5%)	
Border			<0.001*			<0.001*
1	28 (38.4%)	2 (3.2%)		13 (46.4%)	1 (3.1%)	
2	24 (32.9%)	43 (68.3%)		10 (35.7%)	20 (62.5%)	
3	21 (28.8%)	18 (28.6%)		5 (17.9%)	11 (34.4%)	
Air bronchogram			0.032*			0.178
0	72 (98.6%)	57 (90.5%)		28 (100.0%)	30 (93.8%)	
1	1 (1.4%)	6 (9.5%)		0 (0.0%)	2 (6.3%)	
Lymphadenopathy			0.032*			0.369
0	72 (98.6%)	57 (90.5%)		27 (96.4%)	29 (90.6%)	
1	1 (1.4%)	6 (9.5%)		1 (3.6%)	3 (9.4%)	

Chi-square tests were used to compare the differences in categorical variables while a two-sample t-test was used to compare the differences in continuous variables.
NM, non-metastasis group; PM, lung metastasis group.
**p < 0.05.*

probability of metastasis was calculated for each. The area under the receiver characteristic curve (AUC) was calculated to quantify the predictive accuracy of the three models in the training and validation datasets. We also calculated the accuracy, sensitivity, specificity, positive predictive value, and negative predictive value for each model.

We compared the relative strengths of the single-slice and whole-lesion histogram models and then used the more efficient model in combination with the morphological features to construct the morphological-histogram model. A morphological-histogram nomogram was then constructed for clinical application. A receiver operating characteristic (ROC) curve was used to describe the discrimination abilities of the nomogram. An AUC above 0.75 is considered as good (21). Nomogram performance was graphically demonstrated by calibration plots in both the training and validation datasets. Finally, decision curve analysis (DCA) was applied to assess the clinical usefulness of the nomogram.

RESULTS

Patient Characteristics

The patients characteristics and statistically significant CT morphological features are shown in **Table 1**

TABLE 2 | Comparison of single-slice and whole-lesion histogram parameters of PM and NM in the training and validation datasets.

Parameters	Training dataset			Validation dataset		
	NM	PM	p	NM	PM	*p*
S-ASD	−2.976 ± 15.859	−2.902 ± 8.077	0.973	−2.312 ± 9.593	0.490 ± 6.293	0.181
S-STD	161.602 ± 101.977	135.420 ± 60.213	0.076	129.270 ± 68.500	130.685 ± 59.258	0.932
S-Average ratio	0.428 ± 0.685	1.345 ± 1.414	<0.001*	0.523 ± 0.888	0.993 ± 0.887	0.045*
S-Mean	−58.878 ± 340.285	−15.669 ± 330.362	0.456	−73.787 ± 164.636	−37.878 ± 88.316	0.289
S-Skewness	0.088 ± 1.089	−0.854 ± 0.816	<0.001*	0.231 ± 1.035	−0.871 ± 0.995	<0.001*
S-Kurtosis	0.987 ± 4.616	1.163 ± 3.397	0.802	1.364 ± 3.860	1.727 ± 5.520	0.772
S-Area	0.775 ± 0.357	1.236 ± 0.416	<0.001*	0.795 ± 0.419	1.128 ± 0.373	0.002*
S-Volume	0.428 ± 0.685	1.345 ± 1.414	<0.001*	0.523 ± 0.888	0.993 ± 0.887	0.045*
S-Median	−50.810 ± 344.826	−73.250 ± 146.112	<0.001*	−299.820 ± 296.139	−29.190 ± 114.492	<0.001*
S-Maximum	190.590 ± 450.767	156.170 ± 176.689	0.57	123.320 ± 600.592	197.130 ± 112.848	0.498
S-Minimum	−58.970 ± 233.980	−69.970 ± 252.732	0.035*	−623.040 ± 210.656	−532.810 ± 289.115	0.178
W-ASD	364.124 ± 909.629	218.649 ± 176.266	0.214	211.019 ± 314.358	204.203 ± 179.826	0.917
W-STD	161.978 ± 101.572	135.300 ± 60.134	0.07	128.909 ± 68.194	130.383 ± 59.421	0.929
W-Average ratio	0.960 ± 0.466	1.610 ± 0.522	<0.001*	0.960 ± 0.404	1.485 ± 0.450	<0.001*
W-Mean	−39.095 ± 326.163	−99.504 ± 129.099	0.002*	−292.437 ± 290.740	−56.853 ± 106.652	<0.001*
W-Skewness	0.064 ± 1.016	−0.851 ± 0.824	<0.001*	0.245 ± 1.041	−0.859 ± 1.008	<0.001*
W-Kurtosis	3.723 ± 4.279	4.185 ± 3.455	0.494	4.373 ± 3.957	4.713 ± 5.622	0.791
W-Area	86.586 ± 145.308	206.406 ± 221.798	<0.001*	93.797 ± 112.534	167.841 ± 189.459	0.076
W-Volume	134.501 ± 340.853	394.708 ± 510.214	0.001*	122.925 ± 184.577	301.584 ± 380.789	0.028*
W-Median	−42.164 ± 337.581	−73.365 ± 146.188	<0.001*	−299.839 ± 295.991	−29.359 ± 114.730	<0.001*
W-Maximum	188.810 ± 450.066	156.140 ± 176.460	0.589	138.640 ± 586.255	195.940 ± 110.947	0.59
W-Minimum	−18.330 ± 250.784	−57.700 ± 263.027	0.172	−613.680 ± 225.487	−505.530 ± 311.857	0.134

ASD, Average standard deviation ratio; STD, Standard deviation.
S-, single-slice histogram parameters; W-, whole-lesion histogram parameters. A two-sample t-test was used to compare the differences of those parameters.
**p < 0.05.*

(**Supplementary Table 3** contains complete morphological features comparison), and the histogram parameters are presented in **Table 2**. There were no significant differences between the training and validation datasets except in pleural attachment (**Supplementary Table 2**). The agreement between the two operators was excellent for most characteristics and good for several features (**Supplementary Table 4**).

Significant Morphological Features and Histogram Parameters

Regarding the CT morphological features, the chi-square tests and *t*-tests revealed that nine CT features were associated with lung metastasis, including lobe location ($p = 0.025$), size category ($p < 0.001$), long-axis diameter ($p < 0.001$), short-axis diameter ($p = 0.001$), density ($p < 0.001$), contour ($p = 0.043$), border ($p < 0.001$), air bronchogram ($p = 0.032$), and lymphadenopathy ($p = 0.032$). After stepwise logistic analysis, long-axis diameter (OR = 1.360, 95%CI: 1.198–1.544, $P < 0.001$), density (OR = 11.166, 95%CI: 2.721–45.815, $P < 0.001$) and contour (OR = 0.317, 95%CI: 0.177–0.569, $P = 0.001$) remained independent predictors in the CT morphological model, as shown in **Table 3**.

Regarding the single-slice histogram parameters (S- means the parameters from the single-slice histogram analysis and W- from the whole-slice histogram), *t*-tests revealed that the S-average ratio ($p < 0.001$), S-skewness ($p < 0.001$), S-area ($p < 0.001$), S-volume ($p < 0.001$), S-median ($p < 0.001$), and S-minimum ($p = 0.035$) were significant variables related to PM.

TABLE 3 | Comparison of the models by multivariate logistic regression analysis.

	OR (95%CI)	P	AIC
CT morphological features			127.34
Long-axis diameter	1.360 (1.198–1.544)	<0.001*	
Density	11.166 (2.721–45.815)	<0.001*	
Contour	0.317 (0.177–0.569)	0.001*	
Single-slice histogram			130.90
S-Average ratio	0.268 (0.111–0.642)	0.003*	
S-Area	559.372 (42.344–7389.333)	<0.001*	
S-Median	1.004 (1.002–1.005)	<0.001*	
Whole-lesion histogram			130.25
W-Average ratio	12.764 (4.653–35.018)	<0.001*	
W-Mean	0.977 (0.961–0.994)	0.004*	
W-Median	1.024 (1.008–1.041)	0.009*	
Morphological-histogram			121.74
Density	5.434 (1.161–25.440)	0.032*	
Contour	0.495 (0.286–0.858)	0.012*	
W-Average ratio	9.727 (3.538–26.740)	<0.001*	
W-Mean	0.977 (0.959–0.995)	0.009*	
W-Median	1.023 (1.006–1.042)	0.013*	

OR, odds ratio; CI, confidence interval; AIC, Akaike information criterion.
**p < 0.05.*

After stepwise logistic analysis, the S-average ratio (OR = 0.268, 95%CI: 0.111–0.642, $P = 0.003$), S-area (OR = 559.372, 95%CI: 42.344–7389.333, $P < 0.001$), and S-median (OR = 1.004, 95%CI:

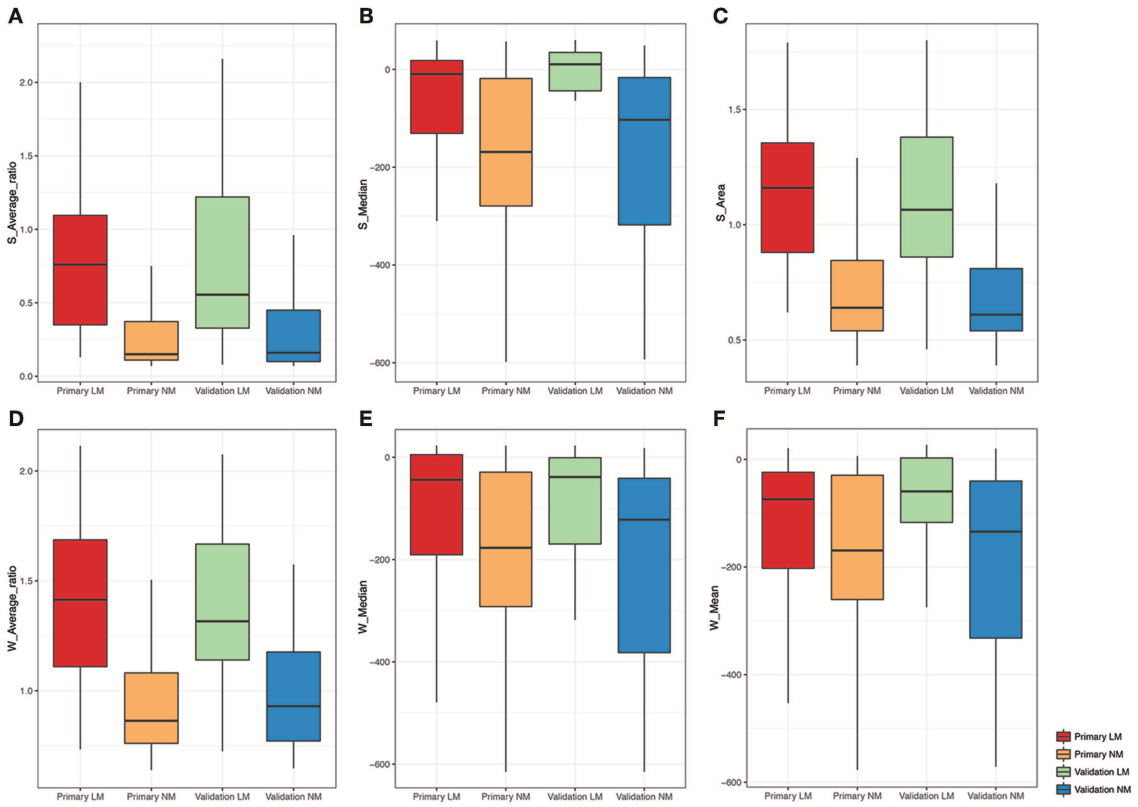


FIGURE 2 | Boxplot of the selected histogram parameters in the LM and NM group. **(A–C)** Boxplot of the S-Average ratio, S-Median, S-Area from the single-slice histogram in the LM and NM group, respectively. **(D–F)** Boxplot of the W-Average ratio, W-Median, W-Mean from the whole-lesion histogram in the LM and NM group, respectively.

1.002–1.005, $P < 0.001$) were selected as independent predictors for the single-slice histogram model.

Regarding the whole-lesion histogram parameters, *t*-tests revealed that the W-average ratio ($p < 0.001$), W-mean ($p = 0.002$), W-skewness ($p < 0.001$), W-area ($p < 0.001$), S-volume ($p = 0.001$), W-median ($p < 0.001$), and S-minimum ($p = 0.035$) were significant parameters associated with PM. After stepwise logistic analysis, the W-average ratio (OR = 12.764, 95%CI: 4.653–35.018, $P = 0.003$), W-mean (OR = 0.977, 95%CI: 0.961–0.994, $P = 0.004$), and S-median (OR = 1.024, 95%CI: 1.008–1.041, $P = 0.009$) were selected as independent predictors for the whole-lesion histogram model. **Figure 2** shows the distributions of the significant histogram parameters in the training and validation datasets.

Comparison of Single-Slice and Whole-Lesion Histogram Analyses

The whole-lesion histogram model (AIC = 130.25) had lower AIC value than the single-slice model (AIC = 130.9) and achieved better discrimination. It yielded an AUC of 0.888 for the training dataset and of 0.865 for the validation dataset, exceeding the AUC values of the single-slice model (AUC = 0.872 for the training dataset and AUC = 0.819 for the validation dataset). The ROC curves of the two models are presented in **Figure 3**.

Development and Validation of the Morphological-Histogram Nomogram

We subjected the CT morphological features and the whole-lesion histogram parameters to stepwise logistic regression analysis. Density (OR = 5.434, 95%CI: 1.161–25.440, $P = 0.032$), contour (OR = 0.495, 95%CI: 0.286–0.858, $P = 0.012$), the *W*-average ratio (OR = 9.727, 95%CI: 3.538–26.740, $P < 0.001$), W-mean (OR = 0.977, 95%CI: 0.959–0.995, $P = 0.009$), and W-median (OR = 1.023, 95%CI: 1.006–1.042, $P = 0.013$) were identified as independent risk factors in the model. The integrated model also achieved the best performance among the models, with an AUC of 0.919 (95%CI: 0.871–0.968, accuracy: 88.2%, sensitivity: 84.9%, specificity: 92.1%, PPV: 92.5%, NPV: 84.1%) for the training dataset and of 0.895 (95%CI: 0.813–0.977, accuracy: 81.7%, sensitivity: 78.5%, specificity: 84.4%, PPV: 81.5%, NPV: 81.8%) for the validation dataset (**Table 4**). The ROC curves of the models for both the training and validation datasets are presented in **Figure 3**.

The morphological-histogram nomogram was successfully constructed, with good discrimination, based on the morphological-histogram model (**Figure 4A**). The calibration plots also presented good accordance between the nomogram prediction and actual outcome for PM and NM in both the training and validation datasets (**Figures 4B,C**). The decision

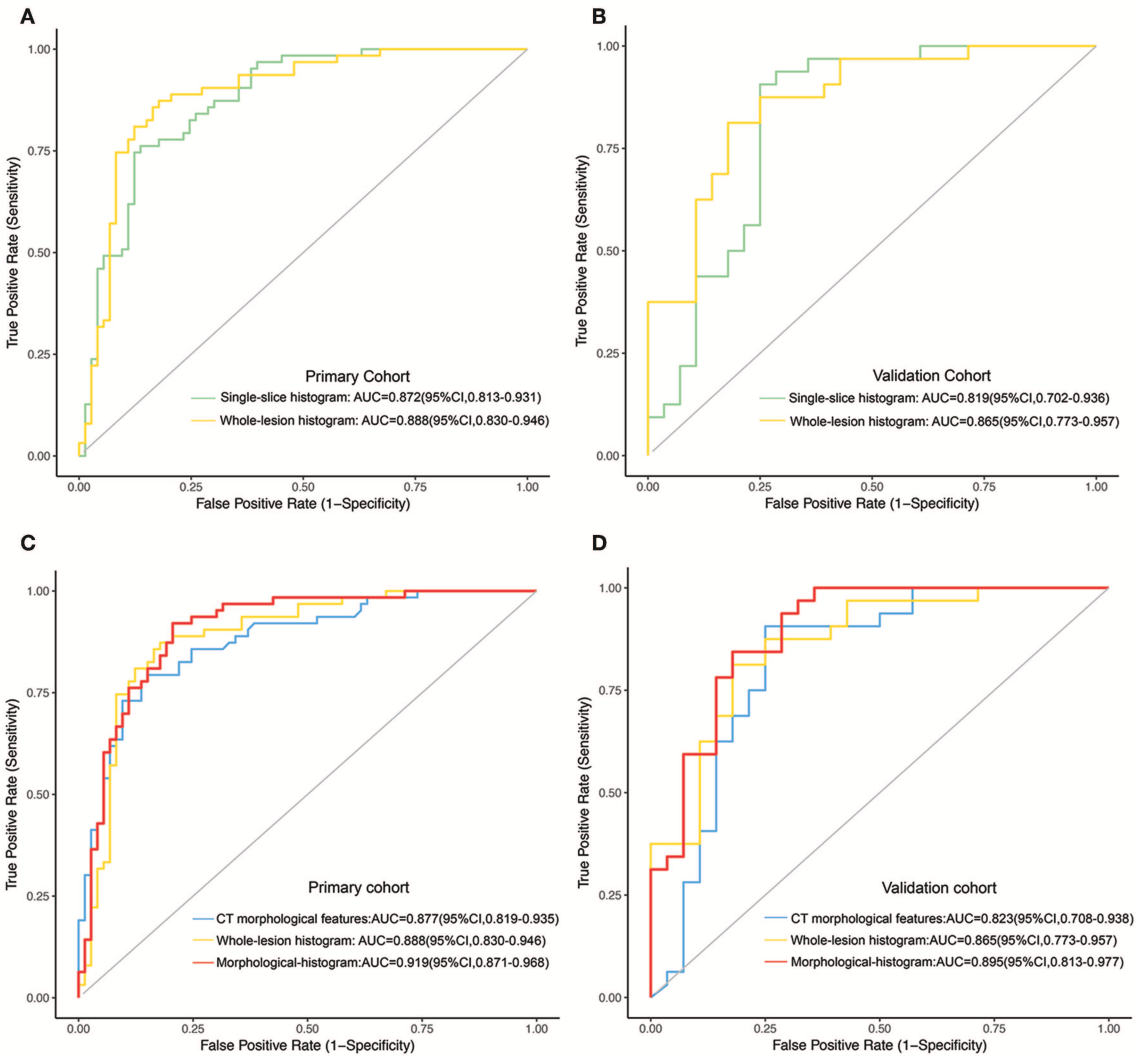


FIGURE 3 | (A,B) The ROC curves of the single-slice histogram model and whole-lesion histogram model in the training **(A)** and validation sets **(B)**, respectively. **(C,D)** The ROC curves of the CT morphological model, the whole-lesion histogram model and the integrated morphological-histogram model in the training **(C)** and validation cohorts **(D)**, respectively.

curve analysis demonstrated that given a threshold probability ranging from 0 to 100%, the morphological-histogram model was superior to the treat-all and treat-none schemes in predicting lung metastasis (**Figure 4D**).

DISCUSSION

In the present study, we investigated the imaging characteristics of IPNs 5–20 mm in diameter on initial CT in CRC patients and compared the predictive accuracy of whole-lesion and single-slice histogram parameters. We then constructed a morphological-histogram nomogram using a combination of morphological features and whole-lesion histogram parameters for IPNs. This nomogram may be clinically useful for discriminating CRC patients who might benefit from early and curable metastasectomy for metastatic lesions or an appropriate surveillance program.

CT offers direct visualization of lesions and potentially allows a detailed characterization of the morphologic extent of lesions. The careful evaluation of morphologic features is an essential step in pulmonary nodules assessment (13). Although several studies (22–24) have sought to identify significant image features for metastatic nodules, there is no consensus regarding the definition of IPNs, which led to slight differences between our results and previously published ones. In our study, we found that significant morphological features associated with pulmonary metastasis were long-axis diameter, density, and contour.

As reported by many other studies, nodule diameter is a reliable indicator of malignant potential (22, 23, 25). We found that solid nodules are more likely to be metastatic lesions. As more than 95% of nodules that originate from colorectal

TABLE 4 | Accuracy and predictive value between those models.

Training dataset	AUC	95%CI	Sensitivity	Specificity	Accuracy	PPV	NPV
CT morphological features	0.877	0.819–0.935	83.6% (61/73)	79.4% (50/63)	81.6% (111/136)	82.4% (61/74)	80.7% (50/62)
Single-slice histogram	0.872	0.813–0.931	86.3% (63/73)	76.2% (48/63)	81.6% (111/136)	80.8% (63/78)	82.8% (48/58)
Whole-lesion histogram	0.888	0.830–0.946	82.2% (60/73)	87.3% (55/63)	84.6% (115/136)	88.2% (60/68)	80.9% (55/68)
Morphological-histogram	0.919	0.871–0.968	84.9% (62/73)	92.1% (58/63)	88.2% (120/136)	92.5% (62/67)	84.1% (58/69)
Validation dataset	**AUC**	**95%CI**	**Sensitivity**	**Specificity**	**Accuracy**	**PPV**	**NPV**
CT morphological features	0.823	0.708–0.938	78.5% (22/28)	75% (24/32)	76.7% (46/60)	73.3% (22/30)	80% (24/30)
Single-slice histogram	0.819	0.702–0.936	75% (21/28)	71.9% (23/32)	73.3% (44/60)	70% (21/30)	76.7% (23/30)
Whole-lesion histogram	0.865	0.773–0.957	75% (21/28)	84.4% (27/32)	80% (48/60)	80.8% (21/26)	79.4% (27/34)
Morphological-histogram	0.895	0.813–0.977	78.5% (22/28)	84.4% (27/32)	81.7% (49/60)	81.5%(22/27)	81.8%(27/33)

CI, confidence interval; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.
$^{*}p < 0.05$.

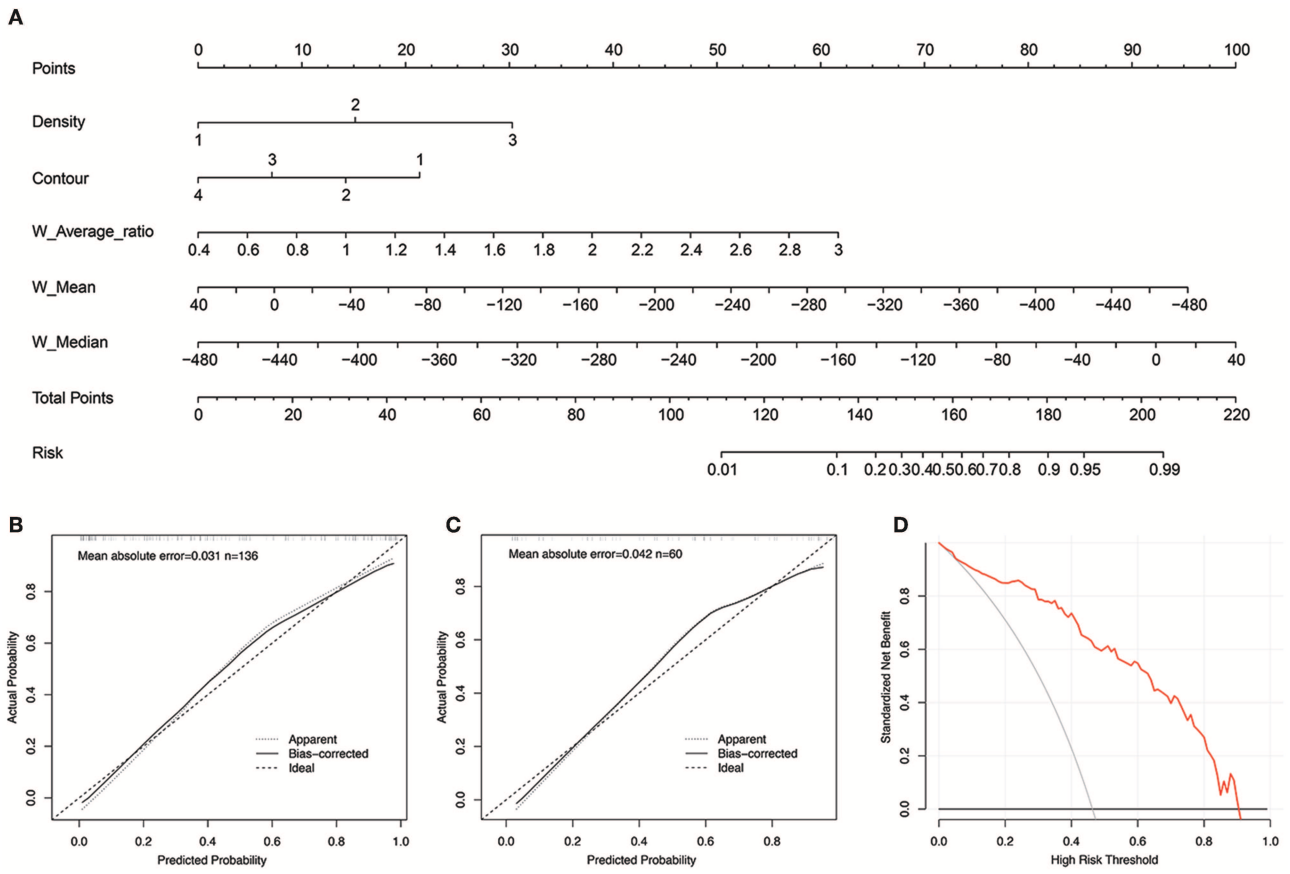


FIGURE 4 | (A) The developed morphological-histogram nomogram for predicting the probability of pulmonary metastases. By summing the scores of each point and locating on the total score scale, the estimated probability of pulmonary metastases could be determined. **(B,C)** The Calibration curves for predicting pulmonary metastases in the training and validation cohort. The y axis represents the actual rate of LM. The x axis represents the predicted probability of LM. The ideal line represents a perfect prediction by an ideal model. The apparent line represents the performance of the nomogram model, of which a closer fit to the ideal line represents a better prediction. **(D)** The decision curves analysis for the morphological-histogram nomogram. The red line represents the net benefit of morphological-histogram model. Across the various threshold probabilities, the morphological-histogram curve showed great net benefit.

cancer are adenocarcinomas (4), metastatic lesions tend to appear as solid pulmonary nodules (SPN) in CT scans, whereas benign lesions, such as inflammation lesions, or organizing pneumonia/fibrosis consistently present as patchy consolidations or mixed-density regions surrounded by ground-glass opacity (GGO) owing to inflammatory cell infiltration (26). Primary lung cancer consistently evolves from pre-invasive lesions (AIS/AAH) that manifested as pure GGO (27) at the early stage. A *post-hoc*

analysis (24) found that a solid consistency and increasing size were statistically associated with malignancy.

In addition, our study found that metastatic nodules tended to be round or oval, consistent with previous research (28). We speculate that as metastatic nodules often exhibit a largely uniform growth rate and homogenous invasion in all directions, these features contribute to a round or quasi-circular contour, whereas non-metastatic lesions, including benign lesions and primary lung cancer, have irregular shapes due to uneven growth rates at various sites (26). Thus, short-interval CT follow-up is highly recommended for IPNs larger than 5 mm in diameter with solid components and approximately regular margins detected on preoperative chest CT.

In addition to the identification of morphological features, the use of texture analysis is a strength of our study. Previous studies have demonstrated that texture analysis can not only distinguish malignant nodules from benign ones (18) but also differentiate *in situ* and minimally invasive lung adenocarcinoma subtypes (19). These studies have shown that texture parameters can reveal the underlying histological changes in tissue below the resolution of the given modality or protocol. In this study, we found that the W-average ratio, W-mean, and W-median, which represent the zone of CT attenuation within the ROI, were substantially higher in the metastasis group than in the non-metastasis group. Thus, short-interval CT follow-up is highly recommended for IPNs larger than 5 mm in diameter with solid components and approximately regular margins detected on preoperative chest CT. This speculation is also in line with another finding of our study that vascular convergence was more common and the enhancement degree was higher in the metastasis group than in the non-metastasis group. However, as texture analysis is a mathematical method, the biological mechanisms underlying the textural features are complex and not completely understood (29). In cases where vascular convergence or the enhancement degree is insufficient to differentiate metastatic lesions, the values from the CT attenuation zone might exhibit local variation and more sensitive preservation of spatial information (30).

Another finding of our study was that the whole-lesion texture analysis outperformed the single-slice analysis in evaluating pulmonary nodules, consistent with a previous study (31). Whole-lesion analysis may provide a more comprehensive understanding of the stereo structure of the whole lesion and thereby reflect the integral heterogeneity better than can single-slice analysis. Despite the time-consuming process of the contouring around the whole lesion, it seems more cost-efficient to use this method as it provides improved prediction relative to single-slice analysis and a more definite diagnosis, allowing timely treatment and maximizing the benefits to the patient.

For clinical use, we constructed a risk stratification nomogram for the clinician to predict the risk of PM for an individual CRC patient. As the early and accurate diagnosis of pulmonary metastasis has been recognized as one of the most important steps in treating potential curable lesions with surgery, we propose that patients with a high risk of PM be considered candidates for thoracotomy for resectable lesions to enhance local control and improve the survival rate. We also hope this model can help low-risk patients avoid aggressive follow-up and reduce the burden of radiation exposure. We believe that the clinical use of the nomogram can contribute to reliable diagnoses and help clinicians optimize therapeutic plans for IPNs at an early stage after detection.

Our study has several limitations. First, as a retrospective study, thin-slice contrast-enhanced CT images from our database were used, which limited the number of cases for analysis. And the inclusion and exclusion criteria also limits the implementation of the study in clinical practice. Second, only histogram parameters were extracted in this study. In our previous research (20), 203 radiomic features, including first- and second-order parameters, attained a prognostic value in the differentiation of pulmonary metastasis with an AUC of 0.888, which is slightly higher than that obtained using the histogram parameters (AUC = 0.887). However, the process of extracting radiomic features through MATLAB is intricate and demanding for radiologists and clinicians, which constrains its clinical utilization. The volume histogram analysis performed here allowed the simple, efficient, and automatic acquisition of a density histogram and achieved an accuracy comparable to that of the radiomics analysis. Thus, volume histogram analysis may be more appropriate for imperative clinical decisions, and radiomics analysis can be used as a supplementary method when needed. Another limitation is that the development and validation were performed in a single institution. External validation and multi-center clinical trials are therefore needed for further generalization.

In conclusion, the results of our study demonstrated that histogram parameters may serve as non-invasive imaging biomarkers for differentiating pulmonary metastasis from non-metastatic lesions. When complemented with morphological features, the morphological-histogram nomogram can greatly benefit the diagnosis of pulmonary metastasis in CRC patients.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Medical ethics committee of Fudan University Shanghai Cancer Center. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

TT and SW carried out the concepts and design of the study. DS provided the patients information. YL confirmed the pathology results. XE, YY, HL, and JW provided assistance for data acquisition and statistical analysis. WP provided the permission of imaging acquisition. SW and TH carried our literature research and manuscript editing. TH and

SW contributed equally to this work. All authors have reviewed the final version of the manuscript and approved it for publication.

FUNDING

This study has received funding by the National Natural Science Foundation of China (Grant No. 81971687) and the Shanghai Committee of Science and Technology of China (Grant No. 19YF1409900).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01241/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin.* (2017) 67:177–93. doi: 10.3322/caac.21395
2. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin.* (2017) 67:7–30. doi: 10.3322/caac.21387
3. Kobayashi H, Mochizuki H, Sugihara K, Morita T, Kotake K, Teramoto T, et al. Characteristics of recurrence and surveillance tools after curative resection for colorectal cancer: a multicenter study. *Surgery.* (2007) 141:67–75. doi: 10.1016/j.surg.2006.07.020
4. Watanabe K, Saito N, Sugito M, Ito M, Kobayashi A, Nishizawa Y. Incidence and predictive factors for pulmonary metastases after curative resection of colon cancer. *Ann Surg Oncol.* (2013) 20:1374–80. doi: 10.1245/s10434-012-2747-y
5. Hachimaru A, Maeda R, Suda T, Takagi Y. Repeat pulmonary resection for recurrent lung metastases from colorectal cancer: an analysis of prognostic factors. *Interact Cardiovasc Thorac Surg.* (2016) 22:826–30. doi: 10.1093/icvts/ivv382
6. Cao G, Cheng D, Ye L, Pan Y, Yang F, Lyu S. Surgical resection of pulmonary metastases from colorectal cancer: 11 years of experiences. *PLoS ONE.* (2017) 12:e0175284. doi: 10.1371/journal.pone.0175284
7. Lee JH, Ahn BK, Nam YS, Lee KH. Clinical usefulness of preoperative and postoperative chest computed tomography for colorectal cancer. *J Korean Soc Coloproctol.* (2010) 26:359–64. doi: 10.3393/jksc.2010.26.5.359
8. Nordholm-Carstensen A, Wille-Jorgensen PA, Jorgensen LN, Harling H. Indeterminate pulmonary nodules at colorectal cancer staging: a systematic review of predictive parameters for malignancy. *Ann Surg Oncol.* (2013) 20:4022–30. doi: 10.1245/s10434-013-3062-y
9. Grossmann I, Avenarius JK, Mastboom WJ, Klaase JM. Preoperative staging with chest CT in patients with colorectal carcinoma: not as a routine procedure. *Ann Surg Oncol.* (2010) 17:2045–50. doi: 10.1245/s10434-010-0962-y
10. Maithel SK, Ginsberg MS, D'Amico F, DeMatteo RP, Allen PJ, Fong Y, et al. Natural history of patients with subcentimeter pulmonary nodules undergoing hepatic resection for metastatic colorectal cancer. *J Am Coll Surg.* (2010) 210:31–8. doi: 10.1016/j.jamcollsurg.2009.09.032
11. De Wever W, Meylaerts L, De Ceuninck L, Stroobants S, Verschakelen JA. Additional value of integrated PET-CT in the detection and characterization of lung metastases: correlation with CT alone and PET alone. *Eur Radiol.* (2007) 17:467–73. doi: 10.1007/s00330-006-0362-7
12. Yamauchi Y, Izumi Y, Hashimoto K, Inoue M, Nakatsuka S, Kawamura M, et al. Needle-tract seeding after percutaneous cryoablation for lung metastasis of colorectal cancer. *Ann Thorac Surg.* (2011) 92:e69–71. doi: 10.1016/j.athoracsur.2011.04.072
13. Snoeckx A, Reyntiens P, Desbuquoit D, Spinhoven MJ, Van Schil PE, van Meerbeeck JP, et al. Evaluation of the solitary pulmonary nodule: size matters, but do not ignore the power of morphology. *Insights Imaging.* (2018) 9:73–86. doi: 10.1007/s13244-017-0581-2
14. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
15. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
16. Sumikawa H, Johkoh T, Yamamoto S, Yanagawa M, Inoue A, Honda O, et al. Computed tomography values calculation and volume histogram analysis for various computed tomographic patterns of diffuse lung diseases. *J Comput Assist Tomogr.* (2009) 33:731–8. doi: 10.1097/RCT.0b013e31818da65c
17. Koyama H, Ohno Y, Yamazaki Y, Nogami M, Kusaka A, Murase K, et al. Quantitatively assessed CT imaging measures of pulmonary interstitial pneumonia: effects of reconstruction algorithms on histogram parameters. *Eur J Radiol.* (2010) 74:142–6. doi: 10.1016/j.ejrad.2009.02.006
18. Dennie C, Thornhill R, Sethi-Virmani V, Souza CA, Bayanati H, Gupta A, et al. Role of quantitative computed tomography texture analysis in the differentiation of primary lung cancer and granulomatous nodules. *Quant Imaging Med Surg.* (2016) 6:6–15. doi: 10.3978/j.issn.2223-4292.2016.02.01
19. Cohen JG, Reymond E, Medici M, Lederlin M, Lantuejoul S, Laurent F, et al. CT-texture analysis of subsolid nodules for differentiating invasive from *in-situ* and minimally invasive lung adenocarcinoma subtypes. *Diagn Interv Imaging.* (2018) 99:291–9. doi: 10.1016/j.diii.2017.12.013
20. Hu T, Wang S, Huang L, Wang J, Shi D, Li Y, et al. A clinical-radiomics nomogram for the preoperative prediction of lung metastasis in colorectal cancer patients with indeterminate pulmonary nodules. *Eur Radiol.* (2018) 29:439–49. doi: 10.1007/s00330-018-5539-3
21. Ying GS, Maguire M, Quinn G, Kulp MT, Cyert L, Vision in preschoolers study G. ROC analysis of the accuracy of noncycloplegic retinoscopy, retinomax autorefractor, and suresight vision screener for preschool vision screening. *Invest Ophthalmol Vis Sci.* (2011) 52:9658–64. doi: 10.1167/iovs.11-8559
22. Pomerri F, Pucciarelli S, Maretto I, Perrone E, Pintacuda G, Lonardi S, et al. Significance of pulmonary nodules in patients with colorectal cancer. *Eur Radiol.* (2012) 22:1680–6. doi: 10.1007/s00330-012-2431-4
23. Quyn AJ, Matthews A, Daniel T, Amin AI, Yalamarthi S. The clinical significance of radiologically detected indeterminate pulmonary nodules in colorectal cancer. *Colorectal Dis.* (2012) 14:828–31. doi: 10.1111/j.1463-1318.2011.02722.x
24. Nordholm-Carstensen A, Jorgensen LN, Wille-Jorgensen PA, Hansen H, Harling H. Indeterminate pulmonary nodules in colorectal-cancer: do radiologists agree? *Ann Surg Oncol.* (2015) 22:543–9. doi: 10.1245/s10434-014-4063-1
25. Kim CH, Huh JW, Kim HR, Kim YJ. Indeterminate pulmonary nodules in colorectal cancer: follow-up guidelines based on a risk predictive model. *Ann Surg.* (2015) 261:1145–52. doi: 10.1097/SLA.0000000000000853
26. Xiang W, Xing Y, Jiang S, Chen G, Mao H, Labh K, et al. Morphological factors differentiating between early lung adenocarcinomas appearing as pure ground-glass nodules measuring $\leq$10 mm on thin-section computed tomography. *Cancer Imaging.* (2014) 14:33. doi: 10.1186/s40644-014-0033-x
27. Lee SM, Park CM, Goo JM, Lee HJ, Wi JY, Kang CH. Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using CT features. *Radiology.* (2013) 268:265–73. doi: 10.1148/radiol.13120949

28. Yang Q, Wang Y, Ban X, Wu J, Rong D, Zhao Q, et al. Prediction of pulmonary metastasis in pulmonary nodules (ions appearing as ground-glass nodules: differentiation by using CT features. screening. form*Radiol Med.* (2017) 122:837–49. doi: 10.1007/s11547-017-0790-2
29. Suo S, Cheng J, Cao M, Lu Q, Yin Y, Xu J, et al. Assessment of heterogeneity difference between edge and core by using texture analysis: differentiation of malignant from inflammatory pulmonary nodules and masses. *Acad Radiol.* (2016) 23:1115–22. doi: 10.1016/j.acra.2016.04.009
30. Ferreira Junior JR, Koenigkam-Santos M, Cipriano FEG, Fabro AT, Azevedo-Marques PM. RadiomicS-based features for pattern recognition of lung cancer histopathology and metastases. *Comput Methods Programs Biomed.* (2018) 159:23–30. doi: 10.1016/j.cmpb.2018.02.015
31. Ng F, Kozarski R, Ganeshan B, Goh V. Assessment of tumor heterogeneity by CT texture analysis: can the largest crosS-sectional area be used as an alternative to whole tumor analysis? *Eur J Radiol.* (2013) 82:342–8. doi: 10.1016/j.ejrad.2012.10.023

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer DY declared a shared affiliation, though no other collaboration, with the authors to the handling Editor.

frontiers in Oncology

ORIGINAL RESEARCH
published: 22 November 2019
doi: 10.3389/fonc.2019.01265

Evaluation of Lymph Node Metastasis in Advanced Gastric Cancer Using Magnetic Resonance Imaging-Based Radiomics

Wujie Chen[1,2,3†], Siwen Wang[4,5†], Di Dong[4,5†], Xuning Gao[1,2], Kefeng Zhou[1,2], Jiaying Li[1,2], Bin Lv[1,3], Hailin Li[4,5], Xiangjun Wu[4,5], Mengjie Fang[4,5], Jie Tian[4,6] and Maosheng Xu[1,2*]*

[1] First College of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou, China, [2] Department of Radiology, The First Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, China, [3] Department of Gastroenterology, The First Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, China, [4] CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China, [5] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China, [6] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing, China

OPEN ACCESS

Edited by:
Chunxiao Guo,
University of Texas MD Anderson Cancer Center, United States

Reviewed by:
Zhi-Cheng Li,
Shenzhen Institutes of Advanced Technology (CAS), China
Jinhua Yu,
Fudan University, China
Yuming Jiang,
Stanford University, United States

***Correspondence:**
Jie Tian
jie.tian@ia.ac.cn
Maosheng Xu
xums166@zcmu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 11 September 2019
Accepted: 01 November 2019
Published: 22 November 2019

Citation:
Chen W, Wang S, Dong D, Gao X, Zhou K, Li J, Lv B, Li H, Wu X, Fang M, Tian J and Xu M (2019) Evaluation of Lymph Node Metastasis in Advanced Gastric Cancer Using Magnetic Resonance Imaging-Based Radiomics. Front. Oncol. 9:1265. doi: 10.3389/fonc.2019.01265

Objective: To develop and evaluate a diffusion-weighted imaging (DWI)-based radiomic nomogram for lymph node metastasis (LNM) prediction in advanced gastric cancer (AGC) patients.

Overall Study: This retrospective study was conducted with 146 consecutively included pathologically confirmed AGC patients from two centers. All patients underwent preoperative 3.0 T magnetic resonance imaging (MRI) examination. The dataset was allocated to a training cohort ($n = 71$) and an internal validation cohort ($n = 47$) from one center along with an external validation cohort ($n = 28$) from another. A summary of 1,305 radiomic features were extracted per patient. The least absolute shrinkage and selection operator (LASSO) logistic regression and learning vector quantization (LVQ) methods with cross-validations were adopted to select significant features in a radiomic signature. Combining the radiomic signature and independent clinical factors, a radiomic nomogram was established. The MRI-reported N staging and the MRI-derived model were built for comparison. Model performance was evaluated considering receiver operating characteristic (ROC) analysis, calibration curves, and decision curve analysis (DCA).

Results: A two-feature radiomic signature was found significantly associated with LNM ($p < 0.01$, training and internal validation cohorts). A radiomic nomogram was established by incorporating the clinical minimum apparent diffusion coefficient (ADC) and MRI-reported N staging. The radiomic nomogram showed a favorable classification ability with an area under ROC curve of 0.850 [95% confidence interval (CI), 0.758–0.942] in the training cohort, which was then confirmed with an AUC of 0.857 (95% CI, 0.714–1.000) in internal validation cohort and 0.878 (95% CI, 0.696–1.000) in external validation cohort. Meanwhile, the specificity, sensitivity, and accuracy were 0.846, 0.853, and 0.851 in internal validation cohort, and 0.714, 0.952, and 0.893 in external validation cohort, compensating for the MRI-reported N staging and MRI-derived model. DCA demonstrated good clinical use of radiomic nomogram.

Conclusions: This study put forward a DWI-based radiomic nomogram incorporating the radiomic signature, minimum ADC, and MRI-reported N staging for individualized preoperative detection of LNM in patients with AGC.

Keywords: lymph node metastasis, magnetic resonance imaging, diffusion-weighted imaging, advanced gastric cancer, radiomics

INTRODUCTION

Gastric cancer is a common and debilitating disease negatively impacting the physical and mental health of patients, worldwide. The onset of early gastric cancer is concealed, and most of them have become advanced gastric cancer (AGC) related to poor prognosis when clinically discovered (1). Evidence from studies shows that perioperative treatment of AGC (neoadjuvant chemotherapy/radiotherapy and adjuvant chemotherapy/radiotherapy) has been proven superior to surgery alone in many Western countries. The Chinese Society of Clinical Oncology also indicates that preoperative chemotherapy can well-improve the tumor remission rate and R0 resection rate with good safety in Asian countries based on D2 lymphadenectomy studies (2). As a crucial factor affecting the prognosis quality and survival of AGC patients, knowing the lymph node metastasis (LNM) status in advance has potential guiding significance for the decision making of therapeutic strategies including neoadjuvant chemotherapy, surgery, or intraoperative lymph node dissection (1–3). Morphological changes of lymph node architecture have been regarded as the reasonable and clinically acknowledged criteria for the determination of LNM currently (3). However, these changes do not correspond exactly to pathology. For example, small lymph nodes have metastasized, while large lymph nodes may be simply caused by inflammation (2, 4). Both errors offer a glimpse into the potential pitfalls of current LNM analysis methods. Therefore, a method allowing more accurate identification of LNM status should be considered as an urgent issue for clinical decision making.

Diffusion-weighted imaging (DWI) describes a magnetic resonance imaging (MRI) sequence which analyzes the Brownian movement of water molecules *in vivo* to determine morphological and functional parameters (5). Currently, DWI is a powerful modality to differentiate malignant and benign legions with the assumption that malignant lesions generally display higher cellularity. However, the correlation between DWI signal and LNM is not completely uniform, so the current accuracy of DWI-based analysis still falls below the clinical requirement in most cases (6).

Radiomics is a burgeoning field which involves converting imaging data into potential high-dimensional radiomic features through a large series of automatic feature extraction and data characterization algorithms (7–9). Quantitative radiomic feature analysis is now a widely recognized method in capturing distinct phenotypic differences along with changes in internal structure from a microscopic perspective (10). An increasing number of high-quality datasets and advanced pattern recognition algorithms have contributed to the rapid growth and development of radiomics (11). Furthermore, previous studies (5, 6, 9, 12) have indicated that certain quantitative radiomic signature had a surprising correlation with the prediction and evaluation of cancers. However, there is no article about DWI-based radiomic models for LNM prediction in AGC yet. Thereby, a combination of radiomics and DWI may provide a reliable method of precision medicine for the individualized prediction of LNM in patients with AGC.

Abbreviations: ADC, apparent diffusion coefficient; AGC, advanced gastric cancer; AUC, area under the curve; DCA, decision curve analysis; DWI, diffusion-weighted imaging; ICC, intraclass correlation coefficient; LASSO, least absolute shrinkage and selection operator; LNM, lymph node metastasis; LVQ, learning vector quantization; MRI, magnetic resonance imaging; OOB, out-of-bag; ROC, receiver operating characteristic; VOI, volume of interest.

OVERALL STUDY

Research Materials

Patients

Ethical approval for this retrospective study was granted by the ethics committee of the First Affiliated Hospital of Zhejiang Chinese Medical University and Hangzhou Hospital of Traditional Chinese Medicine. We waived the requirement for informed consent. This study consecutively enrolled 146 pathologically diagnosed AGC patients with total or partial radical gastrectomy from February 2016 to December 2018. **Supplementary Figure 1** shows the detailed recruitment diagram for study population from the two centers. The inclusion and exclusion criteria are defined as follows.

The inclusion criteria were the following: (a) patients with confirmed AGC according to the American Joint Committee on Cancer staging manual (1), and (b) a standard 3.0 T MRI was performed <2 weeks before surgical resection.

The exclusion criteria were the following: (a) patients with combined malignant neoplasm, distant metastasis, or preoperative therapy (radiotherapy, chemotherapy, or chemoradiotherapy); (b) incomplete clinical information or pathological information; (c) inflammatory diseases, including infections, ischemic heart disease, hereditary gastric cancer, collagen disease, and bowel perforation or obstruction; (d) the total number of intraoperative lymph node dissections was <16; and (e) low MRI resolution or small tumor lesion (<1 cm).

MRI Acquisition and Tumor Segmentation

All patients were given written informed consent before MRI examinations. Patients attending the inspection fasted for at least 8 h and drank 700–1,000 ml warm water within 5 min to fill the stomach cavity. Each patient was asked to cooperate

with the respiratory training before examination to ensure the normal inspection operation and reduce motion artifacts. A full diagnostic abdominal MRI protocol was executed. The MRI scans, covering the entire stomach region from the diaphragmatic dome to the level of the renal hilum, were performed during a breath-hold, with the patient supine in all the phases. All the patients underwent MRI scans successfully without any discomfort.

The images were exported from the Institutional Picture Archiving and Communication System (PACS, Carestream). MRI was performed using a whole-body 3.0 T scanner (Discovery 750, GE Healthcare, Milwaukee, WI, USA). Eight-channel head phased array coils and conventional sequences were used to obtain all the sequences. The scanning parameters of Axial DWI Shim are as follows: gradient factor b values are 0, 1,000 s/mm^2, matrix 128 × 130, TE = minimum, number of layers are 26 (maximum slices are 38), thickness of layer is 6.0 mm, spacing between layers is 2.0 mm, NEX for T2 is 4.00.

Manual segmentation of the entire tumor volume of interest (VOI) was conducted with ITK-SNAP software (version 3.6; www.itksnap.org) on the axial DWI sequence. VOI included the inner border of the lesion on whole axial slices and avoided necrotic tissue and surrounding adipose tissue (5). The T2-weighted images and contrast-enhanced T1-weighted images were used as references for the VOI segmentation on DWI sequence.

Three-dimensional volume images were delineated by two radiologists (WC and XG, with 7 and 25 years of experience in MRI abdominal diagnosis, respectively). They were both blind to pathological information of patients. WC performed tumor segmentation for all 146 patients and then repeated the segmentation procedure after 2 weeks on 30 randomly selected patients to test the intrareader consistency. XG only segmented the above 30 cases to assess the interreader consistency of the radiomic features.

Clinical Factors

Clinical factors for center 1 patients in this study are summarized in **Table 1**, including age, sex, the primary site of the tumor, tumor size, MRI-reported T staging, MRI-reported N staging, pathological T staging, average apparent diffusion coefficient (ADC) value, minimum ADC value, and combined markers (CA19-9, CA72-4, and CEA). The clinical factors for center 2 patients are given in **Supplementary Table 1**. The detailed grouping criteria are given as follows.

MRI-reported N staging

Patients were classified as N-positive if a regional lymph node with a measurement of >8 mm on its shortest axis was found, or if a regional lymph node had a higher signal intensity than muscle. The absence of enlarged (>8 mm) or hyperintense lymph nodes was defined as N-negative, which was consistent with the definition of radiological positive nodal status in most previous studies.

Primary site of the tumor

In the coronal position, the stomach was divided into upper, middle, and lower parts according to the tripartite connection of the greater curvature and the lesser curvature.

MRI-reported T staging

"T4 staging" defines a tumor lesion that infiltrates the serous layer, while T3 or T2 denotes a tumor that has not invaded the serous layer.

Combined markers

A combined marker was defined as positive when either of the three markers (CA19-9, CA72-4, and CEA) was positive, and all maker results came from the examination 1 week before surgery.

Radiomic Analysis Procedures

A dataset of 118 AGC patients from center 1 were separated into a training cohort ($n = 71$) and an internal validation cohort ($n = 47$) at a ratio of 3:2 randomly. Patients from center 2 constituted an external validation cohort ($n = 28$). As shown in **Figure 1**, the radiomics workflow consists of four steps, including tumor masking, radiomic feature extraction, radiomic signature construction, and radiomic nomogram development and evaluation.

Radiomic Feature Extraction

Radiomic features in this study were extracted from tumor VOIs on DWI images with algorithms implemented in Python 2.7 (https://www.python.org). The radiomic features (summarized in **Supplementary Table 2**) were composed of three groups: shape features, first-order features, and texture features.

To test the reproducibility and stability of extracted features, intraclass correlation coefficients (ICCs) were calculated (**Supplementary Material 1.1**). Features with ICC values >0.75 were reserved due to their good reproducibility. Then, all radiomic features were normalized.

Feature Selection and Radiomic Signature Construction

Radiomic feature selection as well as radiomic signature construction were carried out in the training cohort. The least absolute shrinkage and selection operator (LASSO) logistic regression was conducted by 5-fold cross-validation for feature reduction. Then, radiomic features were ranked according to their importance to LMN status using learning vector quantization (LVQ). LVQ is a kind of supervised neural network algorithm using a small number of weighted vectors to represent original data based on Euclidean distance measurements (13). Comparative out-of-bag (OOB) bootstrapping estimates with logistic regression models were performed 10 times for each feature subset consisting of the top 5, 10, 15, 20, and 25 features from LVQ, respectively. The average testing area under curve (AUC) and average bias between training AUC and testing AUC from 10 measurements were used as an approach to confirm the number of features in the optimal feature subset. Backward stepwise elimination with Akaike information criterion was then applied. Finally, selected radiomic features weighted by

TABLE 1 | Clinical and imaging characteristics of patients with AGC.

Clinical factors	Training cohort ($n = 71$)		*p* value	Validation cohort ($n = 47$)		*p* value
	LNM (+)	LNM (–)		LNM (+)	LNM (–)	
Age, mean ± SD, years	64.7 ± 12.22	66.06 ± 11.06	0.7458	61.24 ± 13.67	67.77 ± 7.51	0.1566
Sex, no. (%)			0.1615			0.6921
Male	43 (81.1)	11 (61.1)		28 (82.4)	10 (76.9)	
Female	10 (18.9)	7 (38.9)		6 (17.6)	3 (23.1)	
Primary site, no. (%)			0.5205			0.1796
Upper	12 (22.6)	6 (33.3)		6 (17.6)	3 (23.1)	
Middle	19 (35.8)	7 (38.9)		15 (44.1)	2 (15.4)	
Under	22 (41.5)	5 (27.8)		13 (38.2)	8 (61.5)	
Tumor size, no. (%)			0.1134			0.6771
<5.0 cm	21 (39.6)	11 (61.1)		15 (44.1)	7 (53.8)	
≥5.0 cm	32 (60.4)	7 (38.9)		19 (55.9)	6 (46.2)	
MRI-reported T staging, no. (%)			0.1726			0.4597
T2–3	17 (32.1)	9 (50.0)		10 (29.4)	6 (46.2)	
T4	36 (67.9)	9 (50.0)		24 (70.6)	7 (53.8)	
MRI-reported N staging, no. (%)			0.0172*			0.0489
Positive	42 (79.2)	9 (50.0)		29 (85.3)	7 (53.8)	
Negative	11 (20.8)	9 (50.0)		5 (14.7)	6 (46.2)	
pT staging, no. (%)			0.0005*			<0.0001*
T2–3	8 (15.1)	11 (61.1)		2 (5.9)	9 (69.2)	
T4	45 (84.9)	7 (38.9)		32 (94.1)	4 (30.8)	
Average ADC value, mean	1,419 (74.6)	1,387 (25.4)	0.9473	1,428 (72.3)	1,499 (27.7)	0.3788
Minimum ADC value, no. (%)			0.0312*			0.0095*
0 (<700)	7 (13.2)	5 (27.8)		1 (2.9)	4 (30.8)	
1 (700–1,200)	36 (67.9)	6 (33.3)		26 (76.5)	5 (38.4)	
2 (≥1,200)	10 (18.9)	7 (38.9)		7 (20.6)	4 (30.8)	
Combined makers, no. (%)			0.2458			0.4146
Positive	29 (54.7)	7 (38.9)		14 (41.2)	3 (23.1)	
Negative	24 (45.3)	11 (61.1)		20 (58.8)	10 (76.9)	
Radiomic signature			<0.0001*			0.0059*
Median	1.771	0.215		1.913	1.024	
(Interquartile range)	(1.136–2.495)	(−0.258–0.981)		(1.059–2.685)	(0.251–1.461)	
Radiomic nomogram			<0.0001*			<0.0001*
Median	2.073	0.245		2.273	0.452	
(Interquartile range)	(1.064–2.989)	(−0.610–0.843)		(1.868–2.892)	(−0.047–0.957)	

*p values are calculated from univariate analysis between each clinical factor and corresponding LNM status. AGC, advanced gastric cancer; LNM, lymph node metastasis; MRI, magnetic resonance imaging; pT staging, pathological T staging; ADC, apparent diffusion coefficient; SD, standard deviation; *p < 0.05.*

corresponding logistic regression coefficients provided a linear mathematical formula to calculate a radiomic signature.

Performance Evaluation of Radiomic Signature

Pearson correlation coefficients were calculated to verify definite contribution of the radiomic signature in classifying LNM status in the training and internal validation cohorts. Receiver operating characteristic (ROC) curves and AUCs were used to evaluate the performance of radiomic signature in the three cohorts. Sensitivity, specificity, and accuracy results were also calculated.

Development of Radiomic Nomogram

Preoperative clinical factors shown in **Table 1** were taken into consideration to establish a more powerful predictive radiomic nomogram. In univariate analysis for selecting significant clinical factors in the training cohort, Mann-Whitney *U*-test was used for numerical variables, and Chi-square test and fisher's exact test were applied for categorical features. Subsequently, multivariate logistic regression was used to build a radiomic nomogram by integrating radiomic signature and significant clinical factors. The output of the radiomic nomogram is the probability of LNM.

Assessment of Radiomic Nomogram

The radiomic nomogram was assessed by ROC curves and AUC values in the training, internal validation, and external validation cohorts. Calibration curves as well as Hosmer-Lemeshow tests were used to assess the fitting degree of

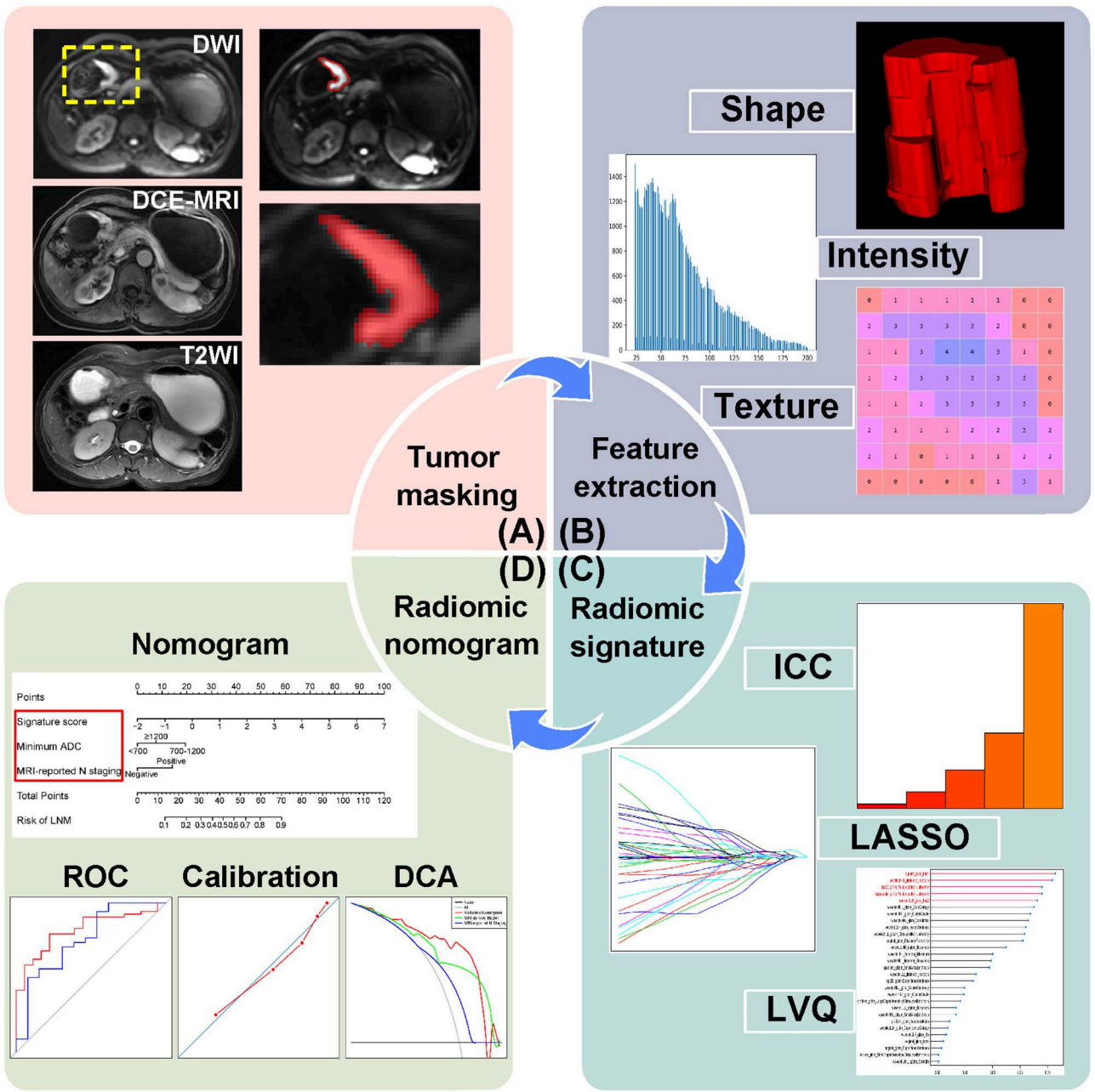


FIGURE 1 | Radiomics workflow of this study. **(A)** Tumor masking of AGC patients based on DWI. **(B)** Radiomic feature extraction, quantifying tumor shape, intensity, and texture. **(C)** Strategies for radiomic signature development. **(D)** Radiomic nomogram with evaluation of ROC, calibration curves, and DCA. AGC, advanced gastric cancer; DWI, diffusion-weighted imaging; DCE-MRI, dynamic contrast enhanced-magnetic resonance imaging; T2WI, T2-weighted images; ICC, intraclass correlation coefficient; LASSO, least absolute shrinkage and selection operator; LVQ, learning vector quantization; ROC, receiver operating characteristic; DCA, decision curve analysis.

radiomic nomogram. An MRI-derived model constructed by significant clinical factors and an MRI-reported N staging scheme was developed for comparison. Sensitivity, specificity, and accuracy results of comparative experiments were also calculated.

Decision curve analysis (DCA) was carried out in the internal validation cohort by quantifying the net benefits at some threshold probabilities and determining clinical use of radiomic nomogram.

Statistical Analysis

A two-sided $p < 0.05$ of every statistical test was deemed significantly different, and all analyses were based on R language

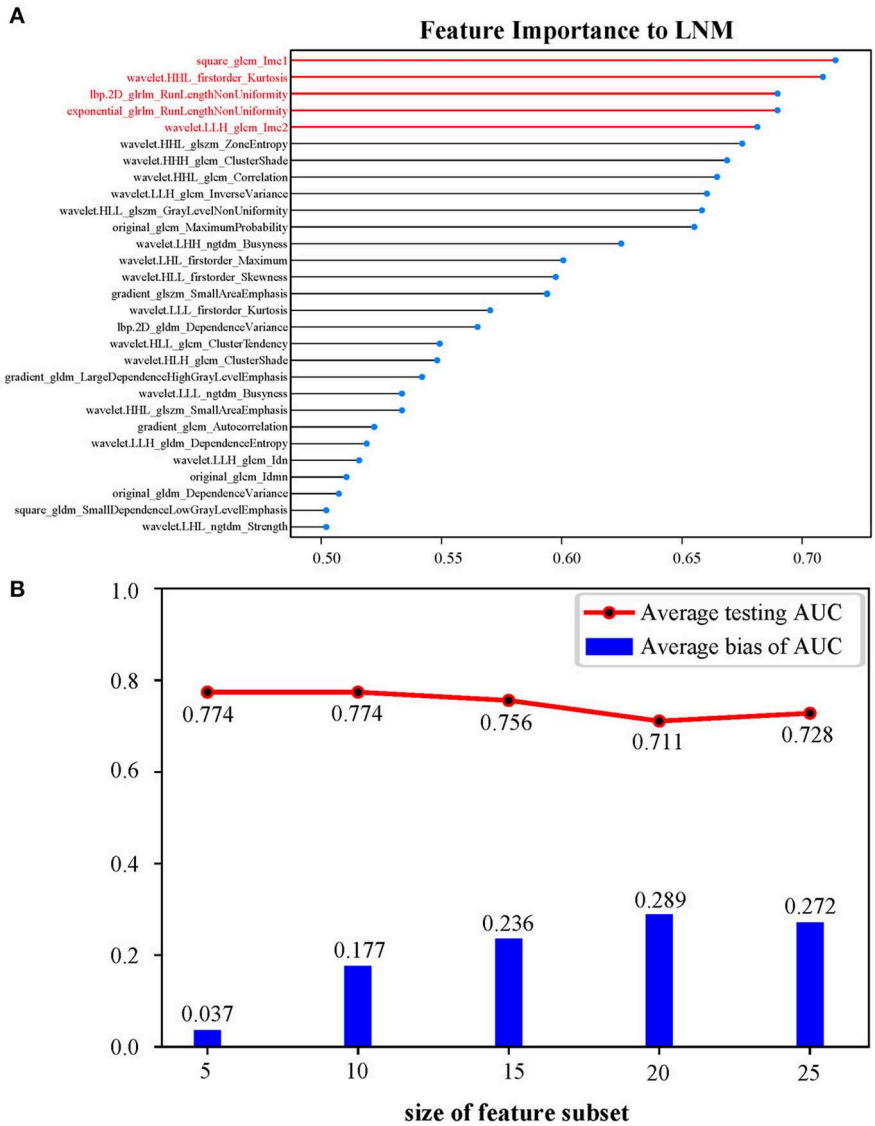


FIGURE 2 | (A) Twenty-nine features ranked in descending order of importance to LNM by LVQ. **(B)** OOB bootstrapping estimates with logistic regression of top 5, 10, 15, 20, and 25 features ranked by LVQ, respectively, confirming an optimal feature subset of five features. Average bias of AUC equals average training AUC minus average testing AUC. LNM, lymph node metastasis; LVQ, learning vector quantization; OOB, out-of-bag; AUC, area under the curve.

(version 3.4.3; https://www.r-project.org). R packages used in our work are described in **Supplementary Material 1.2**.

RESULTS

Clinical Factors

Baseline characteristics of patients from center 1 are shown in **Table 1**. LNM positive patients covered 74.6% (53/71) and 72.3% (34/47) of the training and internal validation cohorts, respectively, with no significant difference ($p = 0.7804$, Chi-square test) in LNM status between the two cohorts. There showed no significant statistical difference in sex ($p = 0.5384$), age ($p = 0.5039$), and all the other clinical factors ($p = 0.1202$–0.7747) between the two cohorts. LNM status had significant associations with MRI-reported N staging ($p = 0.0172$) and minimum ADC ($p = 0.0312$), while other clinical factors were excluded during the univariate analysis.

Feature Selection and Radiomic Signature Building

Among 1,305 original radiomic features per patient, 813 features were first selected after ICC analysis. Then, the multivariate LASSO method indicated 29 potential features (**Supplementary Figures 2A,B**). As shown in **Figure 2**, a logistic regression model consisting of features ranking the top 5 in

TABLE 2 | Performance evaluation of models in three cohorts.

Cohorts	Models	TP	TN	FN	FP	Sensitivity	Specificity	Accuracy	AUC (95% CI)
Training									
	MRI-reported N staging	42	9	11	9	0.792	0.500	0.718	0.646 (0.515–0.777)
	MRI-derived model	28	15	25	3	0.528	0.833	0.606	0.736 (0.602–0.871)
	Radiomic signature	42	14	11	4	0.792	0.778	0.789	0.821 (0.720–0.922)
	Radiomic nomogram	36	17	17	1	0.679	0.944	0.746	0.850 (0.758–0.942)
Internal validation									
	MRI-reported N staging	29	6	5	7	0.853	0.462	0.745	0.657 (0.504–0.811)
	MRI-derived model	21	12	13	1	0.618	0.923	0.702	0.818 (0.688–0.948)
	Radiomic signature	26	6	8	7	0.765	0.462	0.681	0.758 (0.591–0.925)
	Radiomic nomogram	29	11	5	2	0.853	0.846	0.851	0.857 (0.714–1.000)
External validation									
	MRI-reported N staging	17	5	4	2	0.810	0.714	0.786	0.762 (0.562–0.962)
	MRI-derived model	15	7	6	0	0.714	1.000	0.786	0.884 (0.765–1.000)
	Radiomic signature	20	1	1	4	0.952	0.429	0.821	0.741 (0.513–0.971)
	Radiomic nomogram	20	5	1	2	0.952	0.714	0.893	0.878 (0.696–1.000)

MRI, magnetic resonance imaging; TP, true positive; TN, true negative; FN, false negative; FP, false positive; AUC, area under the receiver operating characteristic curve; CI, confidence interval.

LVQ method gained a higher average testing AUC (0.774) and a smaller average bias between training AUC and testing AUC (0.037). After backward stepwise selection, two key features (*square_glcm_Imc1*, $p = 0.0013$; *wavelet.LLH_glcm_Imc2*, $p = 0.0062$) remained and made up the radiomic signature. Detailed explanations for the two radiomic features are given in **Supplementary Material 1.3**. The formula for the radiomic signature is given as below.

$$\begin{aligned}\text{Radiomic signature} = {} & -1.3383 \times square\_glcm \\ & \_Imc1 - 1.0139 \times wavelet.LLH \\ & \_glcm\_Imc2 + 1.5145\end{aligned}$$

The Performance of Radiomic Signature

There was a significant correlation (Pearson's $r = 0.448$, 0.432, and 0.458) between the radiomic signature and LNM status in the three cohorts. A significant difference ($p < 0.0001$) was found in radiomic signature [median (interquartile range)] between LNM and non-LNM groups in training cohort [1.771 (1.136–2.495) vs. 0.215 (−0.258–0.981), respectively]. This difference was confirmed in the validation cohort [1.913 (1.059–2.685) vs. 1.024 (0.251–1.461), $p = 0.0059$]. As estimated, patients with LNM generally got a higher radiomic signature score than those with non-LNM. The distinguishing ability of radiomic signature in training cohort and internal validation cohort was indicated with an AUC of 0.821 [95% confidence interval (CI), 0.720–0.922] and 0.758 (95% CI, 0.591–0.925), respectively. Furthermore, the AUC in external validation cohort achieved 0.741 (95% CI, 0.513–0.971). Detailed sensitivity, specificity, and accuracy results are presented in **Table 2**. Their corresponding 95% CI are attached in **Supplementary Table 3**.

Given the limited sample size, a 10-fold cross-validation in the center 1 cohort was conducted to avoid overfitting. Results given in **Table 3** indicated an average bias across 10-fold of 0.052 between training AUC values and validation AUC values. In addition, feature selection was conducted in each fold. The histogram in **Supplementary Figure 3** summarized the counts of selected feature's appearance, showing that the two radiomic features (*square_glcm_Imc1*, *wavelet.LLH_glcm_Imc2*) used in our radiomic signature appeared most frequently and were the most stable.

TABLE 3 | Ten-fold cross-validation to build radiomic signature in center 1 cohort.

Index	AUC		Bias	Number of features
	Training	Validation		
1	0.841	0.556	0.285	7
2	0.795	0.700	0.095	2
3	0.796	0.833	−0.037	2
4	0.788	1.000	−0.212	2
5	0.802	0.833	−0.031	2
6	0.710	0.722	−0.012	2
7	0.745	0.533	0.212	2
8	0.868	0.611	0.257	5
9	0.866	0.800	0.066	5
10	0.773	0.875	−0.102	2
Average bias	0.798	0.746	0.052	

Bias equals training AUC value minus validation AUC. Numbers of radiomic features selected in each fold are given. AUC, area under the receiver operating characteristic curve.

Development and Assessment of Radiomic Nomogram

A radiomic nomogram combining the radiomic signature, minimum ADC value, and MRI-reported N staging is shown in **Figure 3**. The formula for the radiomic nomogram is shown

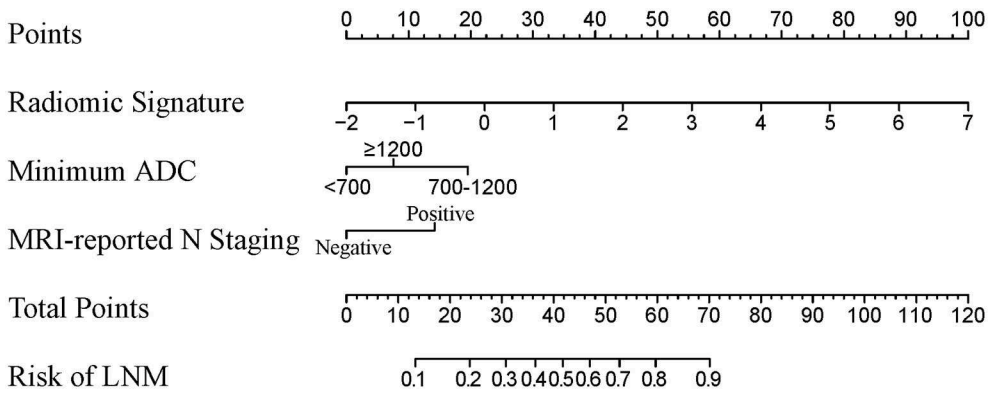


FIGURE 3 | A radiomic nomogram integrated radiomic signature, clinical minimum ADC value, and MRI-reported N staging. The value of each predictor can be converted into a risk score according to the "Points." After adding up the individual risk score of these predictors in "Total Points," we can get the corresponding prediction probability of LNM in "Risk of LNM" at the bottom. ADC, apparent diffusion coefficient; MRI, magnetic resonance imaging; LNM, lymph node metastasis.

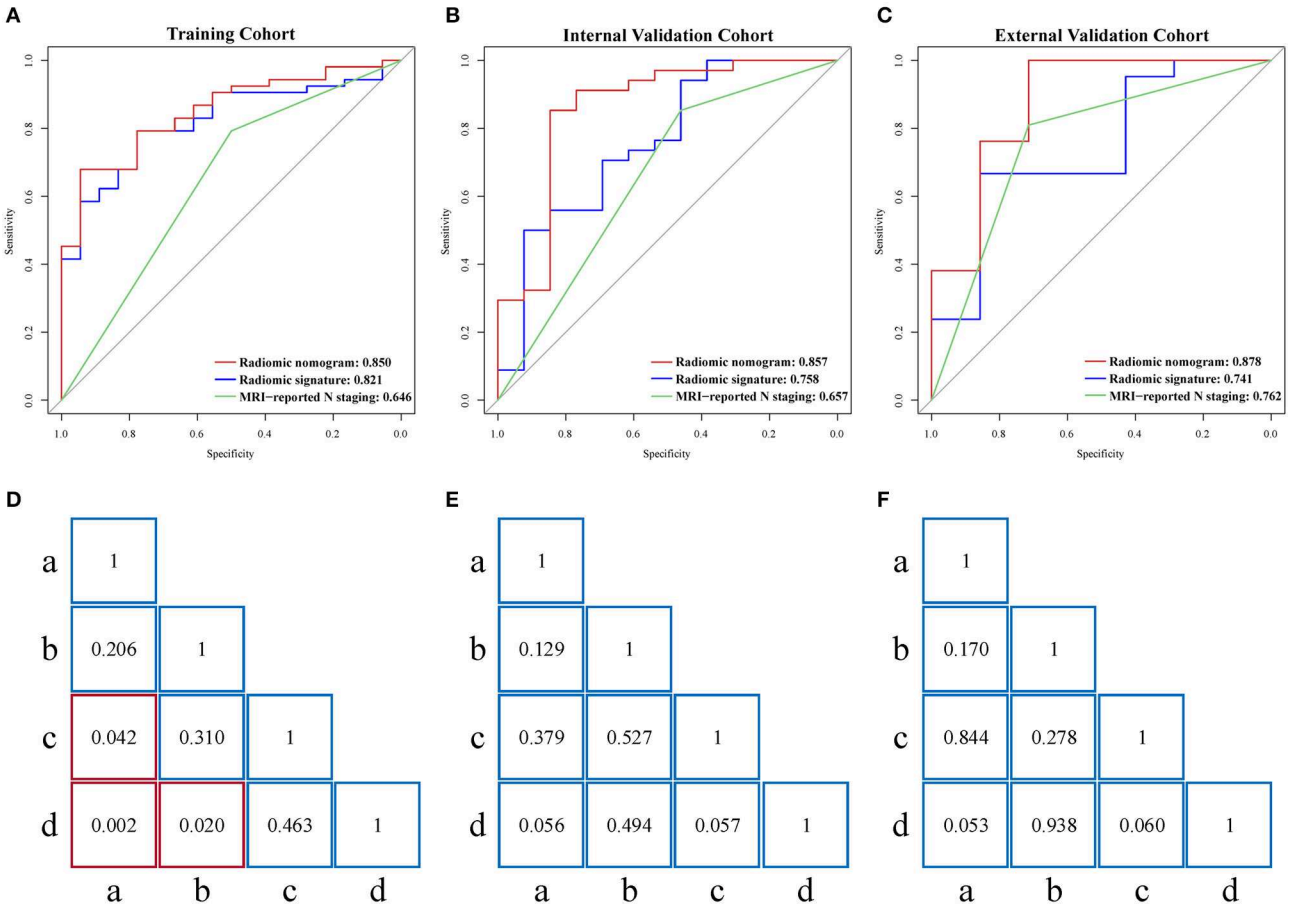


FIGURE 4 | (A–C) Performance of radiomic nomogram, radiomic signature, and MRI-reported N staging scheme in the training, internal validation, and external validation cohorts. **(D–F)** Delong tests for AUCs of four models in the three cohorts. Red box represents $p < 0.05$. "a", MRI-reported N staging; "b", MRI-derived model; "c", radiomic signature; "d", radiomic nomogram; MRI, magnetic resonance imaging; AUC, area under the curve.

as below, where "*IF(minimum ADC = 1)*" represents 700 ≤ minimum ADC < 1,200, "*IF(minimum ADC = 2)*" means minimum ADC ≥ 1,200, and "*IF(MRI-reported N staging = 1)*" represents positive MRI-reported N staging. MRI-derived model was built by minimum ADC and MRI-reported N staging.

$$\begin{aligned}\text{Radiomic nomogram} &= 0.8592 \times \text{Radiomic signature} \\ &+ 1.5085 \times IF\,(minimum\ ADC = 1) \\ &+ 0.5829 \times IF\,(minimum\ ADC = 2) \\ &+ 1.0957 \times IF\,(MRI\text{-}reported\ N\ staging \\ &= 1) - 1.5102\end{aligned}$$

There was a significant correlation (Pearson's $r = 0.530$, 0.602, and 0.677) between radiomic nomogram and LNM status in the three cohorts. ROC curves are given in **Figures 4A–C**. Sensitivity, specificity, and accuracy results are presented in **Table 2**. In the internal validation cohort, our radiomic nomogram showed good discrimination performance of LNM status and surpassed the routinely used MRI-reported N staging, reaching an AUC of 0.857 vs. 0.657, with an accuracy of 0.851 vs. 0.745, a specificity of 0.846 vs. 0.462, and a same sensitivity of 0.853. Compared with the MRI-derived model, our radiomic nomogram still showed superior predictive ability with an AUC of 0.857 vs. 0.818, an accuracy of 0.851 vs. 0.702, and a sensitivity of 0.853 vs. 0.618, although falling behind a little in specificity. In the external validation cohort, the radiomic nomogram also outperformed MRI-reported N staging in AUC (0.878 vs. 0.762), sensitivity (0.952 vs. 0.810), and accuracy (0.878 vs. 0.786). Similarly, the radiomic nomogram could still compensate the MRI-derived model for sensitivity and accuracy. **Figures 4D–F** show the quantitative AUC comparisons of the four models with Delong test.

As shown in **Supplementary Figure 4**, calibration curves of the radiomic nomogram suggested an agreement between model and actual outputs. Furthermore, DCA (**Figure 5**) indicated that the radiomic nomogram added more benefit when directing treatment decisions if the threshold probability was set between 0.24 and 0.86 compared with treat-none, treat-all, MRI-derived model, and MRI-reported N staging scheme.

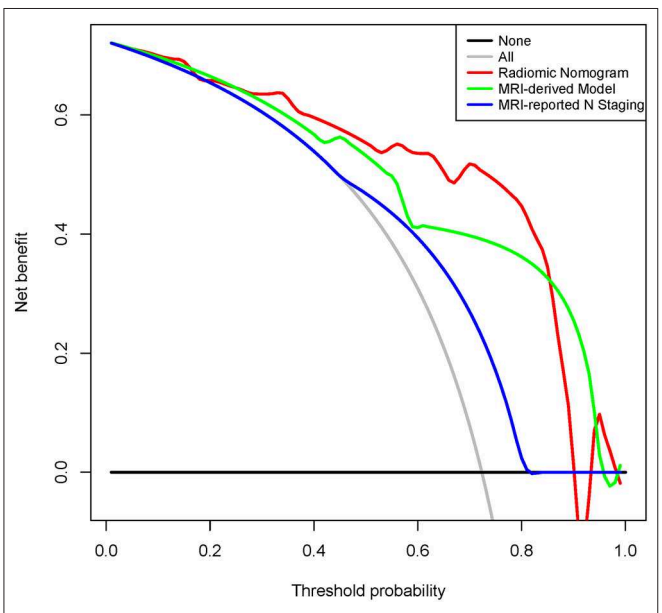


FIGURE 5 | Decision curve analysis for radiomic nomogram, MRI-derived model, and MRI-reported N staging in the internal validation cohort. Red line represents radiomic nomogram. Green line represents MRI-derived model. Blue line represents MRI-reported N staging. Gray line assumes all patients have LNM. Black line assumes no patients have LNM. MRI, magnetic resonance imaging; ADC, apparent diffusion coefficient; LNM, lymph node metastasis.

DISCUSSION

In this study, we established a radiomic nomogram which incorporated the radiomic signature and clinical factors including the minimum ADC value and MRI-reported N staging for non-invasive prediction of LNM in AGC patients. The radiomic nomogram showed better performance in determining and evaluating preoperative LNM status than clinical radiologists. The practical radiomic nomogram could facilitate a more accurate and objective assessment of LNM in AGC while providing personalized support for clinical decision making.

In terms of machine learning radiomics, typical LASSO method followed by OOB bootstrapping estimates of different feature subsets defined by LVQ was adopted to select crucial radiomic features in this study, which were later on fed to the generally used logistic regression for model building. Jiang's study (4) analyzed the association between computed tomography (CT)-based radiomic signature and LNM in gastric cancer using LASSO logistic regression. Taking a step forward, Wang's study (8) used ICC for feature selection and random forest algorithm to construct a radiomic signature. Upon the consistence in feature selection and model building with their studies, our radiomic models not only brought a novel view of LVQ in radiomics methods but also achieved similar model performance.

Radiomic features adopted in this study were both texture features about informational measure of correlation between local grayscale pixels calculated from gray level co-occurrence matrix. Results of cross-validation showed their great stability. Further analysis of these two features revealed that the radiomic signature score increased as the values of *square_glcm_Imc1* and *wavelet.LLH_glcm_Imc2* decreased according to the radiomic signature formula, which represented the uneven texture features of images and high heterogeneity of tumors. This suggested that radiomic signature could reflect a preclinical potential in establishing a connection between image information and LNM status.

LNM is an intricate biological process in AGC, in which the primary tumor lesions undoubtedly play an important role (14–16). Jiang's study (4) established a radiomic nomogram based on CT images and clinicopathological findings to estimate the LNM in patients with gastric cancer. However, the ROIs only covering the single maximum level of the tumor lesion may lead to incomplete radiomic features. Besides, some small lymph nodes have metastasized, while large lymph nodes may be simply caused by inflammation, so the judgment of CT-reported findings could also bring some bias. Compared to CT images, MRI signal variations are more visible to detect and diagnose qualitatively

(17–19). However, cases need to be noted that nodes with a diameter <8 mm or no obvious signal changes were later found to be metastatic nodes, while the opposite were found benign (20, 21). The low specificity in radiological diagnosis of LNM would preferentially overestimate the severity of disease and lead to excessive medical treatment. Without taking sample bias into account, a possible explanation was that while the tumor cells had already invaded into lymph nodes, changes in morphology and MRI signal were unlikely to present during the incubation period (18–21). The results of the current study thus showed the predictive power and potential for radiomics to reveal additional information invisible to the naked eye.

The ADC value mainly reflects tumor cell signal as a functional index that may provide an effective approach for the judgment of malignancy clinically (18, 20). Previous studies (3, 6, 19) have qualitatively studied the ADC value on the target lymph nodes, revealing a great correlation between low ADC value and metastatic nodes. Liu's study (22) showed that LNM had a correlation with ADC values of gastric cancer tissue. Traditional cognition displayed a greater tendency for a target node with a lower ADC to have greater malignancy (18). However, previous research results did not mention whether it was the average or the minimum ADC value, so many articles adopted default average ADC values. However, tumor heterogeneity should not only be expressed by a simple average ADC value but also by minimum ADC values, which would reflect the most heterogeneous ingredient in pathology type. By studying subjects with a minimum ADC value <700 and 600 mm^2/s in this dataset, 52.9% (9/17) and 100% (3/3) of the cases were pathologically confirmed as a negative LNM, respectively. A possible explanation for this result is that the tumor cells have high malignancy and increased cell alignment (23). Besides, unabsorbed hematoma may also lead to an extreme minimum ADC value. Even so, the nomogram indicated improved discrimination for nodal assessment with a reported accuracy of 85% compared to 75 and 50% for endoscopic ultrasound and CT, respectively (24). This was an innovative attempt in the image-data combination era in AGC, and further study would improve upon the construction and development of radiomic nomograms with increased sample sizes and upgraded iterations of technological computer-aided algorithms (25).

Considering the close relationship between T staging and the presence of peritoneal seeding (4, 17), we hypothesized a connection between LNM and T staging. However, the actual results showed no statistical significance ($p = 0.1726$). Some tumor indicators and combined markers have been shown to be associated with LNM in gastric cancer. However, the results showed no significant correlation in this study ($p = 0.2458$). An explanation for this phenomenon was that combined markers were only divided into negative and positive results. Some indicators only showed significance when they were many times higher than the normal value (26).

The quality of the VOI could directly affect final experimental results, as the VOI acquisition was the raw material of all procedures (9). DWI sequence was utilized, as the gastric lesions showed better contrast and clearer circumscription even if some lesions were in high-grade T staging or exhibited invasive growth (16). To avoid the influence of lymph node signals and visual judgment errors, each modality was necessary to combine multiple sequences as a reference to accurately judge the profile of lesions (14, 16, 17). The shape of gastric tumors was irregular in their appearance on cross-section due to the congenital differences in anatomical location and morphology (19). Hence, the VOI of AGC lesions may have some insufficiency in terms of volume- and shape-related radiomic signature.

Despite the advantages offered by the approaches presented herein, there are some limitations to be noted. The inherent selection bias and an incomplete dataset are known issues of retrospective studies. The robustness and reproducibility of the radiomic models, although validated in an external cohort, still need to be optimized using a larger sample size. Subsequent studies should subdivide the N staging into more categories, especially for N3a and N3b staging, which would be of great use for clinical decision making (16, 27).

In conclusion, this study provided a radiomic nomogram incorporating the radiomic signature, minimum ADC value, and MRI-reported N staging, to establish an effective method to improve the preoperative individualized predictive efficacy of LNM in AGC patients conveniently and accurately.

DATA AVAILABILITY STATEMENT

The datasets for this manuscript are not publicly available because of patient information privacy. Requests to access the datasets should be directed to corresponding author (MX), xums166@zcmu.edu.cn.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the First Affiliated Hospital of Zhejiang Chinese Medical University and Hangzhou Hospital of Traditional Chinese Medicine. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JT and MX conceived and launched this study. WC, SW, and DD designed the medical and statistical analyses. WC, KZ, and JL collected cases and acquired clinical information. XG and BL implemented the control of image quality and clinical diagnosis. SW and HL analyzed the data and carried out statistical experiments. SW, XW, and MF provided result interpretation. WC and SW wrote the first draft of this manuscript. DD, JT, and MX revised and edited the final version. XG, KZ, JL, BL, HL, XW, and MF reviewed the manuscript.

FUNDING

This work was supported by the Key Research Project of Zhejiang TCM Science and Technology Plan (2018ZZ010), the National Key R&D Program of China (2017YFA0205200, 2017YFC1308700, and 2017YFC1309100), National Natural Science Foundation of China (81971776, 81673745, 81771924,

81227901, 81671854, 81671851, 81527805, 61671449, and 61622117), the Zhejiang Province Natural Science Foundation (LY15H030033 and LSY19H030001), the Beijing Natural Science Foundation (L182061), and the Youth Innovation Promotion Association CAS (2017175).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01265/full#supplementary-material

REFERENCES

1. Rice TW, Gress DM, Patil DT, Hofstetter WL, Kelsen DP, Blackstone EH. Cancer of the esophagus and esophagogastric junction-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin.* (2017) 67:304–17. doi: 10.3322/caac.21399
2. Wang FH, Shen L, Li J, Zhou ZW, Liang H, Zhang XT, et al. The Chinese Society of Clinical Oncology (CSCO): clinical guidelines for the diagnosis and treatment of gastric cancer. *Cancer Commun.* (2019) 39:10. doi: 10.1186/s40880-019-0349-9
3. In H, Solsky I, Palis B, Langdon-Embry M, Ajani J, Sano T. Validation of the 8th edition of the AJCC TNM staging system for gastric cancer using the national cancer database. *Ann Surg Oncol.* (2017) 24:3683–91. doi: 10.1245/s10434-017-6078-x
4. Jiang Y, Wang W, Chen C, Zhang X, Zha X, Lv V, et al. Radiomics signature on computed tomography imaging: association with lymph node metastasis in patients with gastric cancer. *Front Oncol.* (2019) 9:340. doi: 10.3389/fonc.2019.00340
5. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin Cancer Res.* (2016) 22:5256–64. doi: 10.1158/1078-0432.CCR-15-2997
6. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuze S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol.* (2017) 28:1191–206. doi: 10.1093/annonc/mdx034
7. Meng Y, Zhang Y, Dong D, Li C, Liang X, Zhang CD, et al. Novel radiomic signature as a prognostic biomarker for locally advanced rectal cancer. *J Magn Reson Imaging.* (2018) 48:605–14. doi: 10.1002/jmri.25968
8. Wang Y, Liu W, Yu Y, Liu J, Xue H, Qi Y, et al. CT radiomics nomogram for the preoperative prediction of lymph node metastasis in gastric cancer. *Eur Radiol.* (2019). doi: 10.1007/s00330-019-06398-z. [Epub ahead of print].
9. Bickelhaupt S, Jaeger PF, Laun FB, Lederer W, Daniel H, Kuder TA, et al. Radiomics based on adapted diffusion kurtosis imaging helps to clarify most mammographic findings suspicious for cancer. *Radiology.* (2018) 287:761–70. doi: 10.1148/radiol.2017170273
10. Dong D, Tang L, Li Z, Fang M, Gao J, Shan X, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol.* (2019) 30:431–8. doi: 10.1093/annonc/mdz001
11. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2014) 5:4006. doi: 10.1038/ncomms5006
12. Zhang B, Tian J, Dong D, Gu DS, Dong YH, Zhang L, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin Cancer Res.* (2017) 23:4259. doi: 10.1158/1078-0432.CCR-16-2910
13. Schneider P, Biehl M, Hammer B. Relevance matrices in LVQ. In: *Dagstuhl Seminar Proceedings*. Wadern: Schloss Dagstuhl-Leibniz-Zentrum für Informatik (2007).
14. Ito S, Sano T, Mizusawa J, Takahari D, Katayama H, Katai H, et al. A phase II study of preoperative chemotherapy with docetaxel, cisplatin, and S-1 followed by gastrectomy with D2 plus para-aortic lymph node dissection for gastric cancer with extensive lymph node metastasis: JCOG1002. *Gastric Cancer.* (2017) 20:322–31. doi: 10.1007/s10120-016-0619-z
15. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol.* (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
16. Li Z, Li H, Wang S, Dong D, Yin F, Chen A, et al. MR-based radiomics nomogram of cervical cancer in prediction of the lymph-vascular space invasion preoperatively. *J Magn Reson Imaging.* (2018) 49:1420–6. doi: 10.1002/jmri.26531
17. Giganti F, Tang L, Baba H. Gastric cancer and imaging biomarkers: Part 1 – a critical review of DW-MRI and CE-MDCT findings. *Eur Radiol.* (2018) 29:1743–53. doi: 10.1007/s00330-018-5732-4
18. Hutchinson EB, Avram AV, Irfanoglu MO, Koay CG, Barnett AS, Komlosh ME, et al. Analysis of the effects of noise, DWI sampling, and value of assumed parameters in diffusion MRI models. *Magn Reson Med.* (2017) 78:1767–80. doi: 10.1002/mrm.26575
19. Giganti F, Orsenigo E, Arcidiacono PG, Nicoletti R, Albarello L, Ambrosi A, et al. Preoperative locoregional staging of gastric cancer: is there a place for magnetic resonance imaging? Prospective comparison with EUS and multidetector computed tomography. *Gastric Cancer.* (2016) 19:216–25. doi: 10.1007/s10120-015-0468-1
20. Joo I, Lee JM, Kim JH, Shin CI, Han JK, Choi BI. Prospective comparison of 3T MRI with diffusion-weighted imaging and MDCT for the preoperative TNM staging of gastric cancer. *J Magn Reson Imaging.* (2015) 41:814–21. doi: 10.1002/jmri.24586
21. Hu YF, Huang CM, Sun YH, Su XQ, Cao H, Hu JK, et al. Morbidity and mortality of laparoscopic versus open D2 distal gastrectomy for advanced gastric cancer: a randomized controlled trial. *J Clin Oncol.* (2016) 34:1350–7. doi: 10.1200/JCO.2015.63.7215
22. Liu S, Zhang YJ, Chen L, Guan WX, Guan Y, Ge Y, et al. Whole-lesion apparent diffusion coefficient histogram analysis: significance in T and N staging of gastric cancers. *BMC Cancer.* (2017) 17:665. doi: 10.1186/s12885-017-3622-9
23. Zhang Y, Chen J, Liu S, Shi H, Guan WX, Ji CF, et al. Assessment of histological differentiation in gastric cancers using whole-volume histogram analysis of apparent diffusion coefficient maps. *J Magn Reson Imaging.* (2016) 45:440–9. doi: 10.1002/jmri.25360
24. Sah BR, Owczarczyk K, Siddique M, Cook GJR, Goh V. Radiomics in esophageal and gastric cancer. *Abdom Radiol.* (2018) 44:2048–58. doi: 10.1007/s00261-018-1724-8
25. Han L, Zhu YB, Liu ZY, Yu T, He CJ, Jiang WY, et al. Radiomic nomogram for prediction of axillary lymph node metastasis in breast cancer. *Eur Radiol.* (2019) 29:3820–9. doi: 10.1007/s00330-018-5981-2
26. Yu JX, Zhang SG, Zhao BB. Differences and correlation of serum CEA, CA19-9 and CA72-4 in gastric cancer. *Mol Clin Oncol.* (2016) 4:441–9. doi: 10.3892/mco.2015.712
27. Ikoma N, Estrella JS, Blum M, Das P, Chen HC, Wang XM, et al. Central lymph node metastasis in gastric cancer is predictive of survival after preoperative therapy. *J Gastrointest Surg.* (2018) 22:1325–33. doi: 10.1007/s11605-018-3764-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 22 November 2019
doi: 10.3389/fonc.2019.01250

Radiomics Analysis of Iodine-Based Material Decomposition Images With Dual-Energy Computed Tomography Imaging for Preoperatively Predicting Microsatellite Instability Status in Colorectal Cancer

*Jingjun Wu[1], Qinhe Zhang[1], Ying Zhao[1], Yijun Liu[1], Anliang Chen[1], Xin Li[2], Tingfan Wu[2], Jianying Li[3], Yan Guo[3] and Ailian Liu[1]**

[1] Department of Radiology, First Affiliated Hospital, Dalian Medical University, Dalian, China, [2] Translational Medicine Team, GE Healthcare (China), Shanghai, China, [3] GE Healthcare (China), Shanghai, China

OPEN ACCESS

Edited by:
Chunxiao Guo,
University of Texas MD Anderson Cancer Center, United States

Reviewed by:
Di Dong,
Institute of Automation (CAS), China
Hui Li,
University of Chicago, United States

***Correspondence:**
Ailian Liu
liuailian@dmu.edu.cn

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 29 July 2019
Accepted: 30 October 2019
Published: 22 November 2019

Citation:
Wu J, Zhang Q, Zhao Y, Liu Y, Chen A, Li X, Wu T, Li J, Guo Y and Liu A (2019) Radiomics Analysis of Iodine-Based Material Decomposition Images With Dual-Energy Computed Tomography Imaging for Preoperatively Predicting Microsatellite Instability Status in Colorectal Cancer. Front. Oncol. 9:1250. doi: 10.3389/fonc.2019.01250

Purpose: The aim of this study was to investigate the value of radiomics analysis of iodine-based material decomposition (MD) images with dual-energy computed tomography (DECT) imaging for preoperatively predicting microsatellite instability (MSI) status in colorectal cancer (CRC).

Methods: This study included 102 CRC patients proved by postoperative pathology, and their MSI status was confirmed by immunohistochemistry staining. All patients underwent preoperative DECT imaging scanned on either a Revolution CT or Discovery CT 750HD scanner, and the iodine-based MD images in the venous phase were reconstructed. The clinical, CT-reported, and radiomics features were obtained and analyzed. Data from the Revolution CT scanner were used to establish a radiomics model to predict MSI status (70% samples were randomly selected as the training set, and the remaining samples were used to validate); and data from the Discovery CT 750HD scanner were used to test the radiomics model. The stable radiomics features with both inter-user and intra-user stability were selected for the next analysis. The feature dimension reduction was performed by using Student's *t*-test or Mann–Whitney *U*-test, Spearman's rank correlation test, min–max standardization, one-hot encoding, and least absolute shrinkage and selection operator selection method. The multiparameter logistic regression model was established to predict MSI status. The model performances were evaluated: The discrimination performance was accessed by receiver operating characteristic (ROC) curve analysis; the calibration performance was tested by calibration curve accompanied by Hosmer–Lemeshow test; the clinical utility was assessed by decision curve analysis.

Results: Nine top-ranked features were finally selected to construct the radiomics model. In the training set, the area under the ROC curve (AUC) was 0.961 (accuracy: 0.875; sensitivity: 1.000; specificity: 0.812); in the validation set, the AUC was 0.918 (accuracy: 0.875; sensitivity: 0.875; specificity: 0.857). In the testing set, the diagnostic

performance was slightly lower with AUC of 0.875 (accuracy: 0.788; sensitivity: 0.909; specificity: 0.727). A nomogram based on clinical factors and radiomics score was generated via the proposed logistic regression model. Good calibration and clinical utility were observed using the calibration and decision curve analyses, respectively.

Conclusion: Radiomics analysis of iodine-based MD images with DECT imaging holds great potential to predict MSI status in CRC patients.

Keywords: microsatellite instability, colorectal neoplasms, iodine-based material decomposition image, radiomics, dual-energy computed tomography

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related death worldwide (1). The occurrence and development of CRC are accompanied by a series of genetic abnormalities, of which microsatellite instability (MSI) is an important pathway in carcinogenesis (2). According to previous reports, even though MSI occurs in only approximately 15% of CRCs, it has gained considerable attention by clinicians owing to its significant value for CRC prognosis and treatment (2, 3). Microsatellite stability (MSS) status is maintained by the mismatch repair (MMR) genes, which are applied to repair genetic sequences that have been erroneous during replication in normal tissues. When MMR system is impaired, the error microsatellite sequences will accumulate, resulting in MSI and early onset of CRC (2). Obtaining MSI status is necessary because the MSI CRC tissues possess special biological behaviors, they are more likely to have a better prognosis and benefit from immunotherapy, and they may be resistant to fluorouracil chemotherapy (4). However, the methods for assessing MSI status including immunohistochemistry (IHC) and polymerase chain reaction (PCR) are all based on pathological tissues obtained by invasive methods. And these advanced biological tests have not been widely generalized owing to the limitation of advanced medical equipment in local institutions (5). Thus, development of non-invasive and cost-effective method for predicting MSI status could be meaningful for clinicians to obtain more diagnostic clues and guide further treatment strategies.

Given the growing number of applications in clinical diagnosis, dual-energy computed tomography (DECT) has been considered as a milestone in CT imaging because it can provide quantitative measurements to characterize the lesions (6). DECT can generate accurate iodine-based material decomposition (MD) images, which can reflect the vascularization of various tissues via measuring the contrast material (iodine) concentration (IC) (7–9). And the correlation between IC values and MSI status has been reported in previous studies (10, 11). However, from the iodine-based MD images, we can only routinely obtain the mean value of IC in lesions, and more imaging characteristics such as heterogeneity remain untapped. Radiomics analysis achieved the conversion of medical images to high-dimensional mineable data to quantitatively and comprehensively describe tissues' characteristics from imaging (12). Several scholars have reported that the radiomics features extracted from CT images showed some value in predicting MSI status in CRC patients; however, the diagnostic performance was limited (13, 14). Accordingly, we have presumed that the radiomics analysis of iodine-based MD images might serve as a non-invasive and reproducible way to preoperatively assess MSI status in CRC patients and set up a study to investigate its diagnostic efficacy.

Abbreviations: AUC, area under the ROC curve; CRC, colorectal cancer; DECT, dual-energy computed tomography; DICOM, digital imaging data and communications in medicine; GLCM, gray level co-occurrence matrix; GLZSM, grey-level zone size matrix; IC, iodine concentration; IHC, immunohistochemistry; LASSO, least absolute shrinkage and selection operator; MD, material decomposition; MMR, mismatch repair; MSI, microsatellite instability; MSS, microsatellite stability; PCR, polymerase chain reaction; ROC, receiver operating characteristic; ROI, region of interest.

MATERIALS AND METHODS

Patient Population

Our institutional review board approved this retrospective study with waiver of the informed consent. Patients examined in our institution from January 2016 to March 2019 who met the following criteria were included in our study. Inclusion criteria are as follows: (1) underwent curative-intent surgical resection and diagnosed as CRC by postoperative pathology; (2) underwent abdominal enhanced DECT examination within about 1 week before surgery; and (3) with MSI information tested by IHC staining in pathological report. Exclusion criteria are (1) with any local or systematic anticancer therapy (radiotherapy, chemotherapy, and biotherapy) before CT imaging; (2) without available digital imaging data and communications in medicine (DICOM) files in our system; (3) without available or complete clinical data; and (4) with invisible target lesion on CT images. According to the outcomes of MSI testing in the pathological report, we collected 653 CRC patients including 34 MSI CRC patients (incidence rate of 5.2%) and 619 MSS CRC patients. For further statistical analysis, 34 MSI CRC patients (23 scanned on Revolution CT and 11 scanned on Discovery CT 750HD) and 68 controls with MSS CRC (46 scanned on Revolution CT and 22 scanned on Discovery CT 750HD) in a 1:2 ratio (15) (randomly selected from 619 MSS CRC patients) were ultimately included in our study (61 males and 41 females; age: 63.82 $\pm$ 11.51 years; range 26–87 years). The flowchart of patient selection process is shown in **Figure 1**. The demographics of CRC patients is listed in **Table 1**. The clinical data of all CRC

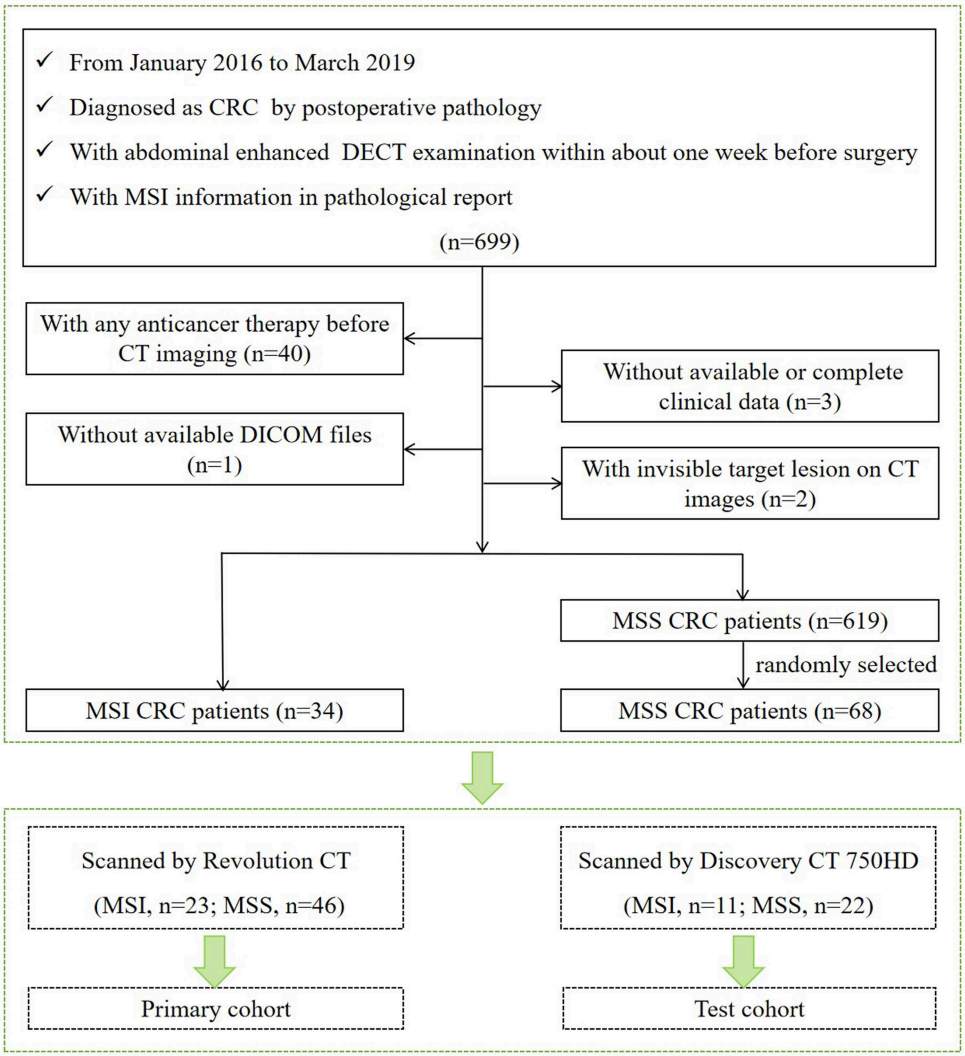


FIGURE 1 | Flowchart of patient selection process.

patients including age, gender, carcinoembryonic antigen (CEA) (normal level, 0–5 U/ml), carbohydrate antigen 19-9 (CA19-9) (normal level, 0–27 U/ml), alcohol history, smoking history, hypertension history, diabetes history, and family history of cancer were recorded. The included CRC patients were divided into two independent cohorts: (1) primary cohort: CRC patients examined on the Revolution CT scanner were used to establish a radiomics model to predict MSI status (70% samples were randomly selected as the training set, and the remaining samples were used to validate); and (2) testing cohort: CRC patients examined on the Discovery CT 750HD scanner were used to test the predictive model.

Microsatellite Instability Status Assessment

The MSI status was assessed by IHC staining of MMR proteins (MLH1, MSH2, PMS2, and MSH6). IHC staining was routinely performed based on postoperative tissues via standard streptavidin biotin-peroxidase procedure. According to the staining results of MMR proteins, patients were classified into the MSI or MSS group. CRC tissues with at least one of four negatively stained MMR proteins were defined as MSI CRC; others with four positively stained proteins were defined as MSS CRC (2).

Iodine-Based Material Decomposition Image Acquisition and Analysis

The abdominal DECT scans were performed on a Revolution CT scanner or Discovery CT 750HD scanner (GE Healthcare) in supine position. The non-enhanced abdominal CT scan was performed first with the conventional CT protocol of using the tube voltage of 120 kVp. The contrast-enhanced CT scans were performed using the dual-energy spectral CT scanning mode using the following scan parameters: helical, rapid switch

TABLE 1 | Demographics of CRC patients.

Characteristic	Primary cohort (Revolution CT) ($n = 69$)		Validation cohort (Discovery CT 750HD) ($n = 33$)	
	MSI ($n = 23$)	MSS ($n = 46$)	MSI ($n = 11$)	MSS ($n = 22$)
Age (years) (mean ± SD)	60.22 ± 11.95	63.5 ± 10.96	61.18 ± 10.83	69.59 ± 11.01
Gender, *n* (%)				
Male	12 (52.2)	28 (60.9)	7 (63.6)	14 (63.6)
Female	11 (47.8)	18 (39.1)	4 (36.4)	8 (36.4)
CEA, *n* (%)				
Normal	17 (73.9)	31 (67.4)	7 (63.6)	9 (40.9)
Abnormal	6 (26.1)	15 (32.6)	4 (36.4)	13 (59.1)
CA19-9, *n* (%)				
Normal	19 (82.6)	39 (84.8)	9 (81.8)	15 (68.2)
Abnormal	4 (17.4)	7 (15.2)	2 (18.2)	7 (31.8)
Alcohol history, *n* (%)				
Yes	2 (8.7)	4 (8.7)	2 (18.2)	2 (9.1)
No	21 (91.3)	42 (91.3)	9 (81.8)	20 (90.9)
Smoking history, *n* (%)				
Yes	1 (4.3)	8 (17.4)	2 (18.2)	2 (9.1)
No	22 (95.7)	38 (82.6)	9 (81.8)	20 (90.9)
Hypertension, *n* (%)				
Yes	7 (30.4)	21 (45.7)	1 (9.1)	5 (22.7)
No	16 (69.6)	25 (54.3)	10 (90.9)	17 (77.3)
Diabetes, *n* (%)				
Yes	3 (13)	8 (17.4)	1(9.1)	1 (4.5)
No	20 (87)	38 (82.6)	10 (90.9)	21 (95.5)
Family history of cancer, *n* (%)				
Yes	2 (8.7)	1 (2.2)	2 (18.2)	0 (0)
No	21 (91.3)	45 (97.8)	9 (81.8)	22 (100)
CT-reported tumor size (cm) (mean ± SD)	2.46 ± 1.41	1.83 ± 1.39	2.83 ± 1.76	2.11 ± 1.61
CT-reported tumor location, *n* (%)				
Right colon	12 (52.2)	14 (30.4)	7 (63.6)	13 (59.1)
Left colon	9 (39.1)	26 (56.5)	2 (18.2)	6 (27.3)
Rectum	2 (8.7)	6 (13)	2 (18.2)	3 (13.6)
CT-reported serous invasion, *n* (%)				
Yes	21 (91.3)	28 (60.9)	8 (72.7)	18 (81.8)
No	2 (8.7)	18 (39.1)	3 (27.3)	4 (18.2)
CT-reported lymph node invasion, *n* (%)				
Yes	17 (73.9)	32 (69.6)	8 (72.7)	9 (40.9)
No	6 (26.1)	14 (30.4)	3 (27.3)	13 (59.1)

CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; CRC, colorectal cancer; MSI, microsatellite instability; MSS, microsatellite stability.

between tube voltages of 80 and 140 kVp in 0.5 ms; tube current, 230–445 mA; detector width, 80 mm; helical pitch, 0.992:1 on the Revolution CT scanner and 1.375:1 on the Discovery CT 750HD scanner; rotation time, 0.6–0.8 s; slice thickness, 1.25 mm; and slice interval, 1.25 mm. For the contrast-enhanced CT scans, 1.2 ml/kg of non-ionic contrast media iohexol (Omnipaque 300 mg/ml, GE Healthcare) was used. The contrast medium was administered via the antecubital vein at an injection rate of 3 ml/s. The arterial phase, venous phase, and delayed phase scans were obtained after 30, 60, and 120 s following the administration of contrast agents. The CT scans covered the abdomen and pelvis from the dome of diaphragm to pubic symphysis. After CT scans, the iodine-based MD images in the venous phase were reconstructed at 1.25-mm image slice thickness and interval using the Gemstone Spectral Imaging (GSI) software on an advanced workstation 4.6 (AW 4.6; GE Healthcare).

Image analysis was performed by an abdominal radiologist with 3 years of experience and independently verified by another trained radiologist with 5 years of experience to reduce possible bias. Their discrepant interpretations were resolved via consultation. These observers were blinded to all clinical and pathological information of CRC patients. The following data extracted from CT images were analyzed and recorded: (a) tumor size, defined as the maximum axial diameter of tumors on images; (b) tumor location, subclassified as right colon, left colon, and rectum; (c) CT-reported serous invasion, defined as irregular projections from the serosal surface, and/or clouding of the pericolic fat, and/or loss of the normal fat planes, and/or thickened contiguous fascial reflections; (d) CT-reported lymph node invasion, defined as enlarged lymph node (short-axis diameter $>$ 1 cm), and/or clustered at least three lymph nodes (16).

Tumor Segmentation and Radiomics Feature Extraction

The ROI was placed by two experienced abdominal radiologists independently. Radiologist 1 (with 5 years of experience) performed the segmentation of all patients twice with a 6-month interval. Radiologist 2 (with 3 years of experience) performed the segmentation of all patients once. From the iodine-based MD images of venous phase, the two radiologists selected the slice with the largest axial diameter of CRC tumor and its adjacent upper and lower slices. Then, they manually outlined the boundary of the visible tumor on the selected slices via an open-source software ITK-SNAP (version 3.6.0) (17). The ROIs were required to include the area of necrosis and bleeding within the tumor and excluded perienteric fat and intestinal contents. To correct for acquisition-related differences of differing voxel resolutions in the two different CT scanners, voxel dimensions (mm) of each iodine-based MD image dataset were isotropically resampled to a common voxel spacing $0.5 \times 0.5 \times 0.5$ mm^3 (x, y, z) via linear interpolation algorithm (18, 19). Next, a total of 606 radiomics features for each CRC patient were extracted via Artificial Intelligent Kit (GE Healthcare) in concordance with the reference manual by the "Image Biomarker Standardization Initiative." These features were divided into four groups: (1) first-order histogram features ($n = 42$); (2) second-order texture features: gray level co-occurrence matrix (GLCM) ($n = 240$), Haralick features ($n = 10$); (3) grey-level zone size matrix (GLZSM) ($n = 11$); and (4) Gaussian transform ($n = 303$). The inter-user

variability for radiologist 1 and intra-user variability between radiologist 1 and radiologist 2 in tumor segmentation were analyzed via intraclass correlation coefficient (ICC) method [type: single rater; definition: absolute agreement; model: inter-user ICC: two-way random effects; intra-user ICC: two-way mixed effects (20)]. Details of radiomics features are described in **Figure 2**. The formulas of radiomics parameters are shown in **Supplementary I**.

Feature Selection and Prediction Model Building

The dimensionality reduction of all features including clinical, CT-reported, and radiomics features was performed based on training dataset for further analysis. First, the radiomics features with both inter-user and intra-user stability (with ICC values >0.90) were selected via ICC analysis. Second, the continuous features with significant differences ($p < 0.05$) between MSI and MSS groups were selected by Student's *t*-test (for normally distributed data) or Mann–Whitney *U*-test (for non-normally distributed data). Then, the categorical features (gender, CEA, CA19-9, alcohol history, smoking history, hypertension history, diabetes history, family history of cancer, tumor location, CT-reported serous invasion, and CT-reported lymph node invasion) were encoded by using one-hot encoding. One-hot encoding uses N-bit state registers to encode N status, each of which has its own register bits, and at any time, only one of them is valid. One-hot encoding can convert the category variables into a form readily available to machine learning algorithms (21). For example, the "CT-reported tumor location" has three status, right colon, left colon, and rectum, which were coded as "1, 2, 3" in our study first. Then we used one-hot encoding method to encode right colon, left colon, and rectum as 100, 010, 001, respectively. One-hot encoding method can ensure that "1, 2, 3" represents

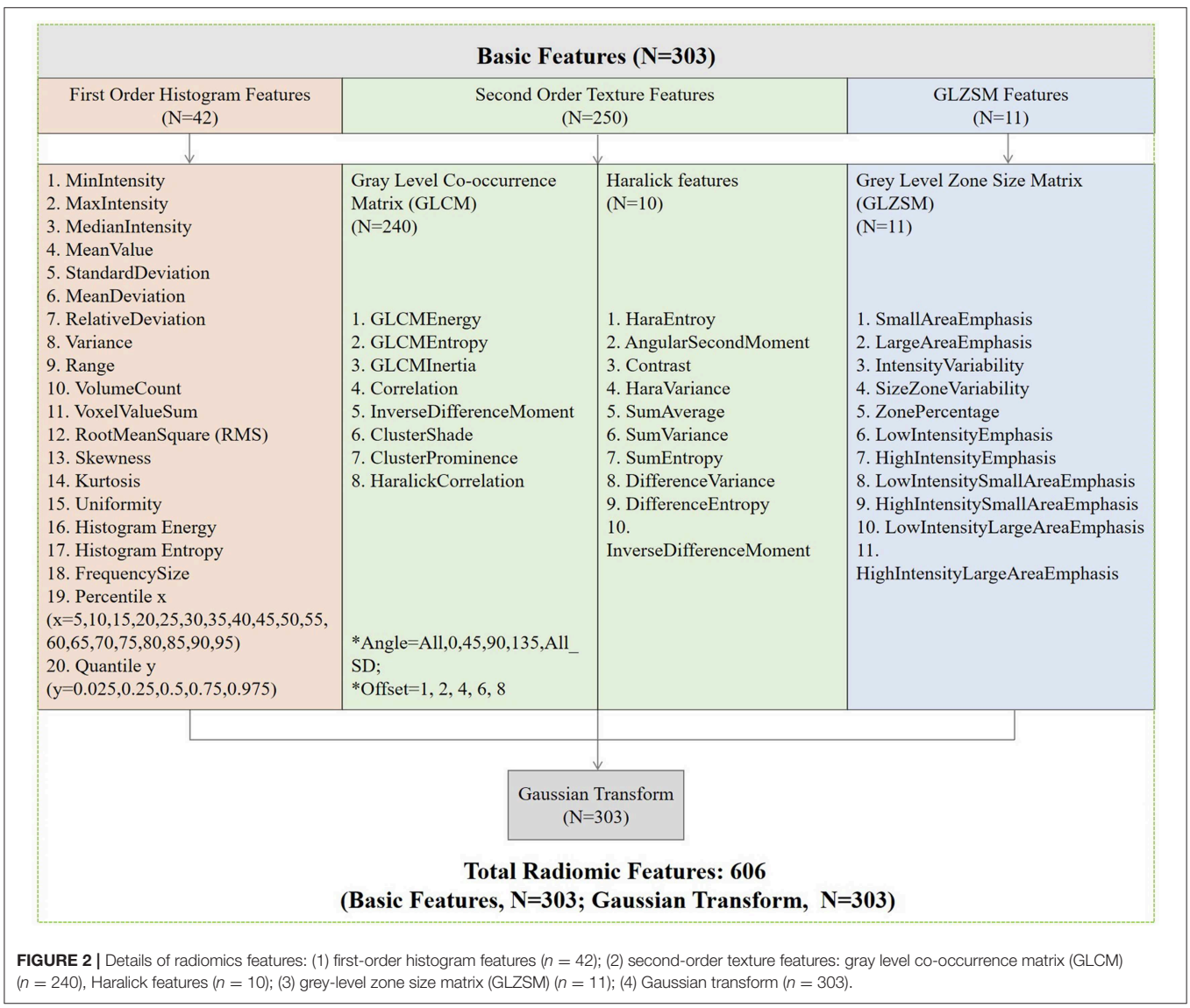


FIGURE 2 | Details of radiomics features: (1) first-order histogram features ($n = 42$); (2) second-order texture features: gray level co-occurrence matrix (GLCM) ($n = 240$), Haralick features ($n = 10$); (3) grey-level zone size matrix (GLZSM) ($n = 11$); (4) Gaussian transform ($n = 303$).

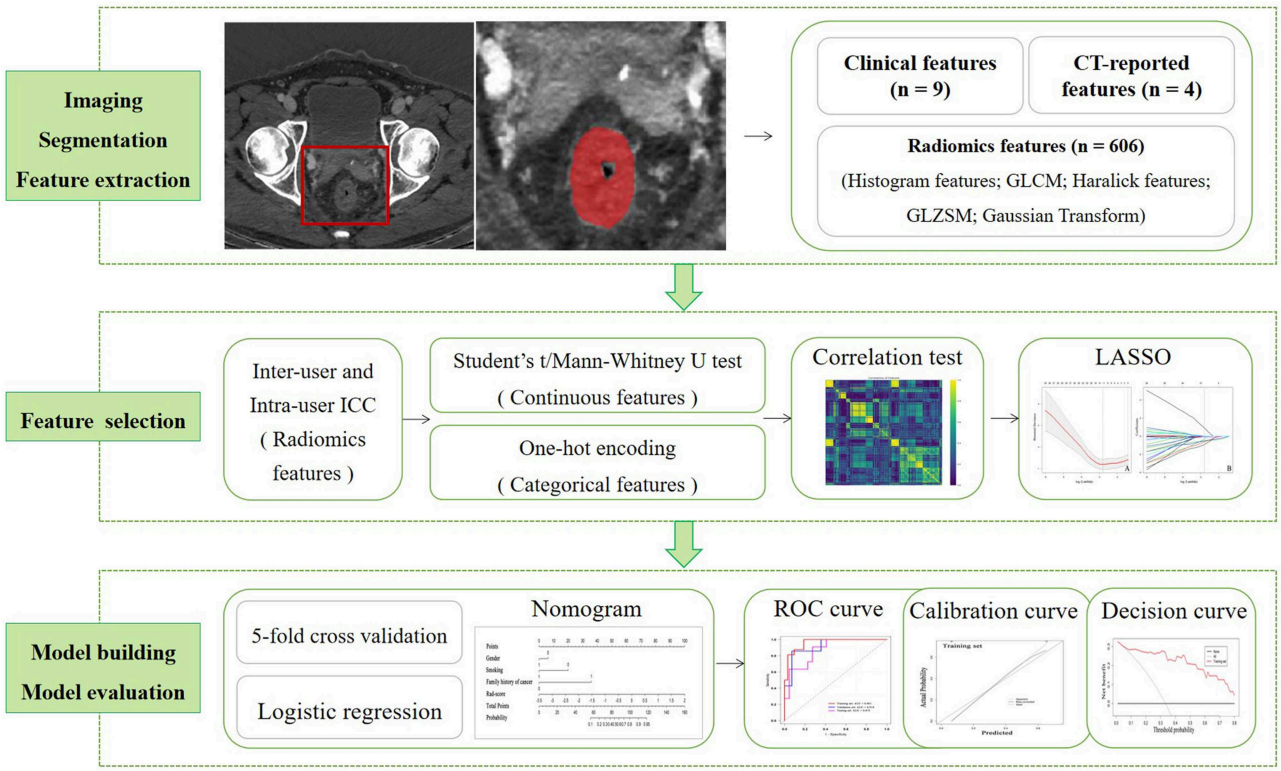


FIGURE 3 | Workflow of radiomics analysis: iodine-based MD imaging and segmentation; feature extraction, feature selection, model building, and model evaluation. MD, material decomposition.

the tumor location instead of the true value of 1, 2, or 3. Third, Spearman's rank correlation test was performed for each feature. Radiomics features with correlation coefficient $\geq$0.9 were selected into the following steps and then transferred with min–max standardization, whereas all features were normalized to a range of 0 to 1. Finally, the least absolute shrinkage and selection operator (LASSO) selection method was further used to identify the top-ranked and most valuable features to build the predictive model.

The selected features were applied to construct multiparameter logistic regression model to predict MSI status. The 5-fold cross-validation technique was used for model selection. The data in training set were divided into five subsets equally. Then, four subsets were selected each time to train, and the remaining one subset was used to test. By changing the subtest set in turn, five loss function values ($L(w)$) during the above five models would be obtained. The average value of $L(w)$ was calculated. When the average $L(w)$ reached a minimum value, the optimization of the logistic regression model would be completed, and the final model would be constructed. The details of $L(w)$ are shown in **Supplementary II**. A nomogram based on clinical factors and radiomics score was generated via the proposed logistic regression model. The probability of MSI status defined as a nomogram score can be calculated for each patient by using the developed nomogram. The data from the Revolution CT equipment were used to establish and validate the radiomics model, and the data from the Discovery CT 750HD equipment were used to test the radiomics model.

Radiomics Model Evaluation

The discrimination performance was accessed by using receiver operating characteristic (ROC) curve analysis. The area under the ROC curve (AUC), accuracy, sensitivity, and specificity was calculated. DeLong's test was used to compare the statistically difference between AUCs. The calibration performance was tested by using the calibration curve accompanied by the Hosmer–Lemeshow test (H-L test). The calibration curves measure the consistency between the predicted MSI status probability and the actual MSI status probability. The H-L test assesses the goodness of fit of the prediction models. The clinical utility of radiomics model was assessed by using decision curve analysis. For decision curve, the horizontal axis indicates the threshold probability with a range of 0.0 to 1.0. The vertical axis indicates the clinical net benefit values. There are two reference lines defined under the assumption that all patients are diagnosed to be either MSI or MSS. A larger area under the decision curve suggests a better clinical utility. All statistical analyses were conducted with R software (version 3.6.0; https://www.r-project.org/). The workflow of radiomics analysis is shown in **Figure 3**.

RESULTS

Inter-user and Intra-user Variabilities

The stable radiomics features (with ICC values >0.90) were as follows: 503 features between the two sets of measurements for radiologist 1, 568 features between the first measurement of radiologist 1 and radiologist 2, and 430 features between the second measurement of radiologist 1 and radiologist 2. Finally, 429 features were considered stable with both inter-user and intra-user stability. These 429 features obtained by radiologist 1 in the first measurement were used for the next analysis.

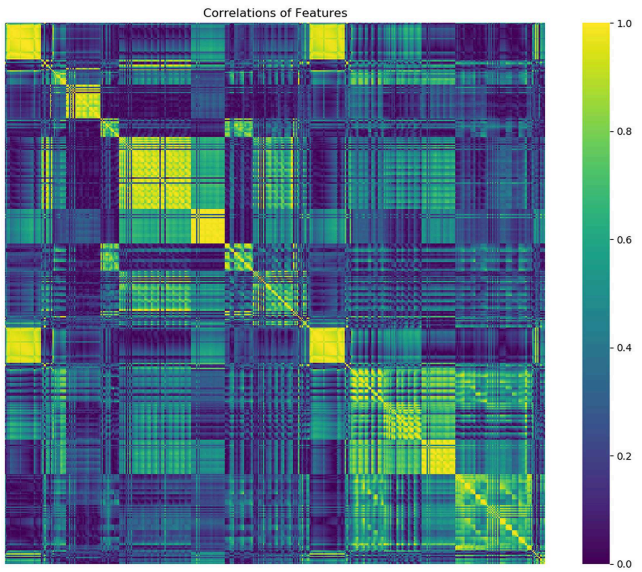


FIGURE 4 | The heat map gives a good visual representation of the feature correlation. The *x* axis and *y* axis indicate features. Color scale on the right side displays the absolute value of the correlation coefficient (higher from 0 to 1, and from blue to yellow).

Feature Selection and Radiomics Model Building

From a total of 429 radiomics features and 13 clinical or CT-reported features, the nine top-ranked features were finally selected for subsequent analysis: gender, smoking, family history of cancer, MaxIntensity, uniformity, GLCMEnergy_AllDirection_offset6_SD_Gaussian, GLCMEnergy_angle90_offset8_Gaussian, GLCMEntropy_AllDirection_offset8_Gaussian, and HaralickCorrelation_AllDirection_offset8_SD_Gaussian. The correlation heat map summarizes the correlations of features (**Figure 4**). Feature selection using the LASSO algorithm is shown in **Figure 5**. The nomogram based on clinical factors and radiomics score is shown in **Figure 6**.

$$
\begin{aligned}
Rad-score\ =\ & \\
-5.63e-01\ \times\ & MaxIntensity \\
-5.08e-01\ \times\ & uniformity \\
-1.76e-02\ \times\ & GLCMEnergy\_AllDirection\_offset6\_SD\_Gaussian \\
-8.25e-02\ \times\ & GLCMEnergy\_angle90\_offset8\_Gaussian \\
-3.70e-02\ \times\ & GLCMEntropy\_AllDirection\_offset8\_Gaussian \\
-6.30e-01\ \times\ & HaralickCorrelation\_AllDirection\_offset8\_SD \\
& \_Gaussian
\end{aligned}
$$

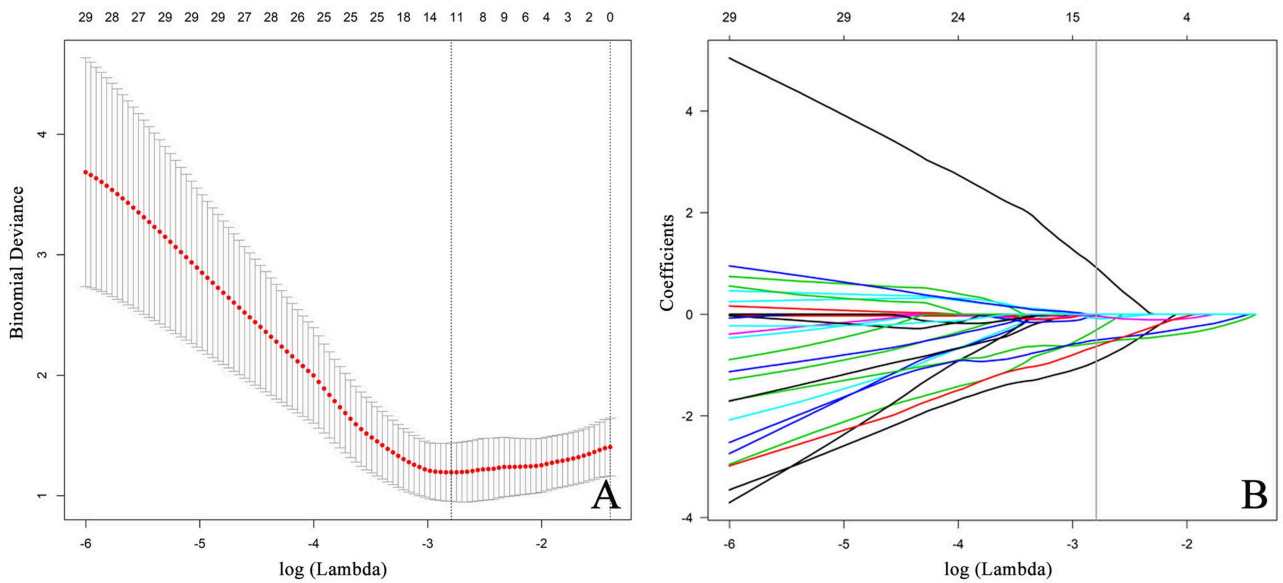


FIGURE 5 | Feature selection using the LASSO algorithm. **(A)** The LASSO tuning parameter (lambda,λ) is iteratively updated by the optimization theory, and the optimal values of λ are indicated by the dotted vertical lines, a value λ of 0.0612 with log(λ) = −2.79 is chosen. **(B)** The LASSO algorithm performs the trend of the coefficients in the feature selection process. A coefficient profile plot is generated by violating the log (λ) sequence. The parameter λ is optimized by a five-fold cross-validation technique, and when the loss function reaches a minimum, 12 variables are selected. The 12 variables correspond to nine features including three clinical features and six radiomics features. The vertical line indicates the coefficient size of each variable and the corresponding log(λ) value when the model is optimal. LASSO, least absolute shrinkage and selection operator.

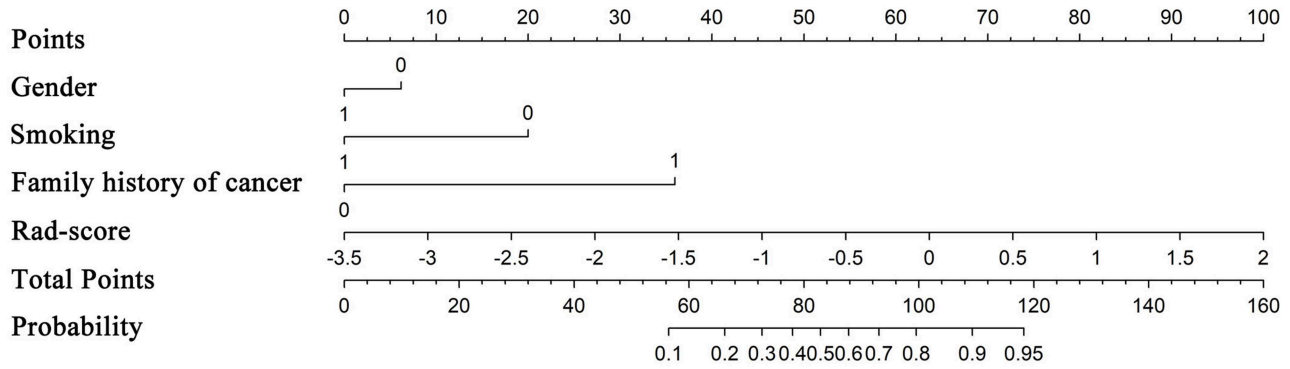


FIGURE 6 | Nomogram based on clinical factors and radiomics score.

Radiomics Model Evaluation

ROC analysis was applied to evaluate the model's performance for predicting MSI status. In the training set, AUC was 0.961 (95%CI [0.861, 0.996]; accuracy: 0.875; sensitivity: 1.000; specificity: 0.812); in the validation set, AUC was 0.918 (95%CI [0.714, 0.992]; accuracy: 0.875; sensitivity: 0.875; specificity: 0.857); and in the testing set, AUC was 0.875 (95%CI [0.715, 0.964]; accuracy: 0.788; sensitivity: 0.909; specificity: 0.727) (**Table 2**). DeLong's test revealed that above AUCs had no significant difference, which suggested that there was no overfitting among training, validation, and testing sets: $p = 0.535$ ($\Delta = 0.043$, 95%CI [0.0416, 0.0894]) between the training and validation sets; $p = 0.198$ ($\Delta = 0.085$, 95%CI [0.0266, 0.1476]) between the training and testing sets; and $p = 0.631$ ($\Delta = 0.042$, 95%CI [0.0224, 0.1492]) between the validation and testing sets. The ROC curves are shown in **Figure 7**. Good calibrations of radiomics models for predicting MSI status in training, validation, and testing sets are shown in **Figure 8**. The H-L test was not significant ($p > 0.05$), demonstrating a good fit (training set: $p = 0.462$; validation set: $p = 0.785$; testing set: $p = 0.568$). The decision curves for radiomics models in training, validation, and testing sets (with net benefit of 17.44, 15.40, and 13.43, respectively) are presented in **Figure 9**.

TABLE 2 | ROC analysis for predicting MSI status.

	Revolution CT		Discovery CT 750HD
	Training set	Validation set	Testing set
AUC	0.961	0.918	0.875
95%CI	[0.861, 0.996]	[0.714, 0.992]	[0.715, 0.964]
Accuracy	0.875	0.875	0.788
Sensitivity	1.000	0.875	0.909
Specificity	0.812	0.857	0.727

ROC, receiver operating characteristic; AUC, area under the ROC curve; CI, confidence interval; MSI, microsatellite instability.

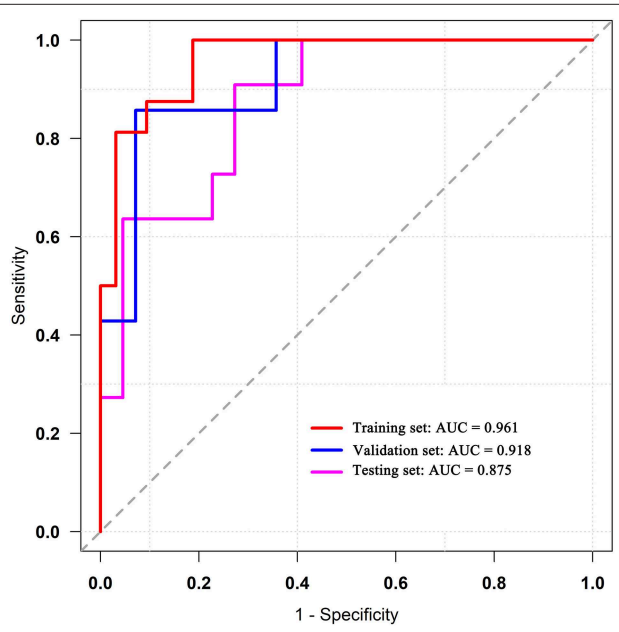


FIGURE 7 | ROC curves of the radiomics models in training, validation, and testing sets. ROC, receiver operating characteristic.

DISCUSSION

In our study, we established a radiomics model based on iodine-based MD images to predict MSI status in CRC patients before surgery. We achieved a good diagnostic performance based on data from the Revolution CT equipment in both the training set (AUC, 0.961) and validation set (AUC, 0.918). And this radiomics model was also suitable for the iodine-based MD images acquired on another CT equipment (Discovery CT 750HD) although with slightly lower diagnostic performance (AUC, 0.875).

Iodine-based MD images can quantitatively reflect the vascularization of tissues. A clear relationship between blood supply and IC values calculated from iodine-based MD images has been confirmed; the richer blood supply is accompanied

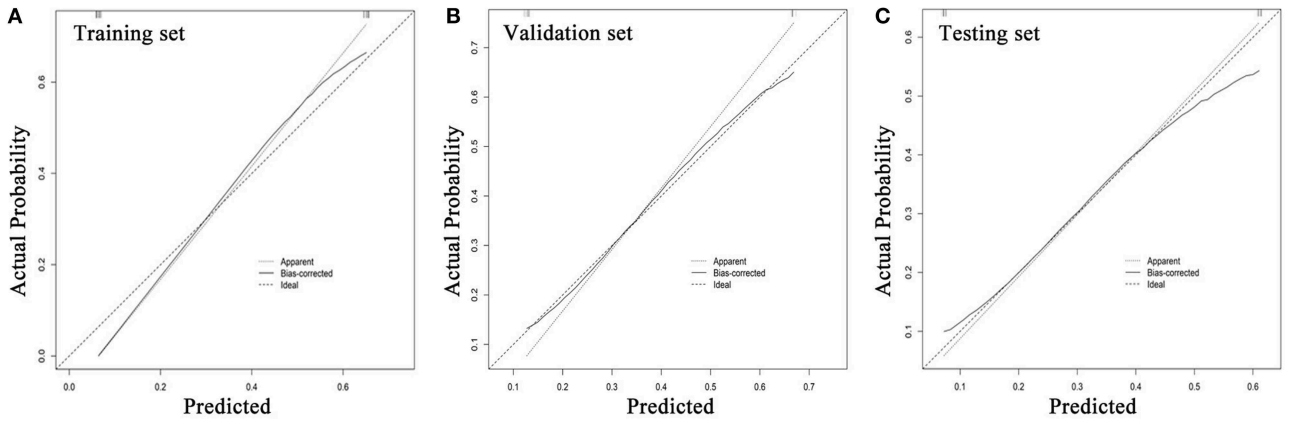


FIGURE 8 | Calibration curves for radiomics models in training **(A)**, validation **(B)**, and testing sets **(C)**. The diagonal dashed reference line represents a perfect estimated MSI status by an ideal model. Solid lines represent estimated MSI status of the model. Good alignment of diagonal dashed reference line and solid line indicates a good performance. MSI, microsatellite instability.

by a higher IC value (8, 22, 23). Our previous study has found that IC value of MSI CRC was significantly lower than that of MSS CRC and demonstrated the clinical value of using the IC values to distinguish MSI status with limited diagnostic performance (10). The measurement of IC in the previous study only reflected the average IC value in ROIs, and more information (such as tumor heterogeneity) was not evaluated. Therefore, radiomics approach, which can provide a wealth of complementary information of the images, should further extend our knowledge and improve the diagnosis (24–26). In our study, six radiomics features were finally selected as the most closely related features to the MSI status. For iodine-based MD images, the MaxIntensity generally represents the most abundant blood supply value within the predefined ROI. This may be explained by the biological characteristics of MSI CRC with less angiogenic phenotype confirmed by previous investigations (11, 27). The uniformity is a measure of the sum of the squares of each intensity value. From the perspective of image smoothness, the higher the intensity value, the higher the uniformity of the image. GLCMEnergy_AllDirection_offset6_SD_Gaussian and GLCMEnergy_angle90_offset8_Gaussian describe the uniformity of the intensity level distribution. GLCMEntropy_AllDirection_offset8_Gaussian describes the randomness of image values. It mainly calculates the average amount of information to encode image values. HaralickCorrelation_AllDirection_offset8_SD_Gaussian measures the linear dependency of grey levels of neighboring pixels; in other words, it measures the similarity of the grey levels in neighboring pixels and tells how correlated a pixel is to its neighbor over the whole image (28, 29). They have all served as recognized parameters to reflect tumor heterogeneity. We reviewed the biological differences between MSI and MSS tumors and tried to explain the imaging heterogeneity observed in this study. De Smedt et al. suggested that the morphological heterogeneity was the most striking feature to distinguish MSI from MSS CRC. Histologically, MSI CRC is often more inclined to present with a mixed morphological patterns including glandular, mucinous, and solid content, which caused the tumor heterogeneity (30). In addition, the higher incidence of internal heterogeneity in MSI CRC may also be explained by a higher density of tumor-infiltrating lymphocytes and a lower cell proliferation rate than MSS CRC (31, 32). Our results that the imaging heterogeneity was a biomarker for MSI tumors were consistent with those of previous studies (14). During radiomics analysis, integrating diverse clinical features plays an important role in improving the performance of the diagnostic model. We recorded the clinical features and CT reported features, which were discrete data except for age and tumor size. We used the one-hot encoding to process category variables, with the main benefits of one-hot encoding as follows: (1) to solve the problem that the classifier is not good at processing category data; (2) to a certain extent also play a role in expanding features; and (3) to choose the most representative new features. In our study, we found that the gender, smoking, and family history of cancer were closely related with the MSI status in CRC patients, and further explorations were required based on larger samples (13, 33).

Radiomics analysis is a promising method to unveil large amount of tumor features hidden in medical images. However, previous studies have reported that the repeatability of radiomics features can be influenced by different CT scanners (34). Our study included data from two different DECT scanners including Revolution CT and Discovery CT 750HD. We first used the data obtained from the Revolution CT scanner to establish a model for preoperatively predicting MSI status; the AUCs of training and validation sets were 0.961 and 0.918, respectively. Subsequently, we analyzed whether this radiomics model was suitable for another DECT scanner (Discovery CT 750HD), and we found that the performance was good with AUC of 0.875. Our results suggested that the radiomics model established in this study was applicable to both Revolution CT and Discovery CT 750HD, and this might be attributed to the stability of the iodine quantification, and that there is little effect of various DECT scanners and acquisition parameters on iodine

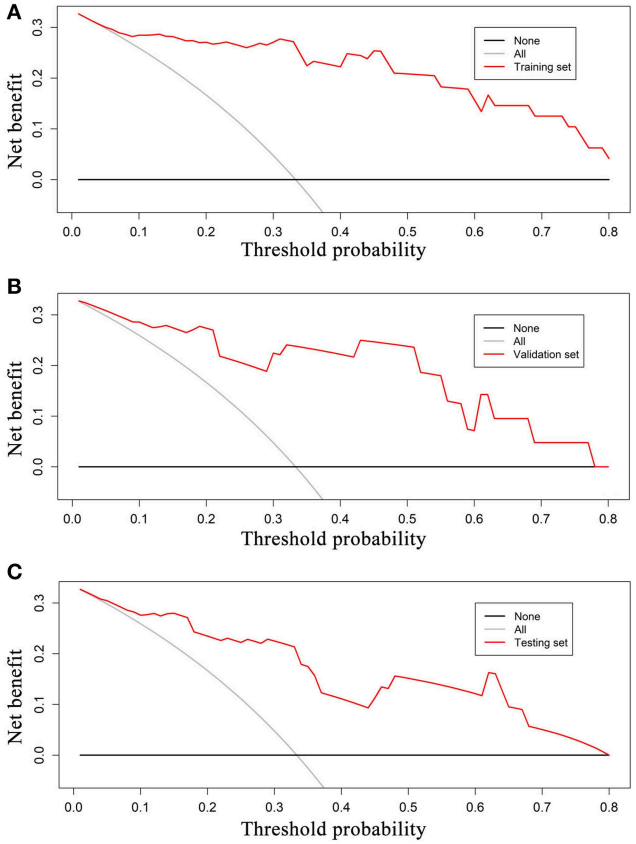


FIGURE 9 | Decision curves for radiomics models in training **(A)**, validation **(B)**, and testing sets **(C)**. The grey reference line indicates the assumption that all patients are diagnosed to be MSI CRC; black dashed reference line indicates the assumption that all patients are diagnosed to be MSS CRC. In training, validation, testing sets, the net benefit is 17.44, 15.40, and 13.43, respectively. A larger area under the decision curve indicates a better clinical utility. MSI, microsatellite instability; CRC, colorectal cancer; MSS, microsatellite stability.

density (35). Hence, further studies are recommended to focus on the radiomics analysis of iodine-based MD images with DECT imaging.

Our study has several limitations. First, the study was retrospective and may result in inherent biases. Second, although IHC test is a reliable way to assess MSI status, the PCR should still be recommended. Third, only a handful of patients were analyzed owing to the low incidence rate of MSI in CRC patients. Further studies are required using a larger sample. Fourth, only three slices of CT images were analyzed, and we plan to compare the performance of using three slices and whole tumors in future investigations. Fifth, some discrepancies caused by manually outlined ROIs are unavoidable, even though we had made efforts to minimize the bias by using two trained radiologists. Sixth, our data were only from a single center. In the future, we will try to collect multicenter data to reinforce the conclusions of our study.

In conclusion, radiomics analysis based on iodine-based MD images with DECT imaging can provide a relatively high diagnostic value for predicting MSI status in CRC patients. This study provides insight into the potential applications of using radiomics analysis of iodine-based MD images produced via DECT in predicting MSI status, and its usefulness for preoperatively providing more information in CRC clinical outcome and treatment decision making.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The First Affiliated Hospital of Dalian Medical University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Data analysis and interpretation, study design, manuscript writing, and manuscript approval were performed by JW, QZ, and AL, and they are accountable for all aspects of the work. CT data and pathological data analysis and interpretation, statistical analysis, and manuscript approval were performed by YZ. CT data analysis and manuscript approval were performed by YL and AC. Statistical analysis and manuscript approval were performed by XL, TW, JL, and YG.

FUNDING

This work was supported by the Program for Training Capital Science and Technology Leading Talents (grant number Z181100006318003). The funding body contributes to the design of the study and analysis of data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01250/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin.* (2017) 67:104–117. doi: 10.3322/caac.21395
2. Gelsomino F, Barbolini M, Spallanzani A, Pugliese G, Cascinu S. The evolving role of microsatellite instability in colorectal cancer: a review. *Cancer Treat Rev.* (2016) 51:19–26. doi: 10.1016/j.ctrv.2016.10.005
3. Chang L, Chang M, Chang HM, Chang F. Microsatellite instability: a predictive biomarker for cancer immunotherapy. *Appl Immunohistochem*

Mol Morphol. (2017) 26:e15–21. doi: 10.1097/PAI.0000000000000575
4. Rd BA, Venook AP, Cederquist L, Chan E, Chen YJ, Cooper HS, et al. Colon cancer, version 1.2017, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw.* (2017) 15:370–98. doi: 10.6004/jnccn.2017.0036
5. Yan W, Hu J, Xie L, Cheng L, Yang M, Li L, et al. Prediction of biological behavior and prognosis of colorectal cancer patients by tumor MSI/MMR in the Chinese population. *Onco Targets Ther.* (2016) 9:7415–24. doi: 10.2147/OTT.S117089
6. Goo HW, Goo JM. Dual-energy CT: new horizon in medical imaging. *Korean J Radiol.* (2017) 18:555–69. doi: 10.3348/kjr.2017.18.4.555
7. Muenzel D, Lo GC, Yu HS, Parakh A, Patino M, Kambadakone A, et al. Material density iodine images in dual-energy CT: detection and characterization of hypervascular liver lesions compared to magnetic resonance imaging. *Eur J Radiol.* (2017) 95:300–6. doi: 10.1016/j.ejrad.2017.08.035
8. Mirus M, Tokalov SV, Wolf G, Heinold J, Prochnow V, Abolmaali N. Noninvasive assessment and quantification of tumour vascularisation using MRI and CT in a tumour model with modifiable angiogenesis - an animal experimental prospective cohort study. *Eur Radiol Exper.* (2017) 1:15. doi: 10.1186/s41747-017-0014-5
9. Chen XH, Ren K, Liang P, Chai Y, Chen KS, Gao JB. Spectral computed tomography in advanced gastric cancer: can iodine concentration non-invasively assess angiogenesis? *World J Gastroenterol.* (2017) 23:1666–75. doi: 10.3748/wjg.v23.i9.1666
10. Wu J, Lv Y, Wang N, Zhao Y, Zhang P, Liu Y, et al. The value of single-source dual-energy CT imaging for discriminating microsatellite instability from microsatellite stability human colorectal cancer. *Eur Radiol.* (2019) 29:3782–90. doi: 10.1007/s00330-019-06144-5
11. Wendum D, Boëlle PY, Rigau V, Sebbagh N, Olschwang S, Mourra N, et al. Mucinous colon carcinomas with microsatellite instability have a lower microvessel density and lower vascular endothelial growth factor expression. *Virchows Arch.* (2003) 442:111–7. doi: 10.1007/s00428-002-0737-3
12. Aerts HJWL, Rios Velazquez E, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Erratum: Corrigendum: decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2014) 5:4006. doi: 10.1038/ncomms5644
13. Fan S, Li X, Cui X, Zheng L, Ren X, Ma W, et al. Computed tomography-based radiomic features could potentially predict microsatellite instability status in stage II colorectal cancer: a preliminary study. *Acad Radiol.* (2019). doi: 10.1016/j.acra.2019.02.009. [Epub ahead of print].
14. Golia Pernicka JS, Gagniere J, Chakraborty J, Yamashita R, Nardo L, Creasy JM, et al. Radiomics-based prediction of microsatellite instability in colorectal cancer at initial computed tomography evaluation. *Abdom Radiol.* (2019) 44:3755–63. doi: 10.1007/s00261-019-02117-w
15. Adebanji A, Nokoe S, Adeyemi A. Effects of sample size ratio on the performance of the quadratic discriminant function. *J Nat Sci Eng Technol.* (2010) 3:97–108.
16. Grant LA, Griffin N. *Grainger & Allison's Diagnostic Radiology Essentials.* London: Churchill Livingstone (2013).
17. Yushkevich PA, Piven J, Cody Hazlett H, Gimpel Smith R, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage.* (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015
18. Chirra P, Bloch NB, Rastinehead A, Purysko A, Madabhushi A, Viswanath SE, et al. Empirical evaluation of cross-site reproducibility in radiomic features for characterizing prostate MRI. *Computer-aided Diagnosis.* (2018). doi: 10.1117/12.2293992
19. Shen TX, Liu L, Li WH, Fu P, Xu K, Jiang YQ, et al. CT imaging-based histogram features for prediction of EGFR mutation status of bone metastases in patients with primary lung adenocarcinoma. *Cancer Imaging.* (2019) 19:34. doi: 10.1186/s40644-019-0221-9
20. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* (2016) 15:S1556370716000158. doi: 10.1016/j.jcm.2016.02.012
21. Qiao Y, Yang X, Wu E. The research of BP neural network based on one-hot encoding and principle component analysis in determining the therapeutic effect of diabetes mellitus. In: *IOP Conference Series: Earth and Environmental Science*. Xian (2019). p. 267.
22. Hawighorst H, Knapstein PG, Knopp MV, Vaupel P, Kaick GV. Cervical carcinoma: standard and pharmacokinetic analysis of time–intensity curves for assessment of tumor angiogenesis and patient survival. *MAGMA.* (1999) 8:55–62. doi: 10.1016/S1352-8661(99)00007-1
23. Zhang XF, Lu Q, Wu LM, Zou AH, Hua XL, Xu JR. Quantitative iodine-based material decomposition images with spectral CT imaging for differentiating prostatic carcinoma from benign prostatic hyperplasia. *Acad Radiol.* (2013) 20:947–56. doi: 10.1016/j.acra.2013.02.011
24. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin.* (2019) 69:127–57. doi: 10.3322/caac.21552
25. Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol.* (2019) 30:431–8. doi: 10.1093/annonc/mdz001
26. Shuo W, Jingyun S, Zhaoxiang Y, Di D, Dongdong Y, Mu Z, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J.* (2019) 53:1800986. doi: 10.1183/13993003.00986-2018
27. Wendum D, Comperat E, Boëlle PY, Parc R, Masliah J, Trugnan G, et al. Cytoplasmic phospholipase A2 alpha overexpression in stromal cells is correlated with angiogenesis in human colorectal cancer. *Mod Pathol.* (2005) 18:212–20. doi: 10.1038/modpathol.3800284
28. Kim JH, Ko ES, Lim Y, Lee KS, Han BK, Ko EY, et al. Breast cancer heterogeneity: MR imaging texture analysis and survival outcomes. *Radiology.* (2016) 282:160261. doi: 10.1148/radiol.2016160261
29. Liu S, Zheng H, Pan X, Chen L, Shi M, Guan Y, et al. Texture analysis of CT imaging for assessment of esophageal squamous cancer aggressiveness. *J Thorac Dis.* (2017) 9:4724–32. doi: 10.21037/jtd.2017.06.46
30. De Smedt L, Lemahieu J, Palmans S, Govaere O, Tousseyn T, Van Cutsem E, et al. Microsatellite instable vs stable colon carcinomas: analysis of tumour heterogeneity, inflammation and angiogenesis. *Brit J Cancer.* (2015) 113:500–9. doi: 10.1038/bjc.2015.213
31. Pauline M, David T, Mohamad H, Bernhard M, Hafid K, Gabriela B, et al. Correlation between density of CD8+ T-cell infiltrate in microsatellite unstable colorectal cancers and frameshift mutations: a rationale for personalized immunotherapy. *Cancer Res.* (2015) 75:3446–55. doi: 10.1158/0008-5472.CAN-14-3051
32. Sinicrope FA, Rego RL, Garrity-Park MM, Foster NR, Sargent DJ, Goldberg RM, et al. Alterations in cell proliferation and apoptosis in colon cancers with microsatellite instability. *Int J Cancer.* (2007) 120:1232–8. doi: 10.1002/ijc.22429
33. Zhang L, Zhao J, Yu B, Song X, Sun G, Han L, et al. Correlations between microsatellite instability, ERCC1/XRCC1 polymorphism and clinical characteristics, and FOLFOX adjuvant chemotherapy effect of colorectal cancer patients. *Cancer Genet.* (2017) 218–219:51–7. doi: 10.1016/j.cancergen.2017.09.004
34. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, Castro-García M, Villas MV, Mansilla Legorburo F, et al. Radiomics of ct features may be nonreproducible and redundant: influence of ct acquisition parameters. *Radiology.* (2018) 288:172361. doi: 10.1148/radiol.2018172361
35. Kim H, Park CM, Kang CK, Yoon J, Chae KJ, Goo JM. Effect of CT Acquisition parameters on iodine density measurement at dual-layer spectral CT. *AJR Am J Roentgenol.* (2018) 211:19381. doi: 10.2214/AJR.17.19381

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 28 November 2019
doi: 10.3389/fonc.2019.01330

T2-Weighted Image-Based Radiomics Signature for Discriminating Between Seminomas and Nonseminoma

Peipei Zhang[1], Zhaoyan Feng[1], Wei Cai[1], Huijuan You[1], Chanyuan Fan[1], Wenzhi Lv[2], Xiangde Min[1] and Liang Wang[1]**

[1] Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2] Julei Technology, Wuhan, China

Objective: To evaluate the performance of a T2-weighted image (T2WI)-based radiomics signature for differentiating between seminomas and nonseminomas.

Materials and Methods: In this retrospective study, 39 patients with testicular germ-cell tumors (TGCTs) confirmed by radical orchiectomy were enrolled, including 19 cases of seminomas and 20 cases of nonseminomas. All patients underwent 3T magnetic resonance imaging (MRI) before radical orchiectomy. Eight hundred fifty-one radiomics features were extracted from the T2WI of each patient. Intra- and interclass correlation coefficients were used to select the features with excellent stability and repeatability. Then, we used the minimum-redundancy maximum-relevance (mRMR) and the least absolute shrinkage and selection operator (LASSO) algorithms for feature selection and radiomics signature development. Receiver operating characteristic curve analysis was used to evaluate the diagnostic performance of the radiomics signature.

Results: Five features were selected to build the radiomics signature. The radiomics signature was significantly different between the seminomas and nonseminomas ($p < 0.01$). The area under the curve (AUC), sensitivity, and specificity of the radiomics signature for discriminating between seminomas and nonseminomas were 0.979 (95% CI: 0.873–1.000), 90.00 (95% CI: 68.3–98.8), and 100.00 (95% CI: 82.4–100.0), respectively.

Conclusion: The T2WI-based radiomics signature has the potential to non-invasively discriminate between seminomas and nonseminomas.

Keywords: magnetic resonance imaging, T2-weighted imaging, testicular neoplasms, testicular germ cell tumors, radiomics

OPEN ACCESS

Edited by:
Rong Tian,
Sichuan University, China

Reviewed by:
Di Dong,
Institute of Automation (CAS), China
Zhenyu Shu,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

***Correspondence:**
Xiangde Min
minxiangde0129@126.com
Liang Wang
wang6@tjh.tjmu.edu.cn

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *23 July 2019*
Accepted: *14 November 2019*
Published: *28 November 2019*

Citation:
Zhang P, Feng Z, Cai W, You H, Fan C, Lv W, Min X and Wang L (2019) T2-Weighted Image-Based Radiomics Signature for Discriminating Between Seminomas and Nonseminoma. Front. Oncol. 9:1330. doi: 10.3389/fonc.2019.01330

INTRODUCTION

Testicular cancer represents 1% of neoplasms and 5% of urological tumors in males. However, testicular cancer is the most common malignancy among men aged between 14 and 44 years (1, 2). Statistics show that there were 71,105 new cases and 9,507 deaths of testicular cancer worldwide in 2018 (3). Approximately 90–95% of testicular cancers are testicular germ cell tumors (TGCTs), which are split into two broad categories: seminomas and nonseminomas (4).

Radical orchiectomy is the main treatment for testicular tumors and can be supplemented by radiotherapy and chemotherapy (4, 5). In view of the different sensitivities of seminomas and nonseminomas to radiotherapy and chemotherapy, characterizing the histologic type of testicular tumors is of great importance (6–8). For patients undergoing orchidectomy, the differentiation of seminomas from nonseminomas would not affect patient management. However, the information gained preoperatively might help physicians to explain the patient's condition and tumor prognosis before surgery, which would help decrease the patient's anxiety. However, for patients who are unwilling to undergo orchiectomy, the seminomas, and nonseminomas must be identified by other non-invasive means, such as imaging examinations, because the guidelines do not recommend that patients with suspected testicular tumors undergo punctures in order to avoid tumor spread and metastasis (5). Therefore, several studies have evaluated the value of sonography or magnetic resonance imaging (MRI) for the non-invasive differentiation of seminomas from nonseminomas (4, 9–11).

Currently, ultrasonography (US) is the initial imaging method for confirming the existence of a testicular mass (5, 12). MRI has emerged as a valuable modality that can be an alternative diagnostic tool, especially in cases of non-diagnostic or equivocal sonographic findings (13). Compared to US, MRI can provide more abundant anatomical and functional information and is less dependent on operator technique. Some MRI features of TGCTs have been found to closely correlate with histopathologic characteristics (4, 9). T2-weighted imaging (T2WI) is an essential component of MRI in oncology. Some previous studies reported that seminomas and nonseminomas have different features on T2WI (8, 9). Most of the previous studies only used qualitative features or limited quantitative features, which may not fully explore the potential value of MRI.

Radiomics uses advanced image processing techniques to extract a large number of quantitative features from imaging data (14–16). It has been applied to various diseases such as lung and head-and-neck cancer (17), gastric cancer (18), colorectal cancer (19), liver fibrosis (20), and prostate cancer (21), etc., and remarkably encouraging results have been reported. However, to date, no study has applied radiomics to the evaluation of testicular diseases.

The purpose of our study was to investigate whether a T2WI-based radiomics signature could differentiate seminomas from nonseminomas.

MATERIALS AND METHODS

Patient Information

Our institutional review board approved this retrospective study. From February 2014 to March 2019, patients were included according to the following inclusion criteria (**Figure 1**): (a) had scrotal lesions on sonography or physical examination, (b) underwent a preoperative 3T MRI examination, (c) underwent radical orchiectomy, and (d) had pathologically confirmed TGCTs. Patients were excluded if apparent susceptibility or movement artifacts existed on the MR images. A total of 39 men (age range, 18–61 years; median age, 29 years) with 39 lesions were included. Nineteen tumors were pathologically confirmed as seminomas, and 20 tumors were pathologically confirmed as nonseminomas. The patients with nonseminomas had embryonal carcinomas ($n = 8$), teratomas ($n = 4$), yolk sac tumor ($n = 1$), and mixed germ cell tumors ($n = 7$) [embryonal carcinomas and teratomas ($n = 3$), seminoma and embryonal carcinoma ($n = 1$), teratoma, yolk sac tumor and embryonal carcinoma ($n = 1$), seminoma, teratoma and yolk sac tumor ($n = 1$), and seminoma, embryonal carcinoma, teratoma, and yolk sac tumor ($n = 1$)]. The classification of the tumor types in the current study was based on the NCCN guideline (22).

MRI Protocol

All the MR images were acquired with a 3T MR scanner (MAGNETOM Skyra, Siemens Healthcare, Erlangen, Germany) and an 18-element body matrix coil in combination with a 32-channel spine coil. The patients were positioned in a feet-first supine position. Transverse, sagittal and coronal T2-weighted turbo spin-echo sequences with the following parameters were used: repetition time/echo time (TR/TE) range 6500-6870/104 ms, slice thickness of 3∼5 mm, interslice gap of 0∼0.5 mm, field of view (FOV) of 180 × 180 mm^2, and a matrix of 384 × 320. Transverse T1-weighted turbo spin-echo sequences were acquired with the following parameters: TR of 750 ms, TE of 13 ms, slice thickness of 3∼5 mm, interslice gap of 0∼0.5 mm, FOV of 300×300 mm^2, and matrix of 320 × 240. Diffusion-weighted imaging (DWI) and dynamic contrast enhanced (DCE) sequences were performed for some patients, but these images were not included in the analysis due to the limited number of scans.

MRI Segmentation and Radiomics Feature Extraction

ITK-SNAP software (version 3.4.0; www.itksnap.org) was used for manual segmentation. Preoperative transverse T2WI was obtained for image analysis. A three-dimensional volume of interest (VOI) covering the tumor was delineated by stacking regions of interest slice-by-slice on the transverse T2WI. Manual segmentation of the tumors on the images was initially performed by a radiologist (Reader 1). Twenty patients were randomly selected from the study cohort. One month later, Reader 1 performed a second segmentation of the 20 patients to assess the intraobserver reproducibility. Another radiologist (Reader 2) performed a manual segmentation of these patients independently to assess the interobserver reproducibility. Both readers were blinded to the histologic results.

The radiomics features were extracted using the PyRadiomics library (https://github.com/Radiomics/pyradiomics.git, version 2.1.2) in Python (version 3.7.0). PyRadiomics is a flexible open-source platform capable of extracting a large panel of engineered features from medical images (23). For the feature extraction method, please reference the PyRadiomics documentation (https://pyradiomics.readthedocs.io/en/latest/). All MRI data were subjected to images normalization and resampled to the same resolution (0.46875 × 0.46875 × 3 mm) before feature extraction. A total of 851 radiomics features were extracted,

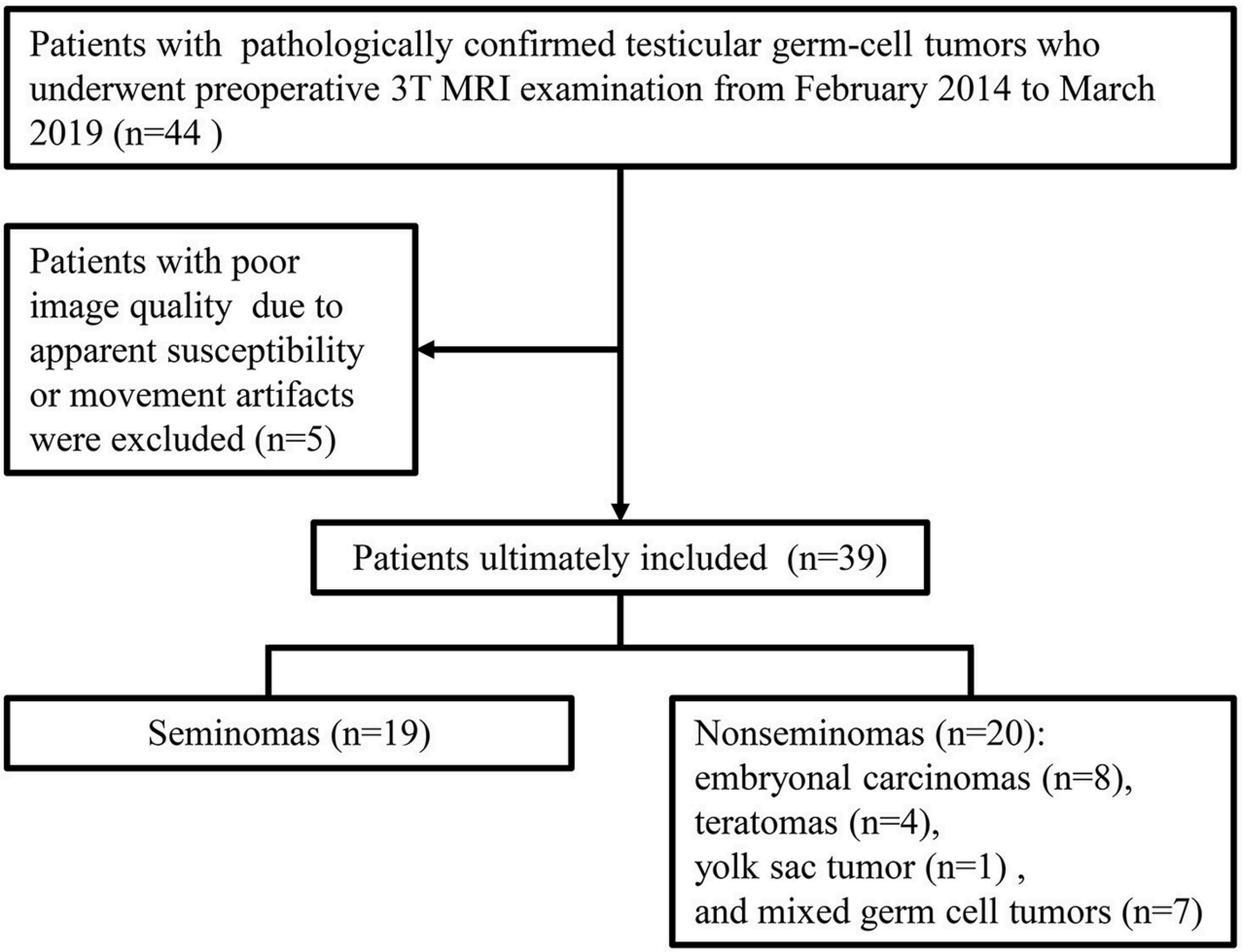


FIGURE 1 | Flow chart of patients' inclusion and exclusion.

including the following four groups: 14 shape features, 18 first-order intensity statistics features, 75 texture features [Gray Level Co-occurrence Matrix (24), Gray Level Size Zone Matrix (16), Gray Level Run Length Matrix (16), Neighboring Gray Tone Difference Matrix (5), and Gray Level Dependence Matrix (14)], and 744 wavelet features.

Statistical Analysis

As high-dimensional features were extracted in the current study, we performed a feature dimension reduction process to select the most relevant features for the classification of testicular lesions to construct a radiomics signature. Features selection included the following steps. First, we used the intra- and interclass correlation coefficient (ICC) to assess the effects of the manual segmentation variations on the value of the features. The ICC was calculated for each radiomics feature. Features with good agreement (ICC $\geq$ 0.8) were regarded as robust features and selected for the following analyses. Second, we compared all the features between seminomas and nonseminomas using the Mann-Whitney *U* test for non-normally distributed features or the independent *t*-test for normally distributed features. Features with $p < 0.05$ were considered significant variables and selected. To control the false-positive rate in multiple comparisons, the false discovery rate-adjusted *p*-value was used in the Mann-Whitney *U* test and the independent *t*-test (24). Third, spearman's correlation coefficient was used to compute the relevance and redundancy of the features. Redundant features indicated by a Spearman's correlation coefficient $\geq$ 0.8 were eliminated. Fourth, we applied the maximum relevance minimum redundancy (mRMR) algorithm to assess the relevance and redundancy of the remaining features (25). The mRMR algorithm was used to select the most relevant features for the classification of testicular lesions, avoiding redundancy between features. By using the mRMR method, the features were ranked according to their relevance-redundancy scores (mRMR scores). The mRMR score of a feature is defined as the mutual information between the status of the lesions and this feature minus the average mutual information of previously selected features and this feature. The top 10 features with high-relevance and low-redundancy were selected for the following analyses. Fifth, the 10 features selected by the above steps were applied to least absolute shrinkage and selection operator (LASSO) logistic

regression model (26). The LASSO logistic regression model with 5-fold cross-validation was adopted for further features selection and radiomics signature construction. LASSO is a regression analysis method that performs feature selection and regularization to improve the mode prediction accuracy and interpretability. Some candidate features coefficients were shrunk to zero and the remaining variables with non-zero coefficients were selected by LASSO. Then, the selected features were linearly combined to construct a radiomics signature.

The differences in the radiomics signature between seminomas and nonseminomas were compared using the Mann-Whitney *U* test. The diagnostic performance of the radiomics signature was evaluated using the receiver operating characteristic (ROC) curve. The area under the curve (AUC), sensitivity, and specificity were calculated. In addition, the diagnostic performance of the top 10 features selected from mRMR was also evaluated using ROC curve analysis. An overview of the radiomics signature development process is presented in **Figure 2**.

The statistical analyses were performed using R software (version 3.3.4; https://www.r-project.org). The following R packages were used: the "corrplot" package was used to calculate Spearman's correlation coefficient; the "mRMRe" package was used to implement the mRMR algorithm; the "glmnet" was used

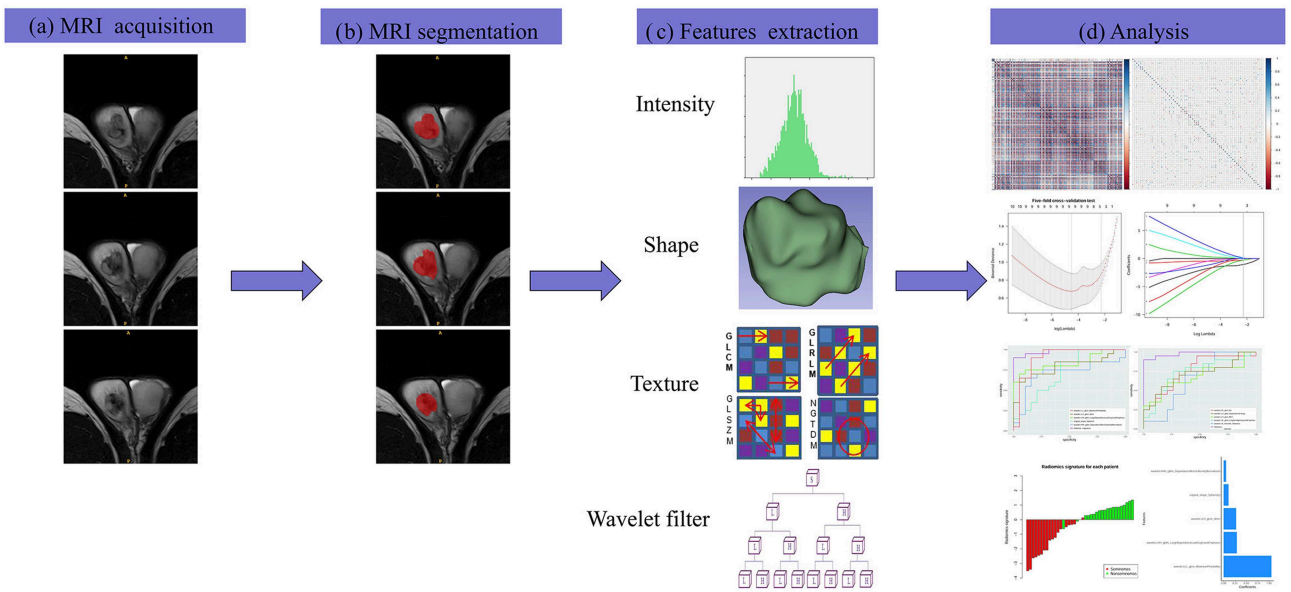


FIGURE 2 | The framework for the radiomics workflow. **(a)** All patients were scanned with preoperative MRI. **(b)** Tumors were delineated by stacking regions of interest slice-by-slice on the transverse T2WI. **(c)** Radiomics features were extracted from the T2WI in a high-throughput manner. **(d)** Data analysis for the features selection and a radiomics signature construction.

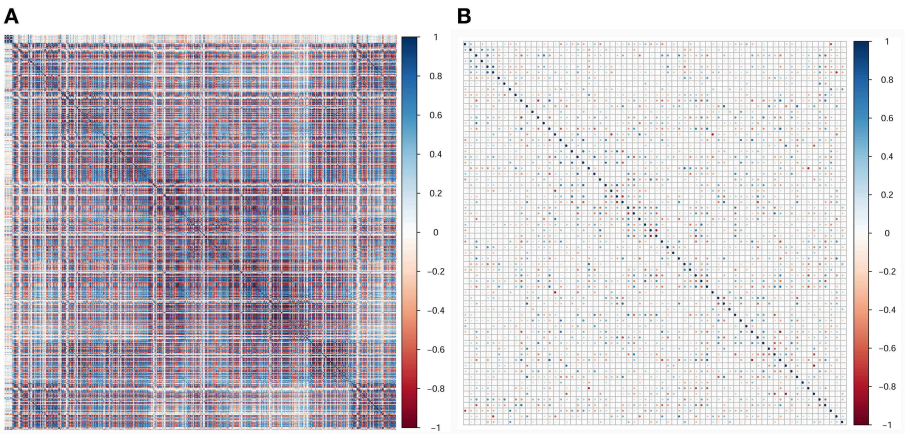


FIGURE 3 | Correlation matrix heatmaps of the features before **(A)** and after **(B)** correlation filtering. Before correlation filtering, a mass of redundant features with high correlation coefficients existed.

to perform the LASSO logistic regression model, and the "pROC" package was used to construct the ROC curve.

RESULTS

In the current study, 851 radiomics features were extracted from the T2WI of each patient. Seven hundred eighty features with an ICC $\geq$ 0.8 were further selected. Two hundred twenty-seven non-significant features were first eliminated using univariate analysis. After removing the redundant features using a Spearman's correlation coefficient threshold value of 0.8, a total of 67 features with low correlation remained. The correlation matrix heatmaps of the features before and after correlation filtering are shown in **Figure 3**. The features were ranked according to their mRMR scores. The top 10 features were selected using the mRMR algorithm (**Table 1**). Through the 5-fold cross-validation of the LASSO algorithm, five features with non-zero coefficients were included to construct the radiomics signature. The feature selection process using the LASSO algorithm is shown in **Figure 4**. The calculation formula to construct the radiomics signature is shown in **Table 2**. The contribution of the five features to the radiomics signature is shown in **Figure 5A**. The radiomics signature of each patient is shown in **Figure 5B**.

The radiomics signature was significantly different between seminomas and nonseminomas ($p < 0.01$). The ROC curves of the radiomics signature and the top 10 features selected from mRMR for discriminating between seminomas and nonseminomas are shown in **Figure 6** and **Table 3**. The AUC, sensitivity, and specificity of the radiomics signature were 0.979 (95% CI: 0.873–1.000), 90.00 (95% CI: 68.3–98.8), and 100.00 (95% CI: 82.4–100.0), respectively. The AUC of the radiomics signature was relativity higher than the AUCs of the top 10 features selected from mRMR.

TABLE 1 | The top 10 features selected by mRMR.

Features	mRMR scores	Groups
wavelet.LLL_glcm_MaximumProbability	0.31769474	Wavelet feature
wavelet.LLH_glcm_Idmn	0.08877716	Wavelet feature
wavelet.LHH_gldm_LargeDependenceLow GrayLevelEmphasis	0.07167124	Wavelet feature
original_shape_Sphericity	0.07024193	Shape feature
wavelet.HHH_gldm_DependenceNon-UniformityNormalized	0.07355068	Wavelet feature
wavelet.LHL_glcm_Idn	0.04066711	Wavelet feature
wavelet.LLH_gldm_DependenceEntropy	0.04644461	Wavelet feature
wavelet.LLH_glcm_MCC	0.02630265	Wavelet feature
wavelet.LHL_glrlm_LongRunHighGray LevelEmphasis	0.02301324	Wavelet feature
wavelet.LHL_firstorder_Skewness	0.02354773	Wavelet feature

DISCUSSION

In this study, an MRI-based radiomics signature was established to preoperatively discriminate between seminomas and

TABLE 2 | Calculation formula for the radiomics signature.

Variables	Coefficients
Intercept	−0.04258474
wavelet.LLL_glcm_MaximumProbability	−1.05440198
wavelet.LLH_glcm_Idmn	−0.27559477
wavelet.LHH_gldm_LargeDependenceLowGrayLevelEmphasis	−0.29108858
original_shape_Sphericity	0.10820225
wavelet.HHH_gldm_DependenceNon-UniformityNormalized	0.05352220

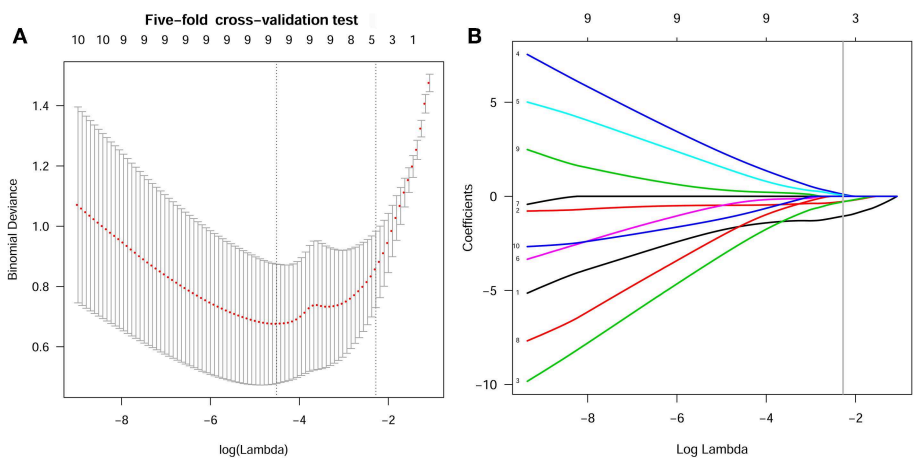


FIGURE 4 | Features selection using the LASSO algorithm. **(A)** Selection of the tuning parameter (Lambda) in the LASSO model using 5-fold cross-validation. Binomial deviances from the LASSO regression cross-validation model were plotted as a function of log(Lambda). The dotted vertical line at the right was drawn at the optimal value based on the minimum criteria and the 1-standard error rule (the 1-SE criteria). An optimal Lambda value of 0.102 with log(Lambda) = −2.280 and 5 non-zero coefficients were selected. **(B)** LASSO coefficient profiles of the 10 texture features. A vertical line was drawn at the optimal value selected using the 5-fold cross-validation process in **(A)**. The 5 features with non-zero coefficients were included to construct the radiomics signature.

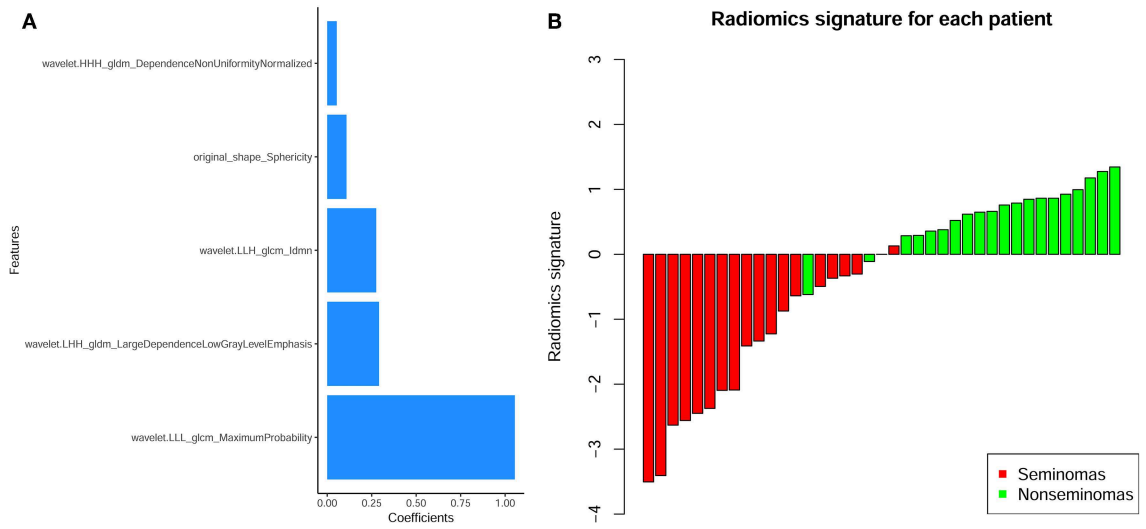


FIGURE 5 | (A) The contribution of the features to the radiomics signature. The histogram shows the contribution of the five features with non-zero coefficients to the radiomics signature. The features that contributed to the radiomics signature are plotted on the y-axis, and their coefficients in the LASSO Cox analysis are plotted on the x-axis. **(B)** Bar charts of the radiomics signature for each patient. The red bars indicate the radiomics signature of seminomas, while the light green bars indicate the radiomics signature of non-seminomas.

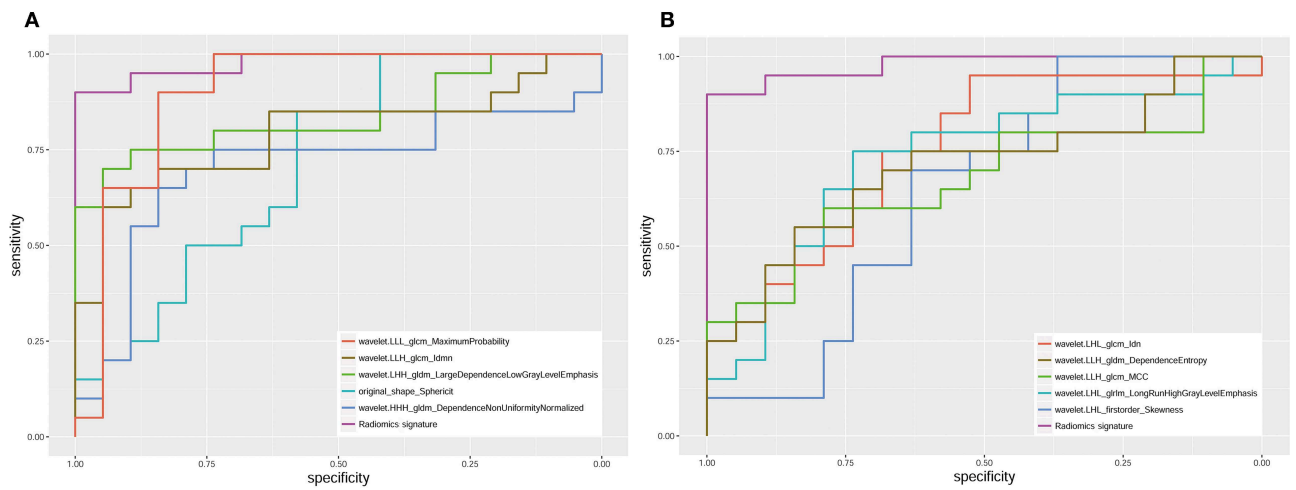


FIGURE 6 | ROC analysis of the radiomics signature and 10 features [**(A)** the top 5; **(B)** the bottom five)] selected from mRMR. The AUC of the radiomics signature was 0.979 (95% CI: 0.873–1.000).

nonseminomas. Our results showed that the radiomics signature could provide an excellent diagnostic performance (AUC = 0.979) by employing a large number of quantitative imaging features (851 features were extracted).

Non-invasively discriminating between seminomas and nonseminomas is of great significance. MRI has been proposed as a valuable supplemental imaging technique for characterizing testicular tumors (4, 9, 11). Tsili AC et al. enrolled 21 patients (10 seminomas and 11 nonseminomas) to investigate the value of MRI for differentiating seminomas from nonseminomas (9). Their results showed that the MRI findings led to a correct histologic diagnosis in 19 (91%) of 21 cases and the researchers concluded that tumor heterogeneity on MRI is indicative of nonseminomas. Another study including 15 seminomas and 11 nonseminomas showed that the mean apparent diffusion coefficient (ADC) values of seminomas were significantly lower than those of nonseminomas, while no significant differences were observed in DCE between seminomas and nonseminomas (4). Min et al. included 14 seminomas and 10 nonseminomas to assess the value of whole-tumor ADC histogram parameters for discriminating between seminomas and nonseminomas (11). Their results showed that the 10th percentile ADC value yielded the highest diagnostic performance. Although some positive results for distinguishing seminomas from nonseminomas have

TABLE 3 | ROC analysis of the features selected from mRMR.

Features	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Radiomics signature	0.979 (0.873–1.000)	90.00 (68.3–98.8)	100.00 (82.4–100.0)
wavelet.LLL_glcm_MaximumProbability	0.903 (0.764–0.974)	90.00 (68.3–98.8)	84.21 (60.4–96.6)
wavelet.LLH_glcm_Idmn	0.792 (0.632–0.905)	60.00 (36.1–80.9)	94.74 (74.0–99.9)
wavelet.LHH_gldm_LargeDependenceLowGrayLevelEmphasis	0.839 (0.687–0.937)	70.00 (45.7–88.1)	94.74 (74.0–99.9)
original_shape_Sphericity	0.718 (0.552–0.850)	85.00 (62.1–96.8)	57.89 (33.5–79.7)
wavelet.HHH_gldm_DependenceNonUniformityNormalized	0.703 (0.535–0.838)	65.00 (40.8–84.6)	84.21 (60.4–96.6)
wavelet.LHL_glcm_Idn	0.758 (0.594–0.880)	95.00 (75.1–99.9)	52.63 (28.9–75.6)
wavelet.LLH_gldm_DependenceEntropy	0.711 (0.543–0.844)	55.00 (31.5–76.9)	84.21 (60.4–96.6)
wavelet.LLH_glcm_MCC	0.679 (0.510–0.819)	55.00 (31.5–76.9)	84.21 (60.4–96.6)
wavelet.LHL_glrlm_LongRunHighGrayLevelEmphasis	0.737 (0.571–0.865)	75.00 (50.9–91.3)	73.68 (48.8–90.9)
wavelet.LHL_firstorder_Skewness	0.647 (0.478–0.793)	100.00 (83.2–100.0)	36.84 (16.3–61.6)

been reported, most previous studies used only some qualitative features or limited quantitative features, which may not fully explore the potential information of MRI, and no established prediction model has been built. In contrast to the above studies, in our study, a large number of quantitative radiomics features were extracted from the images and the most useful features were selected to construct a radiomics signature. Moreover, the sample size included in our study was relatively larger than that in previous studies.

Medical imaging provides valuable information for the diagnosis and evaluation of diseases. The conventional methods only use some qualitative features observable by the naked eyes or basic quantitative features, which cannot fully mine potential information from the images. Radiomics may help find potentially valuable information through the high-throughput extraction of quantitative features (14, 15). The newly proposed radiomics method has been successfully applied to various diseases (17, 18, 27–30). In a recent study, Lewin et al. applied radiomics to predict the pathology of postchemotherapy retroperitoneal nodal masses in germ cell tumors (27). Their results showed that the discriminative accuracy, sensitivity, and specificity of radiomics to identify GCT/teratoma vs. fibrosis was 72, 56.2, and 81.9%, respectively. When combined with clinical variables, the accuracy improved to 88%. In another study, Dong et al. suggested that a CT-based radiomic nomogram had excellent predictive ability for occult peritoneal metastasis in advanced gastric cancer patients (18). In our study, we used radiomics analysis to extract 851 features from T2WI and constructed a radiomics signature that includes features with excellent stability and reproducibility. Our results showed that the radiomics signature provides excellent efficiency for discriminating seminomas from nonseminomas. The AUC, sensitivity, and specificity of the radiomics signature were 0.979, 90.00, and 100.00, respectively. The AUC of the radiomics signature was 7.6–33.2% higher than the AUCs of the top 10 features selected from mRMR.

In this study, we only included T2WI for analysis, because T2WI is an essential component of testicular MRI with high contrast and spatial resolution. Previous studies have reported that seminomas and nonseminomas have different characteristics on T2WI (9). The presence of a relatively homogeneous testicular mass with low signal intensity on T2WI is considered indicative of seminomas. On the other hand, tumor heterogeneity is the most valuable finding in the characterization of nonseminomas. Although some studies have demonstrated the value of DWI and DCE in the characterization of testicular tumors (4, 11), these sequences have their limitations. The geometric distortion, susceptibility, and signal intensity dropout of DWI on tissue-air boundaries, such as the prostate, scrotum, and thyroid gland, are remarkable. Moreover, the DWI sequence usually has a low spatial resolution. These factors will limit the application and efficiency of DWI in characterizing testicular tumors. In recent years, some new techniques have been applied to DWI sequence to reduce geometric distortion and susceptibility artifacts, as well as to improve image resolution. However, few of these techniques had been used in testes; we will explore the value of new DWI techniques in testes in future studies (31–33). DCE-MRI usually requires the injection of a gadolinium-based contrast agent, which may increase the patient's risk for nephrogenic systemic fibrosis. Considering the above reasons and the limited sample size, we did not include DWI and DCE in the analysis.

There are some limitations in this study. First, our sample size was small. Although the number of patients included was higher than that of most previous studies, the sample size was still relatively small due to the low morbidity of testicular tumors. Further large-scale and multicenter studies are therefore warranted to obtain high-level evidence for clinical application. Second, rather than an independent validation cohort, internal validation was used in the current study, because there is insufficient data available to create an independent training cohort and a validation cohort. In this case, a fair way to accurately estimate the diagnostic performance of the radiomics signature is to use cross-validation (34). Third, the MRI sequences employed similar parameters but slightly varied slice numbers and thicknesses to cover some large lesions. To this end, we resampled the images before feature extraction to decrease the variability of the radiomics features extracted from the MRI sequences (35).

In conclusion, in the present study, we established a radiomics signature based on the features extracted from T2WI

to characterize TGCTs. The radiomics signature provides a non-invasive and quantitative method to differentiate between seminomas from nonseminomas. Further studies are warranted to validate our initial results.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology institutional review board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LW, XM, and PZ conception and design. PZ, ZF, WC, HY, CF, and WL acquisition of data. PZ and XM data processing. PZ, ZF, and XM analysis and interpretation of data. LW and XM supervision. PZ and ZF drafting the article. LW, XM, WC, HY, CF, and WL revising article critically for important intellectual content. All authors final approval of the manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (Grant numbers: 81671656, 81801668).

REFERENCES

1. Cheng L, Albers P, Berney DM, Feldman DR, Daugaard G, Gilligan T, et al. Testicular cancer. *Nat Rev Dis Primers*. (2018) 4:29. doi: 10.1038/s41572-018-0029-0
2. Batool A, Karimi N, Wu XN, Chen SR, Liu YX. Testicular germ cell tumor: a comprehensive review. *Cell Mol Life Sci*. (2019) 76:1713–27. doi: 10.1007/s00018-019-03022-7
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2018) 68:394–424. doi: 10.3322/caac.21492
4. Tsili AC, Sylakos A, Ntorkou A, Stavrou S, Astrakas LG, Sofikitis N, et al. Apparent diffusion coefficient values and dynamic contrast enhancement patterns in differentiating seminomas from non-seminomatous testicular neoplasms. *Eur J Radiol*. (2015) 84:1219–26. doi: 10.1016/j.ejrad.2015.04.004
5. Albers P, Albrecht W, Algaba F, Bokemeyer C, Cohn-Cedermark G, Fizazi K, et al. Guidelines on testicular cancer: 2015 update. *Eur Urol*. (2015) 68:1054–68. doi: 10.1016/j.eururo.2015.07.044
6. Boujelbene N, Cosinschi A, Boujelbene N, Khanfir K, Bhagwati S, Herrmann E, et al. Pure seminoma: a review and update. *Radiat Oncol*. (2011) 6:90. doi: 10.1186/1748-717X-6-90
7. Oldenburg J, Fossa SD, Nuver J, Heidenreich A, Schmoll HJ, Bokemeyer C, et al. Testicular seminoma and non-seminoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. (2013) 24 (Suppl. 6):vi125–32. doi: 10.1093/annonc/mdt304
8. Tsili AC, Sofikitis N, Stiliara E, Argyropoulou MI. MRI of testicular malignancies. *Abdom Radiol*. (2019) 44:1070–82. doi: 10.1007/s00261-018-1816-5
9. Tsili AC, Tsampoulas C, Giannakopoulos X, Stefanou D, Alamanos Y, Sofikitis N, et al. MRI in the histologic characterization of testicular neoplasms. *AJR Am J Roentgenol*. (2007) 189:W331–7. doi: 10.2214/AJR.07.2267
10. Dikici AS, Er ME, Alis D, Samanci C, Ustabasioglu FE, Demirdag C, et al. Is there any difference between seminomas and nonseminomatous germ cell tumors on shear wave elastography? A preliminary study. *J Ultrasound Med*. (2016) 35:2575–80. doi: 10.7863/ultra.15.12067
11. Min X, Feng Z, Wang L, Cai J, Yan X, Li B, et al. Characterization of testicular germ cell tumors: whole-lesion histogram analysis of the apparent diffusion coefficient at 3T. *Eur J Radiol*. (2018) 98:25–31. doi: 10.1016/j.ejrad.2017.10.030
12. Dogra VS, Gottlieb RH, Oka M, Rubens DJ. Sonography of the scrotum. *Radiology*. (2003) 227:18–36. doi: 10.1148/radiol.2271001744
13. Tsili AC, Bertolotto M, Turgut AT, Dogra V, Freeman S, Rocher L, et al. MRI of the scrotum: recommendations of the ESUR scrotal and penile imaging working group. *Eur Radiol*. (2018) 28:31–43. doi: 10.1007/s00330-017-4944-3
14. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
15. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
16. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169
17. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by non-invasive imaging using a quantitative radiomics approach. *Nat Commun*. (2014) 5:4006. doi: 10.1038/ncomms5644
18. Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol*. (2019) 30:431–8. doi: 10.1093/annonc/mdz001
19. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
20. Park HJ, Lee SS, Park B, Yun J, Sung YS, Shim WH, et al. Radiomics analysis of gadoxetic acid-enhanced MRI for staging liver fibrosis. *Radiology*. (2019) 292:269. doi: 10.1148/radiol.2019194012
21. Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang YD. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol*. (2017) 27:4082–90. doi: 10.1007/s00330-017-4800-5
22. Motzer RJ, Agarwal N, Beard C, Bhayani S, Bolger GB, Buyyounouski MK, et al. Testicular cancer. *J Natl Compr Canc Netw*. (2012) 10:502–35. doi: 10.6004/jnccn.2012.0050
23. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77:e104–e7. doi: 10.1158/0008-5472.CAN-17-0339
24. Huynh E, Coroller TP, Narayan V, Agrawal V, Romano J, Franco I, et al. Associations of radiomic data extracted from static and respiratory-gated CT scans with disease recurrence in lung cancer patients treated with SBRT. *PLoS ONE*. (2017) 12:e0169172. doi: 10.1371/journal.pone.0169172
25. De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. mRMRe: an R package for parallelized mRMR ensemble feature selection. *Bioinformatics*. (2013) 29:2365–8. doi: 10.1093/bioinformatics/btt383
26. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in

multivariable model building. *Stat Med*. (2007) 26:5512–28. doi: 10.1002/sim.3148
27. Lewin J, Dufort P, Halankar J, O'Malley M, Jewett MAS, Hamilton RJ, et al. Applying radiomics to predict pathology of postchemotherapy retroperitoneal nodal masses in germ cell tumors. *JCO Clin Cancer Inform*. (2018) 2:1–12. doi: 10.1200/CCI.18.00004
28. Song J, Shi J, Dong D, Fang M, Zhong W, Wang K, et al. A new approach to predict progression-free survival in stage IV EGFR-mutant NSCLC patients with EGFR-TKI therapy. *Clin Cancer Res.* (2018) 24:3583–92. doi: 10.1158/1078-0432.CCR-17-2507
29. Peng H, Dong D, Fang MJ, Li L, Tang LL, Chen L, et al. Prognostic value of deep learning PET/CT-based radiomics: potential role for future individual induction chemotherapy in advanced nasopharyngeal carcinoma. *Clin Cancer Res*. (2019) 25:4271–9. doi: 10.1158/1078-0432.CCR-18-3065
30. Min X, Li M, Dong D, Feng Z, Zhang P, Ke Z, et al. Multi-parametric MRI-based radiomics signature for discriminating between clinically significant and insignificant prostate cancer: cross-validation of a machine learning method. *Eur J Radiol*. (2019) 115:16–21. doi: 10.1016/j.ejrad.2019.03.010
31. Kim YJ, Kim SH, Kang BJ, Park CS, Kim HS, Son YH, et al. Readout-segmented echo-planar imaging in diffusion-weighted mr imaging in breast cancer: comparison with single-shot echo-planar imaging in image quality. *Korean J Radiol.* (2014) 15:403–10. doi: 10.3348/kjr.2014.15.4.403
32. Kim LH, Lee EH, Galvez M, Aksoy M, Skare S, O'Halloran R, et al. Reduced field of view echo-planar imaging diffusion tensor MRI for pediatric spinal tumors. *J Neurosurg Spine*. (2019) 5:1–9. doi: 10.3171/2019.4.SPINE19178
33. Corrias G, Raeside MC, Agostini A, Huicochea-Castellanos S, Aramburu-Nunez D, Paudyal R, et al. Pilot study of rapid MR pancreas screening for patients with BRCA mutation. *Eur Radiol*. (2019) 29:3976–85. doi: 10.1007/s00330-018-5975-0
34. LeCarpentier GL, Roubidoux MA, Fowlkes JB, Krucker JF, Hunt KA, Paramagul C, et al. Suspicious breast lesions: assessment of 3D Doppler US indexes for classification in a test population and fourfold cross-validation scheme. *Radiology*. (2008) 249:463–70. doi: 10.1148/radiol.2492060888
35. Larue R, van Timmeren JE, de Jong EEC, Feliciani G, Leijenaar RTH, Schreurs WMJ, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncol*. (2017) 56:1544–53. doi: 10.1080/0284186X.2017.1351624

Conflict of Interest: WL was employed by company Julei Technology.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

REVIEW
published: 28 November 2019
doi: 10.3389/fonc.2019.01296

Study Progress of Radiomics With Machine Learning for Precision Medicine in Bladder Cancer Management

Lingling Ge[1†], Yuntian Chen[2†], Chunyi Yan[1], Pan Zhao[3], Peng Zhang[3], Runa A[4] and Jiaming Liu[3]*

[1] West China Hospital, Sichuan University, Chengdu, China, [2] Radiological Department, West China Hospital, Sichuan University, Chengdu, China, [3] Department of Urology, Institute of Urology, West China Hospital, Sichuan University, Chengdu, China, [4] Department of Obstetrics and Gynecology, West China Second Hospital, Sichuan University, Chengdu, China

Bladder cancer is a fatal cancer that happens in the genitourinary tract with quite high morbidity and mortality annually. The high level of recurrence rate ranging from 50 to 80% makes bladder cancer one of the most challenging and costly diseases to manage. Faced with various problems in existing methods, a recently emerging concept for the measurement of imaging biomarkers and extraction of quantitative features called "radiomics" shows great potential in the application of detection, grading, and follow-up management of bladder cancer. Furthermore, machine-learning (ML) algorithms on the basis of "big data" are fueling the powers of radiomics for bladder cancer monitoring in the era of precision medicine. Currently, the usefulness of the novel combination of radiomics and ML has been demonstrated by a large number of successful cases. It possesses outstanding strengths including non-invasiveness, low cost, and high efficiency, which may serve as a revolution to tumor assessment and emancipate workforce. However, for the extensive clinical application in the future, more efforts should be made to break down the limitations caused by technology deficiencies, inherent problems during the process of radiomic analysis, as well as the quality of present studies.

Keywords: radiomics, machine learning, bladder cancer, full-cycle management, precision medicine

OPEN ACCESS

Edited by:
Lei Deng,
Jacobi Medical Center, United States

Reviewed by:
Di Dong,
Institute of Automation (CAS), China
Lubomir Hadjiiski,
University of Michigan, United States

****Correspondence:***
Jiaming Liu
JM3099@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *19 August 2019*
Accepted: *08 November 2019*
Published: *28 November 2019*

Citation:
Ge L, Chen Y, Yan C, Zhao P, Zhang P, A R and Liu J (2019) Study Progress of Radiomics With Machine Learning for Precision Medicine in Bladder Cancer Management. Front. Oncol. 9:1296. doi: 10.3389/fonc.2019.01296

INTRODUCTION

Bladder cancer ranks ninth of the most common malignancies and the 13th most common predisposing cause of cancer-related mortality all over the world, with over 357,000 new cases and over 130,000 deaths annually (1). Bladder cancer is more likely to develop in patients over 65 years old, which has a high recurrence rate ranging from 50 to 80% (2).

Clinical decision and follow-up management of bladder cancer predominantly depend on the presence or absence of muscle invasion and accurate grade of malignancy and also take specific pathological types into consideration (3, 4).

However, all aspects in the management of bladder cancer including tumor staging, diagnosis, treatment, and prognostic evaluation have still been limited by various factors. The gold standard nowadays for bladder cancer detection is telescopic checking of the bladder (cystoscopy) (5). Considering the high recurrence rate, cystoscopy examinations are required to be performed every 3–6 months to monitor bladder cancer patients for recurrence or progression to a more advanced

stage. However, the expensive and invasive characteristics restrict the frequent use of cystoscopy and further cause significant economic and psychological pressure to patients. Furthermore, cystoscopy has still quite limited accuracy for the detection of tumors in low grade, with the sensitivity of 61% (6), nor the muscle-invasive depth. Another universally applied approach in condition detection is biopsy but it can be restricted by an inability to sample every part of the tumor at any point in time. Moreover, heterogeneous disease spectrums as bladder cancer processes, it can always confound correct classification and staging (5), and then influences the choice of treatment plans and finally makes the risk of undertreatment or overtreatment increase. To simplify the detection process of bladder cancer, the concept of urinary biomarkers has been put forward recently. However, no studies have proposed molecular markers with sufficient sensitivity and specificity to replace cystoscopy (1). Furthermore, this method is unable to determine the extent of surrounding tissue invasion and metastasis, which blocks the way of non-invasive bladder cancer detection.

Apart from the difficulties in accurate tumor detection and clinical grading, the prediction of treatment efficacy is also a major obstacle in the management of bladder cancer. It has been widely accepted that patients with nonmuscle-invasive bladder cancer (NMIBC, stage $\leq$T1) are mostly at early stage and are advised to be treated with TURBT followed by treatment applying Bacillus Calmette-Guérin (BCG) (7), whereas muscle-invasive bladder cancer (MIBC, stage $\geq$T2) patients usually have a poorer prognosis and the treatment plan for these patients is supposed to be radical cystectomy (RC) (8). Despite appropriate cancer control in local lesions, over 50% of patients who have undergone RC meet the disturbance of tumor metastasis in no more than 2 years after cystectomy and thus fail to survive (9). Neoadjuvant chemotherapy ahead of cystectomy has been demonstrated to reducing the odds of developing extravesical lesions when compared to taking RC alone, afterward improves the overall survival (OS) of bladder cancer patients (10, 11). However, there is still short of a reliable method to predict the posttreatment response of a specific individual to whether BCG or neoadjuvant chemotherapy currently. Thus, some patients who receive inappropriate therapies tend to suffer from adverse reactions. Worse, these patients are likely to miss the best time to make an adjustment on the strategies of therapy, consequently pose damage to their physical condition and increase the difficulty of cancer management.

Confronting the above problems and limitations, a novel concept of radiomics has emerged for solving the issues of the generalization of precision medicine and how it can be applied in the field of bladder cancer monitoring. It is a high-throughout quantitative feature extraction method to mine the information contained in the multimodality medical images including computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), and ultrasonography (US) (12), then comprehensively analyze these massive images to extract phenotypic features (also known as radiomics biomarkers) and explore the associations between patients' prognosis and these extracted features and improve the decision-making process. ML algorithms on the basis of "big data" are fueling the powers of radiomics in three main tasks related to bladder cancer imaging: initial detection of the existence and localization of volume; pretreatment characterization including the diagnosis, grading, and staging of tumor; posttreatment monitoring by predicting prognosis or factors irrelevant to treatment plans, such as OS, recurrence, and pathological subtypes (13, 14). Therefore, radiomics methods in combination with an optimal ML method may potentially extend the practical use of precision medicine approaches in radiotherapy by providing a non-invasive, high-efficiency, but low-cost way to predict clinical outcomes (15).

In this paper, we review the present studies in association with our topic and discuss the promising usages and hidden challenges of this novel method adapting acute imaging analysis, combining radiomics with ML in the precise management of bladder cancer. This review paper considers and discusses the issues as follows:

1. The concept of radiomics and its significance
2. Workflow of radiomics
3. Clinical applications of the combination of radiomics with ML in bladder cancer management
4. Challenges and future directions.

THE CONCEPT OF RADIOMICS AND ITS SIGNIFICANCE

The meaning of "precision medicine" indicates that the reasonable strategies of treatment are singled out according to the characteristics of different subtypes. It has substantially changed the treatment strategies in the recent 10 years. To make a precise treatment for an individual, accurate detection, characterization, and monitoring after treatment are very important. Unfortunately, the current tumor assessment is far from our expectation because of variable technology deficiency. One important problem is that radiologists usually use subjective, qualitative features to make tumor assessment, which make the results less reproducible and more unstable. Besides, with the rapid development of gene therapy and immunotherapy, gene expression signature and immune phenotype are also essential parts for a comprehensive tumor assessment. Current evaluation for gene expression and immune phenotype is most based on the biopsy, which is invasive and expensive, let alone the result is confused because of intratumoral heterogeneity. Thus, the demand for a non-invasive, cheap, and stable method to assess and monitor tumor has never been greater.

The computational medical imaging, also known as radiomics, was first invented by Lambin in 2012 (16, 17). It was based on the underlying hypothesis that medical imaging contains much more information than we have already utilized, even including cellular and molecular information of target tissue (18). The aim of radiomics is to analyze and translate medical images into quantitative data and provide an image-based biomarker to aid clinical decisions. Compared with biopsy, radiomics biomarker is invasive, reproducible, and has the ability to make an evaluation of tumors' microenvironment, spatial heterogeneity, and longitudinal assessment for disease progression. In recent years, several studies have presented the

potential usage of radiomics in the development of precision medicine. Among all, many pieces of research demonstrated the immense application value of radiomics in combination with ML algorithms to overcome the drawbacks of precision medicine in the diagnosis and treatment of non-small-cell lung cancer (19). Since the breakthrough in the intelligent management of lung cancer, researches on the application of this new technology to other cancers have been carried out one after another with great advances among them. The characteristics of bladder cancer itself and the need for enhanced imaging analysis technology make it one of the major research hot spots.

WORKFLOW OF RADIOMICS

Radiomics is a multidisciplinary-based technology. There are four main steps to complete a radiomics program (**Figure 1**):

(1) Image acquisition and preprocessing
(2) Volumes of interest (VOIs) segmentation
(3) Feature extraction and quantization
(4) Model building.

Image Acquisition

Image acquisition is the first and an important step in radiomics.

There are two common formats of the image data recorded, including Picture Archiving and Communication System (PACS) and the Digital Imaging and Communication in Medicine (DICOM), and they are used in most of the medical institutions, which provide a great convenience for the radiology study.

The accuracy and reproducibility of the final radiomics model lie on the quality of image acquisition. However, there is no guideline nor consensus on image acquisition. As a result, the acquisition strategies in different research teams can be distinctive, which can cause heterogeneity among separate studies. Meanwhile, it is difficult for researchers to take labeling, annotation, segmentation, and quality assurance seriously. Because these processes require well training, wasting both time and money.

Various imaging modalities such as modern CT, MRI, PET, and US scanners allow for acquisition and image reconstruction in wide variations (20). The radiological images applied for radiomic analysis are obtained from different hospitals or institutions using divided parameters and protocols. Thus, they are supposed to be preprocessed to ensure consistency and comparability.

VOIs Segmentation

The region for image data capture is defined as "The Volume of Interest" (VOI). VOIs segmentation is the core step of radiomics study because it determines which volume is analyzed within a medical image.

The ideal VOI includes the complete information for the target lesions, nothing more nor less. Unfortunately, usually, it is hard for the radiologists to make it because many tumors have indistinct borders (21). Besides, the microenvironment around the lesions also provides useful information of the lesions, but there is no guideline in VOIs segmentation for how much microenvironment should be put in radiomics model. For instance, in a recent radiomics study, the research team use 2-mm peripheral ring on each side of the lesions to involve microenvironment of the lesions while another team use 1-mm peripheral ring (22).

VOIs can be created manually, automatically, or semiautomatically. In the past 5 years, most current radiomics study created VOIs manually. However, it is time-consuming and laborious when utilizing big data in radiomics study. Thus, many pieces of research try to create VOI automatically. There are lots of algorithms on VOI creation. A common segmentation algorithm is the "seed method." The radiologist will place some seeds in VOI, and the computer will create VOI automatically (23). Unfortunately, it only works well when the lesion is uniform. There are many other methods used in radiomics study, such as Graph-Cut Methods (24), Level-Set Methods (25), and Active Contour Algorithms (26). In summary, all algorithms have their own deficiencies and need manual correction. Therefore, there is a great demand for algorithms with maximum automation, minimal human intervention, high time efficiency, and repeatability.

Feature Extraction

Feature extraction is the next step after a VOI is reasonably segmented, which is the essence in the workflow of radiomics. It mainly performs the extraction of high-throughput data of quantitative imaging features to identify VOIs and the selection of useful information to aid in the discrimination of normal and abnormal images.

The quantitative data usually can be classified into four types:

(1) **Shape characteristics:** Description of the shape and geometric features of target lesions. For example, the volume of VOI, maximum diameter along with different orthogonal directions, maximum surface area, tumor compactness and sphericity.
(2) **First-order statistical characteristics:** Description of the distribution of individual but do not describe its spatial arrangement, including the mean, median, maximum and minimum of voxel intensity, skewness (asymmetry), kurtosis (flatness), uniformity, and randomness (entropy).
(3) **Second-order statistical characteristics:** Also known as texture features. It is calculated from a statistical correlation between adjacent voxels, usually by Gray-Level Co-Occurrence Matrix (GLCM) (27). Texture features provide heterogeneity information among the lesions.
(4) **High-order statistical characteristics:** Quantitative data from the filter or mathematical transformation on images, including fragment analysis, Minkowski function, wavelet transform, and Laplace transform of gaussian filtered image. The aim of filer and mathematical transformation is to identify repetitive or non-repetitive patterns, suppress noise, or highlight details.

Model Building

To conclude, the building of radiomics model involves three main aspects, including the selection of radiomic features, choice

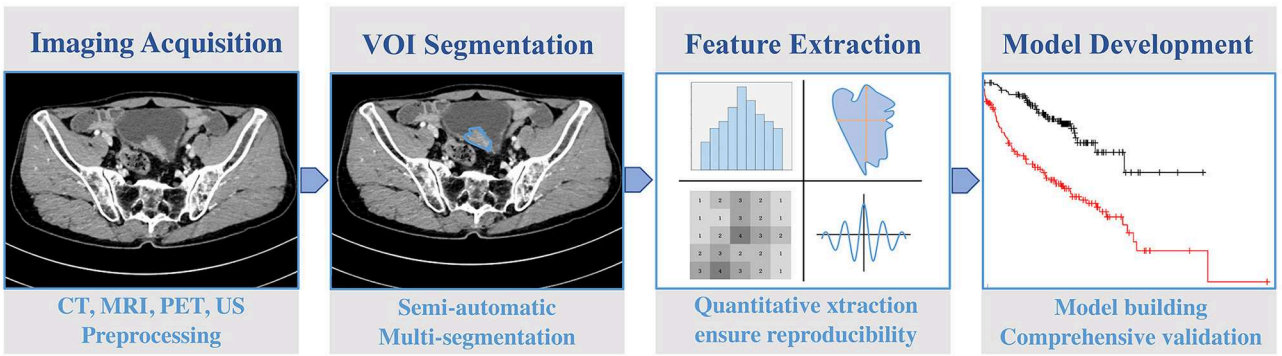


FIGURE 1 | A typical workflow of radiomics in bladder cancer.

of ML methods, and final validation of the developed model. Image analysis software can provide many features, so it is vital to select essential features for further study. At the same time, model selection is as important as feature selection. Selecting a suitable model can help obtain a reliable and stable result, which is very important for clinical decision making. Complete radiomics analysis should include validation, and both internal validation and external validation are indispensable.

Feature Selection

There are two common processes for determining radiomic parameters. One is to make a preliminary analysis of captured features and select most repeatable and reproducible ones (28). Another method is based on features' mathematical definitions. It makes *a priori* selection of features on those definitions and selects targeted parameters (29). In the formal process of analysis, different kinds of image analysis software may output a variety of features ranging from hundreds to thousands (30, 31). The inclusion of all features without selection in the development of radiomics model could cause the result in overfitting inevitably, since some of them might have a high degree of correlation (32). Hence, the inclusion of appropriate features, which are strongly linked to the aiming task but not redundant, is highlighted for improved value for specific clinical applications.

Modeling Methodology

ML is commonly used in radiomics model development, which can be hypothetically defined as a branch of artificial intelligence (AI) (33), which is actually an algorithm trained by inferences from data sets and then helps establish prediction models with high precision and efficiency on the basis of radiomic analysis. As a result, radiomics with ML may improve the clinician decision-making process as it is able to encompass many greater quantities of parameters than manual work and make these various parameters extracted in the workflow of radiomics into comprehensive utilization. With regard to the choice of appropriate radiomic methodology, the identification and application of optimal ML methods for radiomic applications are very crucial steps toward the achievement of clinical relevance. Thus, appropriate ML algorithms should be employed (34) to fuel the detective, diagnostic, and prognostic powers of radiomics in the field of bladder cancer.

Model Validation

Model validation is a dispensable step in model building, which serves as a useful tool to assess the performance and applicability of the developed radiomics model. To make sure that the model is effective for all of the targeted patients, not only the patients selected in the model building process, the internal and/or external validation should be tested. Typically, model performance is measured according to its discrimination, which can be expressed in the form of the receiver operating characteristic (ROC) curve or be calculated as the area under the ROC curve (AUC) in a quantitative way. ROC curve can easily display the ability of disease recognition at any threshold. When comparing two or more models, the ROC curves can draw each model in the same coordinate to identify advantages and disadvantages visually.

CLINICAL APPLICATIONS OF RADIOMICS WITH ML IN BLADDER CANCER MANAGEMENT

Accurate Cancer Staging and Grading

Accurate grading is of vital importance in the follow-up management of bladder cancer and serves as the starting point of radiomics applied in bladder cancer as shown in **Table 1**. Reading radiographic images produced by CT, MRI, PET, or US is essentially a matter of identifying complex patterns, in which computers can be trained to process efficiently, repeatedly, and rapidly. In the recent 5 years, ML techniques have shown latent capacity in accurate grading of bladder cancer. Confronted with the great significance of precise staging in the decision of appropriate treatment plan, Garapati et al. (35) developed a predictive model to serve as a classification tool for layering bladder cancer into two different grading categories and chose T2 as the critical point. They created a data set containing 84 bladder cancer lesions from 76 CT urography (CTU) cases and found that the morphological features, as well as texture features were helpful to stage the lesions of bladder cancer. In this study,

TABLE 1 | Present studies that combined radiomics and machine learning (ML) in bladder cancer.

References	Study type	Application	Cases number	Data modality	ML algorithm	Type of validation	Results
Garapati et al. (35)	Retrospective study	Bladder cancer staging	76	CTU	LDA, NN, SVM, RAF	Two-fold cross validation	Four types of classifier showed equal promise in bladder cancer staging
Zhang et al. (36)	Retrospective study	Bladder cancer grading	61	MRI	SVM-RFE	Single-center validation	The SVM classifier adapting the optimal feature subset performed best (AUC = 0.861; accuracy 82.9%; sensitivity 78.4%; specificity 87.1%)
Wang et al. (37)	Retrospective study	Bladder cancer grading	70	MRI	LASSO algorithm	Ten-fold cross validation	Joint-Model performed best (AUC = 0.9276)
Zheng et al. (38)	Retrospective study	Differentiation of NMIBC and MIBC	199	MRI	LASSO logistic regression algorithm	Single-center validation	The radiomic-clinical nomogram developed on the basis of three-dimensional features showed favorable usage (AUC 0.922)
Wang et al. (39)	Retrospective study	Prediction of mortality after radical cystectomy	117	Clinical data	BPN, RBFN, ELM, RELM, SVM, NB, and KNN	Ten-fold cross validation	The models with RELM and ELM achieved the highest sensitivity and specificity (over 0.8)
Xu et al. (40)	Retrospective study	Recurrence stratification of bladder cancer	71	MRI	SVM-RFE, LASSO algorithm	Five-fold cross validation	The radiomic clinical nomogram achieved more benefits than the radiomics or clinical model alone
Lin et al. (41)	Retrospective study	Prediction of progression-free interval	62	CECT	LASSO algorithm	Single-center validation	Radiomics risk model (AUC 0.956) and transcriptomics risk model (AUC 0.948) showed independent prognostic role to determine the progression
Cha et al. (42)	Retrospective study	Assessment of therapy response	62	CT	DL-CNN	Leave-one-case-out cross validation	DL-CNN has the potential to assist in the treatment response
Cha et al. (43)	Retrospective study	Assessment of treatment response	123	CT	DL-CNN	Single-center validation	The radiomics-based system is advisable to serve as a second option to assist in therapy evaluation
Chalkidou et al. (44)	Retrospective study	Evaluation of sensitivity to neoadjuvant chemotherapy	123	CT	DL-CNN	Single-center validation	The improvement of the physicians' performance was statistically significant (P <.05)

LDA, linear discriminant analysis; NN, neural network; DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient; RFE, recursive feature elimination; BPN, back-propagation neural network; RBFN, radial basis function; ELM, extreme learning machine; RELM, regularized ELM; NB, naive Bayes; KNN, k-nearest neighbor; DL-CNN, deep-learning convolution neural network; MIBC, muscle-invasive bladder cancer; NMIBC, non-muscle-invasive bladder cancer.

the data were divided into two subsets for further two-fold cross validation. As a result, all of the linear discriminant analysis (LDA), neural network (NN), support vector machine (SVM), and random forest (RAF) classifiers included in this study led to relatively unanimous results in their staging accuracy, which effectively demonstrated that ML method can be a promising way to reduce the inaccuracy rate of bladder cancer staging by up to 50% (45) and aid to the implementation in daily care. To achieve better grading of bladder cancer for the sake of appropriate clinical decision, a recent study proposed textural features from diffusion-weighted imaging (DWI) and apparent diffusion coefficient (ADC) maps to distinguish low-grade bladder cancers from high-grade ones and recommended an optimal feature subset selected by SVM with recursive feature elimination (SVM-RFE) for cancer grading using histogram and gray-level co-occurrence matrix (GLCM)-based radiomic features (36). Sixty-one patients with bladder cancer were included in this study to prove that the grading performance in bladder cancer was improved through the candidate or extraction of optimal features, with accuracy, sensitivity, specificity, and AUC achieving 82.9, 78.4, 87.1, and 0.861, respectively.

Recently, there have also been notable advances in MRI technology. Wang et al. (37) built and validated a multiparametric MRI-based radiomic analysis model for the preoperative grading of bladder cancer tumors. This study enrolled 70 bladder cancer patients and applied five radiomic models including T2-weighted imaging (T2WI), DWI, ADC, Max-out, and Joint models, then assessed by ROC curve analysis. By comparing AUC values, the performance in terms of the accuracy, sensitivity, and specificity of the Joint_model in the validation set was obviously superior to that of the other four single-modality models, achieving an AUC of 0.9233 according to the training cohort and 0.9276 in the validation one. As the first study considering pathological grading of bladder cancer applying radiomics, it showed encouraging feasibility for avoiding subjectivity and promote extended future usage in preoperative grade assessment of bladder cancer.

Tumor Classification and Prognosis Prediction

Undoubtedly, clear discrimination between NMIBC and MIBC is crucial for pretreatment decision, posttreatment prognosis, and lasting period consequent management of bladder cancer patients. Considering the large percentage of diagnostic errors caused by conventional cystoscopic examination (36), researches proposed a new radiomics scheme that combined histogram features, co-occurrence matrix (CM) features, and run-length matrix (RLM) features. Based on the novel developed model (4), they assessed the performance of tumor classification based on multiparametric MRI radiomic features for accurate differentiation between NMIBC and MIBC preoperatively in searching for management of bladder cancer patients and finally got a positive result with the AUC and Youden index improving to 0.8610 and 0.7192, respectively.

With regard to the reflection on the heterogeneity of cancer caused by analyzing radiomic features extracted from the whole tumor tissue, Zheng et al. (38) suggested a hypothesis that the radiomics features of the basal part can be used for determining the degree of muscle invasiveness more conclusively. To further validate their assumption, they first developed a radiomic-clinical nomogram incorporating the radiomic signature and extract three-dimensional features for subsequent analysis. The usefulness of this novel nomogram in discriminating NMIBC and MIBC was favored, with an AUC of 0.922 in the training set, and was also confirmed with an AUC of 0.876 in the validation set. By overthrowing existing views, this study came up with a new idea in the extraction of radiomic features and demonstrated the latent capacity of the radiomic-clinical nomogram to serve as an auxiliary tool for bladder cancer classification.

Realizing the high risk of metastasis and mortality of MIBC that is in need of immediate treatment to improve the living quality of patients while faced with the various kinds of ML models emerging, researchers have begun to seek for the urgently needed strategy. A confirmative research was conducted for selection in which seven models including radial basis function (RBFN), back-propagation neural network (BPN), extreme learning machine (ELM), regularized ELM (RELM), SVM, naive Bayes (NB) classifier, and k-nearest neighbor (KNN) were ever explored and compared to predict the 5-year mortality of 117 MIBC patients who had undergone RC (39). The eventual results indicated that the algorithm on the basis of RELM and ELM presented higher mean sensitivity and specificity (over 0.8) relatively in the prediction of mortality in MIBC patients after RC, with an ideally fast learning speed.

Regarding the importance of preoperative prediction of the risk of bladder cancer recurrence, Xu et al. (40) developed a radiomic nomogram for personalized prediction of the first 2 years (TFTY) risk in tumor recurrence. This study took the important baseline variations involving gender, age, cancer grading, MIS of the lesions, size, and the number of the tumor, as well as the recording of previous operations into consideration to ensure the accuracy of the developed model. This multiparametric model finally showed excellent performance in both the validation and training cohorts. The decisive curve exhibited the threshold of risk was more than 0.3; more benefits and higher accuracy were observed by applying the radiomic nomogram than using either the radiomics or clinical model separately.

All of these studies showed the great potential of radiomics aided by ML in the improvement of tumor classification and prognosis prediction of bladder cancer, overcoming the defects traditionally and obviously improved the living quality of bladder cancer patients. Great progression as radiomics made in the prognosis of bladder cancer, the correlations between the imaging features and genomic signatures have rarely been explored up to now in the field of bladder cancer management. Confronting the blank of related researches, Lin et al. (41) put forward a successful case that integrated radiomics and transcriptomics to predict the progression-free interval (PFI) of bladder cancer patients. In this study, both of the radiomics risk model (AUC 0.956) and transcriptomics risk model (AUC 0.948) showed independent

prognostic roles to determine the progression of the bladder tumor, which first provided a novel insight into the microscopic mechanisms of bladder cancer.

Evaluation of Individual Therapy Responses

Once a definite diagnosis is made and the tumor subtype is identified, early evaluation of therapeutic efficacy and response can aid in clinicians'decision on whether to discontinue chemotherapy at an optimal phase. In the research conducted by Cha et al. (42), the feasibility of a radiomics-based prognostic model using CT images obtained before or after treatment in distinguishing bladder cancers with chemotherapy responses or not was explored, which applied deep-learning convolution neural network (DL-CNN) for accurate bladder lesion segmentation. This study indicated that the computerized assessment on the basis of radiomics information extracted from CT images of pretreatment or posttreatment bladder cancer patients had the possibility to aid in the evaluation of therapy response, with the prediction accuracy estimated by AUCs improving by 0.03 in comparison with manual contours. Previously, one of the reasons for optimization has been demonstrated that the complicated tumor size change in response to treatment can be better reflected by the computerized three-dimensional (3-D) segmentation rather than traditional criteria (46). Recognizing the important role played by radiomics and ML in the evaluation of individual therapy response, another feasibility study adopted three radiomics predictive models in the assessment of chemotherapy response and finally reached an agreement among all these models and two expert radiologists' prediction. Instead of replacing artificial analysis, researchers are more supportive to consider the computer-aided system as a second option to assist in the evaluation process. However, the final decision still relies on the judgment of radiologists on the basis of advice given by radiomics-aided models (43).

Regarding the aggressiveness of MIBC, studies were conducted to assess whether a CT-based decision-support system could improve identification of patients who have complete or partial response to neoadjuvant chemotherapy (47). This study investigated 123 subjects, and the AUC was estimated. Compared to the accuracy of the assessment of doctors alone (AUC 0.74), the decision-support system (AUC 0.80) showed statistically significant improvement of therapeutic evaluation.

CHALLENGES AND FUTURE DIRECTIONS

Up to now, there has been a large number of studies to illustrate the rapid development of radiomics and ML algorithms, as well as the effectiveness of their combination in the full-cycle management of bladder cancer, but we also need to realize the limitations when applied in actual use comprehensively.

First, the defects in the aspect of inherent technologies should be recognized. Usefulness as quantitative features based on medical images show in the management of bladder cancer patients, their tendency of causing significant errors and overestimation has also been reported (44). It is prevalent to see that the number of radiological features examined is much larger than the number of patients included in retrospective studies, which may lead to bias in feature selection and even high false-positive result. Furthermore, results obtained from different studies might be hard to compare and evaluate because of the lack of standardized methods for analysis. Larue et al. (48) conducted an overview and put forward a thought-provoking conclusion that there still exist various kinds of challenges to overcome in the whole process of radiomics. In terms of ML, although specific algorithms can improve the accuracy of the prediction with the usage of regression models by extracting complex features in the data set, no amount of algorithm skills or computing power can extrude unwanted information (49).

Second, limitations mentioned in current articles that have already been published should be highlighted and overcome. Quite a large percent of studies are conducted in a single institution, and the sample size is too small to be convincing. Apart from that, some studies lack external validation for model development. For the extensive clinical application, the studies should be designed in a much more comprehensive and delicate way to raise the reliability of research results. To achieve future optimization of this novel method, more investigations should be carried out to test the potential of optimizing the predictive model by the combination of imaging biomarkers with other non-imaging biomarkers.

Third, it has been a long-lasting debate whether such AI-supported systems are much smarter than clinical practitioners (50, 51). However, what we should do is to take the best advantages of the new adjuvant imaging techniques and apply radiomics together with ML to bladder cancer monitoring. Combining all of these imaging analysis methods with the experience of experts will make assurance for the delivery of medical care that outperforms what either of them can achieve alone. Since no code is shared in most papers involving radiomics application, it is impossible to exactly assess the validity of the results up to now. Thus, another direction for future improvement is to achieve the validity of predictive results based on the public full code and imaging data sets and optimize the model performance in a more scientific way.

Moreover, the application of radiomics in bladder cancer remains at the macroscopic level. Although we have witnessed the breakthrough of various researches from basic tumor detection to accurate grading, even recent studies have moved toward the prediction of treatment outcomes. However, gene therapy and immunotherapy, which involve gene expression signature and immune phenotype, have been introduced as revolutionary tools for comprehensive tumor assessment and divert our attention to the microscopic level. A novel concept named "radiogenomics" perfectly presents the integration of genomics with radiomics and serves as an alternative to the invasive biopsy (52). Tremendous value as this research spot shows in several common diseases (53, 54), few studies are conducted to further validate its clinical uses in bladder cancer, which highlights that the combination of radiogenomics and bladder cancer full-cycle management can be an essential breakthrough point for future research.

AUTHOR CONTRIBUTIONS

LG and YC collected and reviewed the literature, and wrote the manuscript. JL helped with the writing design and revised the manuscript. PZhao, PZhan, and RA provided helpful comments on the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by the National Natural Science Fund of China (31801066), National Natural Science Fund of China (31741034), Science and Technology Administration, Sichuan Province (2017SZ0048), Health committee, Sichuan Province (16PJ298) and 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (ZY2016104).

REFERENCES

1. Martinez Rodriguez RH, Buisan Rueda O, Ibarz L. Bladder cancer: present and future. *Med Clin.* (2017) 149:449–55. doi: 10.1016/j.medcle.2017.10.005
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* (2018) 68:7–30. doi: 10.3322/caac.21442
3. Wu S, Zheng J, Li Y, Yu H, Shi S, Xie W, et al. A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer. *Clin Cancer Res.* (2017) 23:6904–11. doi: 10.1158/1078-0432.CCR-17-1510
4. Xu X, Liu Y, Zhang X, Tian Q, Wu Y, Zhang G, et al. Preoperative prediction of muscular invasiveness of bladder cancer with radiomic features on conventional MRI and its high-order derivative maps. *Abdom Radiol.* (2017) 42:1896–905. doi: 10.1007/s00261-017-1079-6
5. Bladder cancer: diagnosis and management of bladder cancer. *BJU Int.* (2017) 120:755–65. doi: 10.1111/bju.14045
6. Isfoss BL. The sensitivity of fluorescent-light cystoscopy for the detection of carcinoma in situ (CIS) of the bladder: a meta-analysis with comments on gold stan- dard. *BJU Int.* (2011) 108:1703–7. doi: 10.1111/j.1464-410X.2011.10485.x
7. Degeorge KC, Holt HR, Hodges SC. Bladder cancer: diagnosis and treatment. *Am Fam Phys.* (2017) 96:507. Available online at: https://www.aafp.org/afp/2017/1015/p507.html
8. Jakse G, Algaba F, Malmstrom P, Oosterlinck W. A second-look TUR in T1 transitional cell carcinoma: why? *Eur Urol.* (2004) 45:539–46. doi: 10.1016/j.eururo.2003.12.016
9. Sternberg CN. The treatment of advanced bladder cancer. *Ann Oncol.* (1995) 6:113–26. doi: 10.1093/oxfordjournals.annonc.a059105
10. Fagg SL, Dawson-Edwards P, Hughes MA, Latief TN, Rolfe EB, Fielding JW. CIS-Diamminedichloroplatinum (DDP) as initial treatment of invasive bladder cancer. *Br J Urol.* (1984) 56:296–300. doi: 10.1111/j.1464-410X.1984.tb05390.x
11. Meeks JJ, Bellmunt J, Bochner BH, Clarke NW, Daneshmand S, Galsky MD, et al. A Systematic Review of Neoadjuvant and Adjuvant Chemotherapy for Muscle-invasive Bladder Cancer. *Eur Urol.* (2012) 62:523–33. doi: 10.1016/j.eururo.2012.05.048
12. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
13. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin.* (2019) 69:127–57. doi: 10.3322/caac.21552
14. Acharya UR, Hagiwara Y, Sudarshan VK, Chan WY, Ng KH. Towards precision medicine: from quantitative imaging to radiomics. *J Zhejiang Univ Sci B.* (2018) 19:6–24. doi: 10.1631/jzus.B1700260
15. Arimura H, Soufi M, Kamezawa H, Ninomiya K, Yamada M. Radiomics with artificial intelligence for precision medicine in radiation therapy. *J Radiat Res.* (2019) 60:150–7. doi: 10.1093/jrr/rry077
16. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
17. Scrivener M, de Jong EEC, van Timmeren JE, Pieters T, Ghaye B, Geets X. Radiomics applied to lung cancer: a review. *Transl Cancer Res.* (2016) 5:398–409. doi: 10.21037/tcr.2016.06.18
18. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2014) 5:4006. doi: 10.1038/ncomms5644
19. Castellino RA. Computer-aided detection (CAD): an overview. *Cancer Imaging.* (2005) 5:17–9. doi: 10.1102/1470-7330.2005.0018
20. Court LE, Fave X, Mackin D, Lee J, Yang J, Zhang L. Computational resources for radiomics. *Transl Cancer Res.* (2016) 5:340–8. doi: 10.21037/tcr.2016.06.17
21. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2015) 278:151169. doi: 10.1148/radiol.2015151169
22. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champia S, Han SR, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol.* (2018) 19:1180–91. doi: 10.1016/S1470-2045(18)30413-3
23. Joseph D, Sachar S, Kishore N, Chandra S. Mechanistic insights into the interactions of magnetic nanoparticles with bovine serum albumin in presence of surfactants. *Colloids Surf B Biointerfaces.* (2015) 135:596–603. doi: 10.1016/j.colsurfb.2015.08.022
24. Kim T, Ahn C, Lee O. Image segmentation by graph cut for radiation images of small animal blood vessels. *Microsc Res Tech.* (2018) 81:1506–12. doi: 10.1002/jemt.23154
25. Chen YT. A novel approach to segmentation and measurement of medical image using level set methods. *Magn Reson Imaging.* (2017) 39:175–93. doi: 10.1016/j.mri.2017.02.008
26. Dey S, Sarkar R, Chatterjee K, Datta P, Barui A, Maity SP. Pre-cancer risk assessment in habitual smokers from DIC images of oral exfoliative cells using active contour and SVM analysis. *Tissue Cell.* (2017) 49:296–306. doi: 10.1016/j.tice.2017.01.009
27. Lubner MG, Smith AD, Sandrasegaran K, Sahani DV, Pickhardt PJ. CT texture analysis: definitions, applications, biologic correlates, and challenges. *Radiographics.* (2017) 37:1483–503. doi: 10.1148/rg.2017170056
28. Zhao X, Nie F, Wang S, Guo J, Xu P, Chen X. Unsupervised 2D dimensionality reduction with adaptive structure learning. *Neural Comput.* (2017) 29:1352–74. doi: 10.1162/NECO_a_00950
29. Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol.* (2017) 10:257–73. doi: 10.1007/s12194-017-0406-5
30. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. *Clin Lung Cancer.* (2016) 17:441–8. doi: 10.1016/j.cllc.2016.02.001
31. Kickingereder P, Burth S, Wick A, Götz M, Eidel O, Schlemmer HP, et al. Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology.* (2016) 280:880–9. doi: 10.1148/radiol.2016160845
32. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: the promise of radiomics. *Phys Med.* (2017) 38:122–39. doi: 10.1016/j.ejmp.2017.05.071
33. Shahnas MH, Yuen DA, Pysklywec RN. Inverse problems in geodynamics using machine learning algorithms. *J Geophys Res Solid Earth.* (2018) 123:296–310. doi: 10.1002/2017JB014846

34. Parmar C, Grossmann P, Bussink J, Lambin, Aerts HJ. Machine learning methods for quantitative radiomic biomarkers. *Sci. Rep.* (2015) 5:13087. doi: 10.1038/srep13087
35. Garapati SS, Hadjiiski L, Cha KH, Chan HP, Caoili EM, Cohan RH, et al. Urinary bladder cancer staging in CT urography using machine learning. *Med Phys.* (2017) 44:5814–23. doi: 10.1002/mp.12510
36. Zhang X, Xu X, Tian Q, Li B, Wu Y, Yang Z. Radiomics assessment of bladder cancer grade using texture features from diffusion-weighted imaging. *J Magn Reson Imaging.* (2017) 46:1281–8. doi: 10.1002/jmri.25669
37. Wang H, Hu D, Yao H, Chen M, Li S, Chen H, et al. Radiomics analysis of multiparametric MRI for the preoperative evaluation of pathological grade in bladder cancer tumors. *Eur Radiol.* (2019) 29:6182–90. doi: 10.1007/s00330-019-06222-8
38. Zheng J, Kong J, Wu S, Li Y, Cai J, Yu H, et al. Development of a noninvasive tool to preoperatively evaluate the muscular invasiveness of bladder cancer using a radiomics approach. *Cancer*. (2019). doi: 10.1002/cncr.32490
39. Wang G, Lam K-M, Deng Z. Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques. *Comp Biol Med.* (2015) 63:124–32. doi: 10.1016/j.compbiomed.2015.05.015
40. Xu X, Wang H, Du P, Zhang F. A predictive nomogram for individualized recurrence stratification of bladder cancer using multiparametric MRI and clinical risk factors. *J Magn Reson Imaging.* (2019) 50:1893–904. doi: 10.1002/jmri.26749
41. Lin P, Wen DY, Chen L, Li X, Li SH, Yan HB, et al. A radiogenomics signature for predicting the clinical outcome of bladder urothelial carcinoma. *Eur Radiol.* (2019). doi: 10.1007/s00330-019-06371-w. [Epub ahead of print].
42. Cha KH, Hadjiiski LM, Samala RK, Chan HP, Cohan RH, Caoili EM, et al. Bladder cancer segmentation in CT for treatment response assessment: application of deep-learning convolution neural network—a pilot study. *Tomogr A J Imaging Res.* (2016) 2:421–9. doi: 10.18383/j.tom.2016.00184
43. Cha KH, Hadjiiski L, Chan HP, Weizer AZ, Alva A, Cohan RH, et al. Bladder cancer treatment response assessment in CT using radiomics with deep-learning. *Sci Rep.* (2017) 7:8738. doi: 10.1038/s41598-017-09315-w
44. Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. *PLoS ONE.* (2015) 10:e0124165. doi: 10.1371/journal.pone.0124165
45. Turker P, Bostrom PJ, Wroclawski ML, van Rhijn B, Kortekangas H, Kuk C, et al. Upstaging of urothelial cancer at the time of radical cystectomy: factors associated with upstaging and its effect on outcome. *BJU Int.* (2012) 110:804–11. doi: 10.1111/j.1464-410X.2012.10939.x
46. Hadjiiski L, Weizer AZ, Alva A, Caoili EM, Cohan RH, Cha K, et al. Treatment response assessment for bladder cancer on CT based on computerized volume analysis, World Health Organization Criteria, and RECIST. *Am J Roentgenol.* (2015) 205:348–52. doi: 10.2214/AJR.14.13732
47. Cha KH, Hadjiiski LM, Cohan RH, Chan HP, Caoili EM, Davenport MS, et al. Diagnostic accuracy of CT for prediction of bladder cancer treatment response with and without computerized decision support. *Acad Radiol.* (2018) 26:1137–45. doi: 10.1016/j.acra.2018.10.010
48. Larue RT, Defraene G, De Ruysscher D, Lambin P, van Elmpt W. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *Br J Radiol.* (2017) 90:20160665. doi: 10.1259/bjr.20160665
49. Stephan G. Machine learning meets databases. *Datenbank Spektrum*. (2017) 17:77–83. doi: 10.1007/s13222-017-0247-8
50. Chen JH, Asch SM. Machine learning and prediction in medicine — beyond the peak of inflated expectations. *N Engl J Med.* (2017) 376:2507–9. doi: 10.1056/NEJMp1702071
51. Yip SS, Aerts HJ. Applications and limitations of radiomics. *Phys Med Biol.* (2016) 61:R150–66. doi: 10.1088/0031-9155/61/13/R150
52. Rosenstein BS. Radiogenomics: identification of genomic predictors for radiation toxicity. *Semin Radiat Oncol.* (2017) 27:300–9. doi: 10.1016/j.semradonc.2017.04.005
53. Badic B, Hatt M, Durand S, Jossic-Corcos CL, Simon B, Visvikis D, et al. Radiogenomics-based cancer prognosis in colorectal cancer. *Sci Rep.* (2019) 9:9743. doi: 10.1038/s41598-019-46286-6
54. Yeh AC, Li H, Zhu Y, Zhang J, Khramtsova G, Drukker K, et al. Radiogenomics of breast cancer using dynamic contrast enhanced MRI and gene expression profiling. *Cancer Imaging.* (2019) 19:48. doi: 10.1186/s40644-019-0233-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 06 December 2019
doi: 10.3389/fonc.2019.01338

The Diagnostic Value of Radiomics-Based Machine Learning in Predicting the Grade of Meningiomas Using Conventional Magnetic Resonance Imaging: A Preliminary Study

Chaoyue Chen[1,2,3], *Xinyi Guo*[3,4], *Jian Wang*[5], *Wen Guo*[1,2,3,4], *Xuelei Ma*[1,2]* *and Jianguo Xu*[3]*

[1] Department of Biotherapy, Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [2] State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, and Collaborative Innovation Center for Biotherapy, Chengdu, China, [3] Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu, China, [4] West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China, [5] School of Computer Science, Nanjing University of Science and Technology, Nanjing, China

OPEN ACCESS

Edited by:
Natalie Julie Serkova,
University of Colorado Denver,
United States

Reviewed by:
Di Dong,
Institute of Automation (CAS), China
Seyedmehdi Payabvash,
Yale University, United States

***Correspondence:**
Xuelei Ma
drmaxuelei@gmail.com
Jianguo Xu
drjianguoxu@gmail.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 25 June 2019
Accepted: 15 November 2019
Published: 06 December 2019

Citation:
Chen C, Guo X, Wang J, Guo W, Ma X and Xu J (2019) The Diagnostic Value of Radiomics-Based Machine Learning in Predicting the Grade of Meningiomas Using Conventional Magnetic Resonance Imaging: A Preliminary Study. Front. Oncol. 9:1338. doi: 10.3389/fonc.2019.01338

Objective: The purpose of the current study is to investigate whether texture analysis-based machine learning algorithms could help devise a non-invasive imaging biomarker for accurate classification of meningiomas using machine learning algorithms.

Method: The study cohort was established from the hospital database by reviewing the medical records. Patients were selected if they underwent meningioma resection in the neurosurgery department between January 2015 and December 2018. A total number of 40 texture parameters were extracted from pretreatment postcontrast T1-weighted (T1C) images based on six matrixes. Three feature selection methods were adopted, namely, distance correlation, least absolute shrinkage and selection operator (LASSO), and gradient boosting decision tree (GBDT). Multiclass classification methods of linear discriminant analysis (LDA) and support vector machine (SVM) algorithms were employed to establish the classification models. The diagnostic performances of models were evaluated with confusion matrix based on which the areas under the curve, accuracy, and Kappa value of models were calculated.

Result: Confusion matrix showed that the LDA-based models represented better diagnostic performances than SVM-based models. The highest accuracy among LDA-based models was 75.6%, shown in the combination of Lasso + LDA. The optimal models for SVM-based models was Lasso+SVM, with accuracy of 59.0% in the testing group. One of the SVM-based models, GBDT+SVM, was overfitting, suggesting that this model was not suitable for application.

Conclusion: Machine learning algorithms with texture features extracted from T1C images could potentially serve as the assistant imaging biomarkers for presurgically grading meningiomas.

Keywords: radiomics, machine learning, magnetic resonance imaging, meningioma, tumor grade

INTRODUCTION

According to the survey conducted by the Central Brain Tumor Registry of the United States (CBTRUS), meningiomas are one of the most frequent intracranial tumors in adults, with an incidence of 8.14/100,000, accounting for 36.8% of the primary central nervous system tumors (1). In most cases, meningiomas are histologically recognized as low-grade meningioma (WHO grade I) with benign behaviors, but approximately 10–20% of meningiomas are recognized as high-grade meningioma (WHO grades II and III), exhibiting aggressive behaviors (2–4). The treatment and prognosis for meningioma are intimately related to the histopathological grade (5). Surgical resection is the first-line treatment for all types of meningiomas, the extent of surgical resection is the most important prognostic factor for high-grade meningioma outcomes. According to the previous investigations, adjuvant radiotherapy is associated with statistically improved overall survival (OS) and progression-free survival (PFS) outcomes (6–9). Moreover, the prognoses of different grades of meningiomas are dramatically different that higher grades meningiomas are correlated to higher recurrence rate (7–25, 29–52, and 50–94%, respectively) and poor survival outcomes (5, 10). Given these differences in treatment and prognosis, the accurate presurgical assessment on tumor grade is clinically important to facilitate treatment decisions.

Lacking specific blood biomarkers, magnetic resonance imaging (MRI) is the most importent imaging technique in the detection and presurgical assessment of intracranial meningiomas. Previous studies demonstrated that preoperative MRI was useful for assessing the grades and evaluating histopathological characteristics of meningiomas (11–14). However, the image patterns of different grades of meningiomas could mimic each other in some cases, resulting in limited diagnostic accuracy and highlighting the urgency of new radiological evaluation methods (15). Texture analysis is a subset of radiomics. With the ability of mathematically converting medical images into mineable quantitative statistics, it has been considered as the emerging field providing a non-invasive assessment on tumor heterogeneity (16). Theoretically, the texture parameters can objectively calculate the structural and spectral characteristics of pixel intensities within an area to extract quantitative metrics that are impossible to assess visually (17, 18). Compared with traditional visual assessment, texture analysis can describe the image with quantitative statistics more sensitively and accurately (19).

Texture analysis has shown promising diagnostic ability in meningioma grading in previous studies (14, 20–22). Additionally, the quantitative evaluation of texture features has been applied into machine learning technology to differentiate high-grade meningiomas from low-grade meningiomas (20, 21). In the current study, we applied multiple classification methods to systematically grade meningiomas. Six models were established and evaluated, aiming to preliminarily investigate the value of radiomics-based machine learning technology in in preoperative prediction of meningioma grades.

MATERIALS AND METHODS

Patient Selection

This retrospective study was led in the neurosurgery department of our hospital. We viewed the electronic medical records to search for patients with detailed pathological reports on meningiomas between January 2015 and December 2018. The presurgical high-quality MR images of patients were also exported with standard format through PACS (Picture Archiving and Communication System). After the initial evaluation on images and patient profiles, we excluded some patients due to the following reasons: (1) images with motion artifacts; (2) relevant tumor treatment history (like radiotherapy or surgery) in other hospitals; (3) recorded intracranial diseases history, such as subarachnoid hemorrhage, cerebral infarction, and so on. Finally, a total number of 150 meningioma patients were introduced in our study. Clinical information and pathological reports were also recorded for further analysis. It is worth noting that pathological grading was corrected based on the 2016 WHO classification system, adjusted by a senior neuropathologist with 10 years of experience.

The institutional review board approved this retrospective study. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The obligatory written informed consent was obtained from participants enrolled in this study (written informed consent for patients <16 years old was signed by parents or guardians). The patients agreed to undertake examination and were informed that the statistics (including MR image), which could be used for academic purpose in the future, would be stored in our institutional database. The Ethics Committee of Sichuan University and neurosurgery department of our institution have given approval for statistics export and utilization for this study.

MRI Acquisition

After consulting with senior radiologists and neurosurgeons, postcontrast T1-weighted (T1C) images were selected for further analysis due to clear depiction of tumor location and boundary (**Figure 1**). The MR scan was conducted in the MR Research Center of our hospital with 3.0T Siemens Trio Scanner. High-quality three-dimensional T1-weighted images were obtained by using a magnetization prepared rapid gradient-echo (MPRAGE) sequence by the following protocols: TR/TE/TI = 1,900/2.26/900 ms, Flip angle = 9°, 176 axial slices with thickness = 1 mm, axial FOV = 25.6 × 25.6 cm^2, and data matrix = 256 × 256. The contrast-enhanced image was acquired with gadopentetate dimeglumine (dose: 0.1 mmol/kg) as the contrast agent.

Texture Features Extraction

The texture analysis was conducted with LIFEx software by two neurosurgeons following the software instructions (23). The authors contoured along the tumor tissue slice by

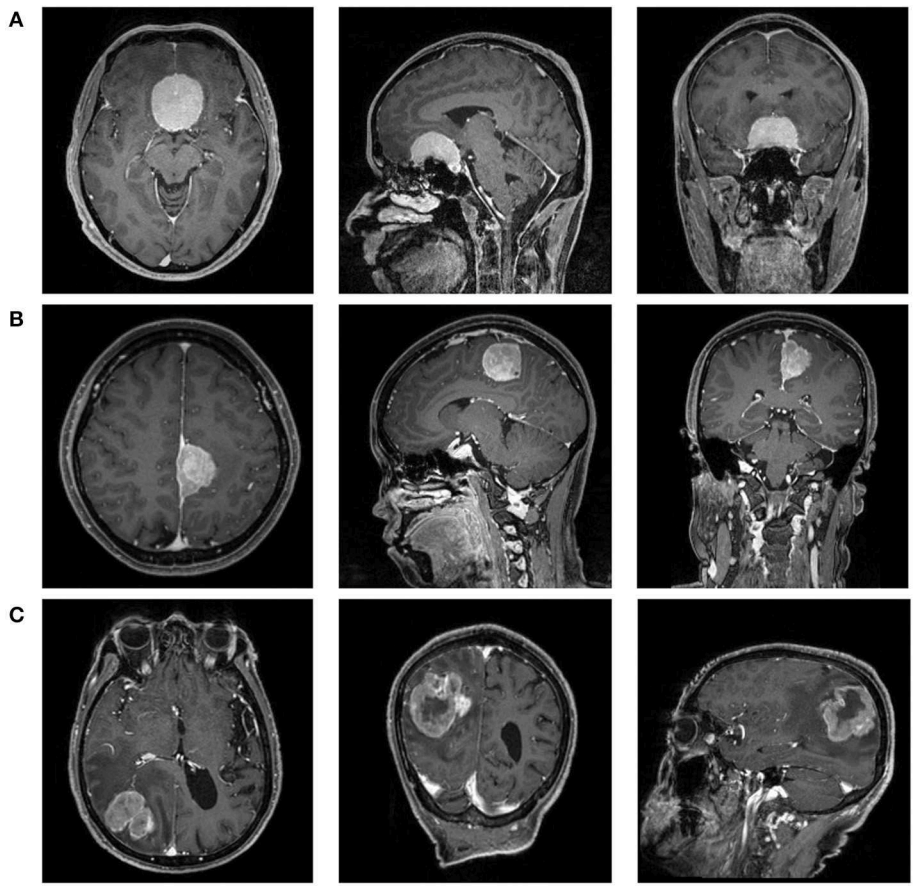


FIGURE 1 | The magnetic resonance images [postcontrast T1-weighted (T1C)] of a patient with **(A)** WHO I meningioma, **(B)** WHO II meningioma, and **(C)** WHO III meningioma.

slice to draw the region of interest (ROI), and the three-dimensional texture features were automatically generated with default setting. Any disagreement regarding the tumor location or border of lesions were resolved by consulting senior neurosurgeons and the senior radiologist. Forty quantified texture features were extracted, including features from histogram-based matrix and shape-based matrix from the first order and features from gray-level co-occurrence matrix (GLCM), gray-level zone length matrix (GLZLM), neighborhood gray-level dependence matrix (NGLDM), and gray-level run length matrix (GLRLM) from second or higher order (**Supplement Material 1**). The definitions of texture parameters were summarized in **Supplement Material 2**. The association between texture parameters was evaluated using Pearson correlation coefficient test.

Machine Learning Classification

The classification models were built with different combinations of three selection methods [distance correlation, least absolute shrinkage and selection operator (LASSO), and gradient boosting decision tree (GBDT)] and two multiclass classification algorithms [linear discriminant analysis (LDA) and support vector machine (SVM)]. The feature selection was essential to the diagnostic performance given that diagnostic values on all features were discrepant, and that optimal features can statistically eliminate overfitting. Moreover, it can contribute to decreased running time and increased accuracy of the models. With selected features retrieved from different methods, the statistics were employed into algorithms separately. Two multiple classification algorithms were adopted in the current study, including LDA and SVM, representing the linear classifier and non-linear classifier, respectively. The patients were randomly separated into two parts in the proportion of 4:1 as the training group and the testing group. Confusion matrixes and areas under the curve (AUC) of each model were calculated to evaluate the performance of the models. The algorithms deployment procedure was repeated 100 times to obtain the realistic distribution of classification accuracies.

The regular statistical analyses in this study were conducted using SPSS software (version 21; IBM, Chicago), including Mann-Whitney *U*-test and Pearson correlation coefficient. The machine learning algorithms were programmed using Python Programming Language and scikit-learn package.

RESULTS

Characteristics of the Study Cohort

A total number of 150 patients were involved in the current study, among whom 61 were diagnosed with WHO I meningioma, 59 with WHO II meningiomas, and 30 with WHO III meningiomas. The mean ages of patients were 49.38, 54.41, and 56.93 years, respectively. The gender ratio for patients was 62:88 (male:female). The clinical characteristics of patients and tumors were summarized in **Table 1**.

Characteristics of Texture Parameters

The results of Mann-Whitney suggested that there was no statistically significant difference among the parameters extracted by two neurosurgeons, implying that the results could be considered reliable and reproducible (**Supplement Material 3**). The Pearson correlation coefficient suggested that most texture features were correlated with each other rather than independent (**Figure 2**).

Diagnostic Performance of Models

In the feature selection, some mutual features were selected when using different methods, suggesting that they were the most significant features in discrimination (**Table 2**). Generally, the LDA-based models represented better performance than SVM-based models. The accuracy rates for the LDA-based models were 73.0, 75.6, and 73.3 in the testing group whereas for the SVM-based models were 57.6 and 59.0%. Overfitting was observed in one model, SVM+GBDT, suggesting that this model was inappropriate in application. The AUC, Kappa value, and accuracy of each model were represented in **Table 3**.

Figure 3 illustrated the performance of the LDA-based models in terms of the distribution of the canonical functions for one of the 100 independent training cycles. **Figure 4** illustrated the examples of distributions of the LDA function determined for the lesions for one cycle.

TABLE 1 | Characteristics of patients and lesions.

	Low-grade meningioma	High-grade meningioma	
	WHO I meningioma	**WHO II meningioma**	**WHO III meningioma**
Number	61	59	30
Age	49.38	54.41	56.93
Gender (*n*, %)			
Male	16 (26.23%)	32 (54.24%)	14 (46.67%)
Female	45 (73.77%)	27 (45.76%)	16 (53.33%)
Maximum diameter (cm)	4.06 ± 1.53	5.75 ± 1.50	6.93 ± 2.03
Location (*n*, %)			
Cerebral convexity	32 (52.46%)	40 (67.80%)	21 (70.00%)
Falx	11 (18.03%)	8 (13.56%)	2 (6.67%)
Skull base	18 (29.51%)	11 (18.64%)	7 (23.33%)
Days between MR scan and surgery	8.7 days	7.2 days	6.7 days

DISCUSSION

The prediction of the histopathological meningioma grade is important because it is closely related to survival outcomes and treatment strategies. According to the instructions of the National Comprehensive Cancer Network (NCCN) guideline, the recommended treatment for WHO grade I meningioma was surgical resection or observation; for WHO grade II meningioma, it was gross total resection combined with/without radiotherapy; and for a WHO grade III meningioma, it was radical surgery with radiotherapy (24). Therefore, the accurate preoperative diagnosis should assist clinicians in making a personalized treatment plan to improve the quality of life. In the current study, we investigated the diagnostic value of texture analysis-based machine learning technology in meningioma grade. The texture features adopted into the classifiers were extracted from T1C images, which brought the possibility to utilize the technology in standard routine care imaging analyses.

Texture analysis provides information on the heterogeneity of tumor imaging, such as tumor cellularity, degenerative changes, and neovascularization, which are hard to assess visually. By analyzing the spectral distribution of pixels, abnormal tumor microenvironment and pathology could be represented as a series of statistics (25). It has been reported that an imaging technology extends beyond radiology to histopathology, like prediction on gene mutation and tumor grading (26–32). As for the different grade meningiomas, the characteristics of enhanced pattern have been reported in previous researches. Specifically, MRI features, such as positive capsular enhancement, indistinct tumor–brain interface, and heterogeneous tumor enhancement, were suggested to be related to a higher tumor grade (33, 34). These MRI features could be reflected in GLZLM_ZLNU, one of the mutual selected features in our study. This feature calculates the non-uniformity of the gray-levels or the length of the homogeneous zones, reflecting the heterogeneity within the delineated area. Fluctuance on value of features from second or higher order represented irregular changes in the gray pixels of aggressive meningiomas due to the heterogeneous structure inside tumor tissue (11). Therefore, it is reasonable to consider that this MRI feature was closely correlated to this texture feature. Another mutual feature, SHAPE_Volume (ml), suggested that the tumor volume was also in relation to grade, according to the differences in tumor diameter. However, it is worthy to note that most features were correlated with each other; the specific reason is still unclear that is why GLZLM_ZLNU was selected as the strongest correlated feature while others were not. Future researches are required to explore this question.

The value of radiomics-based machine learning in meningioma grading has been explored before. Retrieved parameters, feature selection method, sample size, and classification algorithms determined the performance of models. However, all of these studies, as well as this study, was seriously limited by the small sample size due to the rather low incidents of grade III meningioma. Therefore, most of them simply classified them into low-grade and high-grade (21, 35). Only one study explored the multiple classification

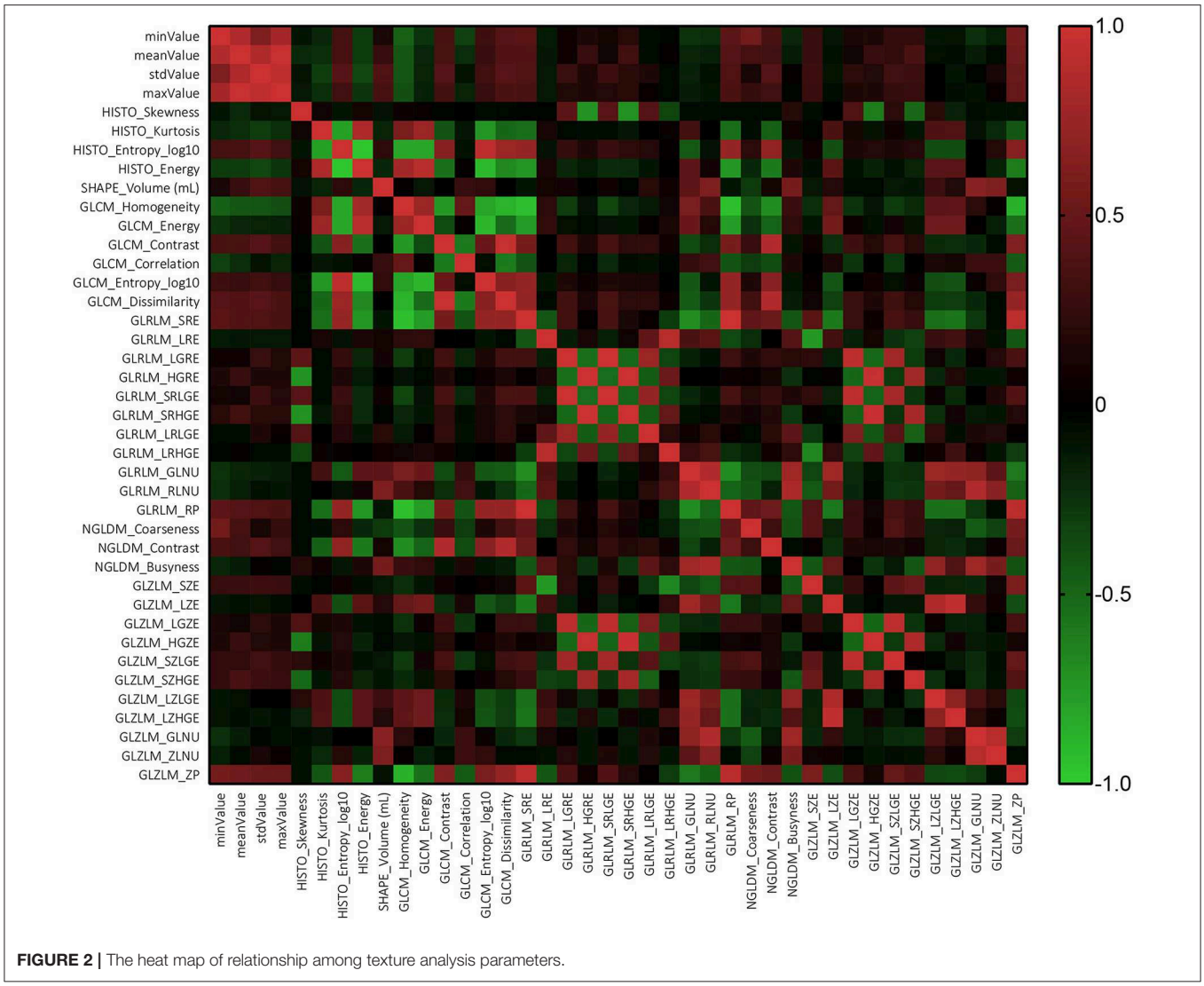


FIGURE 2 | The heat map of relationship among texture analysis parameters.

TABLE 2 | Selected features using distance correlation, LASSO, and GBDT.

Selection method	Selected features
Distance Correlation	HISTO_Kurtosis, HISTO_Entropy, HISTO_Energy, SHAPE_Volume, GLCM_Energy, GLCM_Entropy_log10, NGLDM_Contrast, GLZLM_ZLNU
LASSO	minValue, meanValue, stdValue, SHAPE_Volume (ml), GLCM_Contrast, GLRLM_HGRE, GLRLM_LRHGE, GLRLM_GLNU, GLRLM_RLNU, GLZLM_LZE, GLZLM_HGZE, GLZLM_SZHGE, GLZLM_LZHGE, GLZLM_GLNU, GLZLM_ZLNU
GBDT	minValue, HISTO_Skewness, SHAPE_Volume (ml), GLCM_Homogeneity, GLCM_Energy, GLCM_Correlation, GLCM_Entropy_log10, GLCM_Dissimilarity, GLRLM_LRLGE, GLRLM_RLNU, NGLDM_Contrast, GLZLM_SZE, GLZLM_LZHGE, GLZLM_ZLNU

LASSO, least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree.

models in discrimination, which established models with the parameters extracted from ADC map and decision trees algorithms, demonstrating the equivalent diagnostic performance of machine learning technology compared to experienced neuroradiologists (accuracy = 79.51%, Kappa value = 0.6393) (14). As for this study, we employed different multiple classification algorithms and texture features from different sequences. However, we should note that the differences between the models were not strong enough to select the optimal one, specifically considering that the investigated models seemed to perform quite comparably and that the variance in AUC might be partially attributed to the small sample size. Therefore, our results could only be regarded as a hypothesis and need to be verified in future studies.

LDA and SVM were employed as classification algorithms in the current study. Both of them are considered state-of-the-art in pattern recognition technology, representing two

TABLE 3 | Diagnostic performance of classification models.

Models		Training group				Validation group			
		WHO Grade I	WHO Grade II	WHO Grade III	Kappa value	WHO Grade I	WHO Grade II	WHO Grade III	Kappa value
LDA	Distance Correlation	0.928	0.865	0.882	0.578 (Accuracy = 75.4%)	0.884	0.820	0.846	0.563 (Accuracy = 73.0%)
	LASSO	0.955	0.914	0.915	0.693 (Accuracy = 80.8%)	0.934	0.846	0.783	0.603 (Accuracy = 75.6%)
	GBDT	0.928	0.950	0.908	0.570 (Accuracy = 73%)	0.886	0.854	0.887	0.572 (Accuracy = 73.3%)
SVM	Distance Correlation	0.870	0.831	0.876	0.356 (Accuracy = 61.1%)	0.845	0.798	0.845	0.274 (Accuracy = 57.6%)
	LASSO	0.898	0.806	0.877	0.373 (Accuracy = 62.0%)	0.840	0.772	0.833	0.298 (Accuracy = 59.0%)
	GBDT (Overfitting)	–	–	–	–	–	–	–	

LDA, Linear Discriminate Analysis; SVM, Support Vector Machine; LASSO, Least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree.

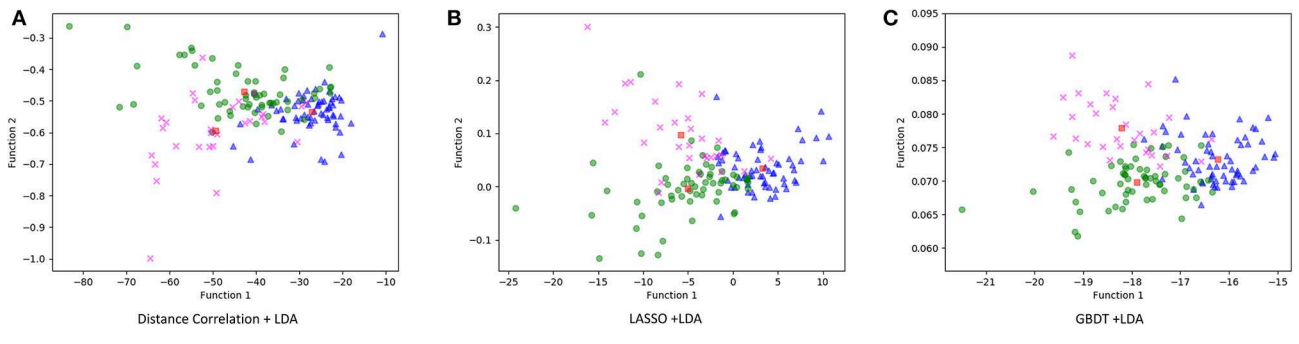


FIGURE 3 | Distribution of the discriminant functions of LDA models. **(A)** Distance correlation + LDA; **(B)** least absolute shrinkage and selection operator (LASSO) + LDA; and **(C)** gradient boosting decision tree (GBDT) + LDA.

different types of classifiers (36). LDA is the linear classifier, consisting of the shape of the decision boundary of straight line in the first case and straight line in second, whereas SVM is the non-linear classifier, of which the shape of the decision boundary is a plane in the first case and a plane in the second (36, 37). Computational time and complexity usually increase together when trying to improve the performance. Therefore, the importance on the trade-off between computational burden and performance has been highlighted to require a suitable selection method. Previous studies performed feature selection with Friedman test or Mann-Whitney *U*-test to choose the most significant features into classifiers, suggesting that the selected features could improve the classifier performances (11, 14, 35). The results of our study showed that all LDA-based models represented better performances than SVM-based models, and that the improvement using different selected models was limited. It seemed that the algorithms have more priority than the selection method in the improvement of diagnostic performances. Therefore, in futures studies, researchers should focus on the algorithm selection, and novel algorithms should be investigated.

There were some limitations in the current study. First, this is a single-institution retrospective study that enrolled 150 patients. The patient sample was relatively small, and the selection bias was inevitable. Second, the texture features into classifiers were extracted from T1C sequence, while the value of features from other sequences was unclear. Given that the research using parameters extracted from ADC images represented better performance, future researches were required to investigate whether the diagnostic performance could be improved when combined with features from other sequences and advanced MR technology. Third, novel radiomics parameters have been identified in recent years, while our studies only involved traditional texture parameters. Compared to many other studies in the same field, the number of radiomics features ($n = 40$) is fairly small. Fourth, we did not perform comparison to the performance of a human reader with classifiers. Fifth, only few classification algorithms were evaluated in our study. Machine learning has been developed rapidly in recent years, and new algorithms are being programmed. Sixth, we did not verify the efficacy of machine learning-based models in external datasets. We tried to search the public datasets, but all of them were

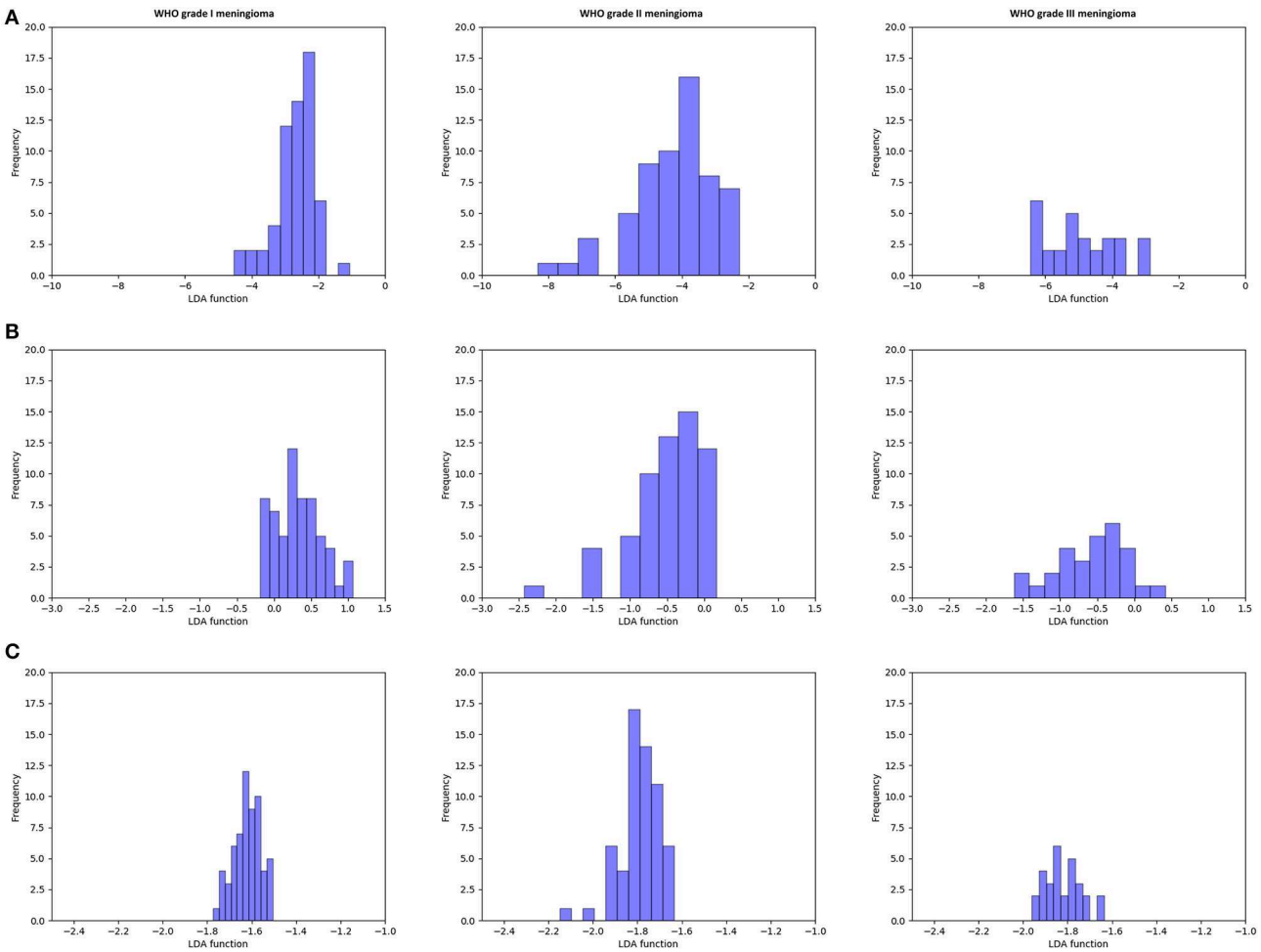


FIGURE 4 | Example of distributions of the linear discriminant analysis (LDA)-based models determined for the lesions for one cycle. **(A)** Distance correlation + LDA; **(B)** least absolute shrinkage and selection operator (LASSO) + LDA; and **(C)** gradient boosting decision tree (GBDT) + LDA.

for gliomas. The software used to extract texture parameters and package to perform machine learning is the open-source package, providing a potential for other researchers to reproduce our researches.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee of Sichuan University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

CC participated in collecting MR images, extracting statistics, and drafting the manuscript. XG participated in collecting MR images and extracting statistics. JW established and deployed the machine learning models. WG participated in collecting MR images and extracting statistics. XM and JX participated in conceptualization and revised some intellectual content in the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01338/full#supplementary-material

Supplement Material 1 | The texture analysis parameters extracted from images.

Supplement Material 2 | The definitions of texture analysis parameters.

Supplement Material 3 | The results of Mann-Whitney *U*-test in the examination of repeatability of texture analysis parameters extraction.

REFERENCES

1. Ostrom QT, Gittleman H, Liao P, Vecchione-Koval T, Wolinsky Y, Kruchko C, et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. *Neuro Oncol*. (2017) 19(Suppl_5):v1–v88. doi: 10.1093/neuonc/nox158
2. Rogers L, Barani I, Chamberlain M, Kaley TJ, McDermott M, Raizer J, et al. Meningiomas: knowledge base, treatment outcomes, and uncertainties. A RANO review. *J Neurosurg*. (2015) 122:4–23. doi: 10.3171/2014.7.JNS131644
3. Goldbrunner R, Minniti G, Preusser M, Jenkinson MD, Sallabanda K, Houdart E, et al. EANO guidelines for the diagnosis and treatment of meningiomas. *Lancet Oncol*. (2016) 17:e383–91. doi: 10.1016/S1470-2045(16)30321-7
4. Whittle IR, Smith C, Navoo P, Collie D. Meningiomas. *Lancet*. (2004) 363:1535–43. doi: 10.1016/S0140-6736(04)16153-9
5. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. (2016) 131:803–20. doi: 10.1007/s00401-016-1545-1
6. Sun SQ, Cai C, Murphy RK, DeWees T, Dacey RG, Grubb RL, et al. Management of atypical cranial meningiomas, part 2: predictors of progression and the role of adjuvant radiation after subtotal resection. *Neurosurgery*. (2014) 75:356–63; discussion 363. doi: 10.1227/NEU.0000000000000462
7. Komotar RJ, Iorgulescu JB, Raper DM, Holland EC, Beal K, Bilsky MH, et al. The role of radiotherapy following gross-total resection of atypical meningiomas. *J Neurosurg*. (2012) 117:679–86. doi: 10.3171/2012.7.JNS112113
8. Aizer AA, Arvold ND, Catalano P, Claus EB, Golby AJ, Johnson MD, et al. Adjuvant radiation therapy, local recurrence, and the need for salvage therapy in atypical meningioma. *Neuro Oncol*. (2014) 16:1547–53. doi: 10.1093/neuonc/nou098
9. Wang C, Kaprealian TB, Suh JH, Kubicky CD, Ciporen JN, Chen Y, et al. Overall survival benefit associated with adjuvant radiotherapy in WHO grade II meningioma. *Neuro Oncol*. (2017) 19:1263–70. doi: 10.1093/neuonc/nox007
10. Moliterno J, Cope WP, Vartanian ED, Reiner AS, Kellen R, Ogilvie SQ, et al. Survival in patients treated for anaplastic meningioma. *J Neurosurg*. (2015) 123:23–30. doi: 10.3171/2014.10.JNS14502
11. Lu Y, Xiong J, Yin B, Wen J, Liu L, Geng D. The role of three-dimensional pseudo-continuous arterial spin labelling in grading and differentiating histological subgroups of meningiomas. *Clin Radiol*. (2018) 73:176–84. doi: 10.1016/j.crad.2017.08.005
12. Svolos P, Tsolaki E, Theodorou K, Fountas K, Kapsalaki E, Fezoulidis I, et al. Classification methods for the differentiation of atypical meningiomas using diffusion and perfusion techniques at 3-T MRI. *Clin Imaging*. (2013) 37:856–64. doi: 10.1016/j.clinimag.2013.03.006
13. Gutman DA, Dunn WD, Grossmann P, Cooper LA, Holder CA, Ligon KL, et al. Somatic mutations associated with MRI-derived volumetric features in glioblastoma. *Neuroradiology*. (2015) 57:1227–37. doi: 10.1007/s00234-015-1576-7
14. Lu Y, Liu L, Luan S, Xiong J, Geng D, Yin B. The diagnostic value of texture analysis in predicting WHO grades of meningiomas based on ADC maps: an attempt using decision tree and decision forest. *Eur Radiol*. (2019) 29:1318–28. doi: 10.1007/s00330-018-5632-7
15. Yin B, Liu L, Zhang BY, Li YX, Li Y, Geng DY. Correlating apparent diffusion coefficients with histopathologic findings on meningiomas. *Eur J Radiol*. (2012) 81:4050–6. doi: 10.1016/j.ejrad.2012.06.002
16. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
17. Soni N, Priya S, Bathla G. Texture analysis in cerebral gliomas: a review of the literature. *AJNR Am J Neuroradiol*. (2019) 40:928–34. doi: 10.3174/ajnr.A6075
18. Depeursinge A, Foncubierta-Rodriguez A, Van De Ville D, Müller H. Three-dimensional solid texture analysis in biomedical imaging: review and opportunities. *Med Image Anal*. (2014) 18:176–96. doi: 10.1016/j.media.2013.10.005
19. Kassner A, Thornhill RE. Texture analysis: a review of neurologic MR imaging applications. *AJNR Am J Neuroradiol*. (2010) 31:809–16. doi: 10.3174/ajnr.A2061
20. Hale AT, Stonko DP, Wang L, Strother MK, Chambless LB. Machine learning analyses can differentiate meningioma grade by features on magnetic resonance imaging. *Neurosurg Focus*. (2018) 45:E4. doi: 10.3171/2018.8.FOCUS18191
21. Park YW, Oh J, You SC, Han K, Ahn SS, Choi YS, et al. Texture analysis and machine learning may accurately predict the grade and histological subtype in meningiomas using conventional and diffusion tensor imaging. *Eur Radiol*. (2018) 29:4068–76. doi: 10.1007/s00330-018-5830-3
22. Laukamp KR, Thiele F, Shakirin G, Zopfs D, Faymonville A, Timmer M, et al. Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI. *Eur Radiol*. (2019) 29:124–32. doi: 10.1007/s00330-018-5595-8
23. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res*. (2018) 78:4786–9. doi: 10.1158/0008-5472.CAN-18-0125
24. Nabors LB, Portnow J, Ammirati M, Baehring J, Brem H, Butowski N, et al. NCCN guidelines insights: central nervous system cancers, version 1.2017. *J Natl Compr Canc Netw*. (2017) 15:1331–45. doi: 10.6004/jnccn.2017.0166
25. Ahn SY, Park CM, Park SJ, Kim HJ, Song C, Lee SM, et al. Prognostic value of computed tomography texture features in non-small cell lung cancers treated with definitive concomitant chemoradiotherapy. *Invest Radiol*. (2015) 50:719–25. doi: 10.1097/RLI.0000000000000174
26. Kocak B, Durmaz ES, Ates E, Ulusan MB. Radiogenomics in clear cell renal cell carcinoma: machine learning-based high-dimensional quantitative CT texture analysis in predicting PBRM1 mutation status. *AJR Am J Roentgenol*. (2019) 212:W55–63. doi: 10.2214/AJR.18.20443
27. Oh JE, Kim MJ, Lee J, Hur BY, Kim B, Kim DY, et al. Magnetic resonance-based texture analysis differentiating KRAS mutation status in rectal cancer. *Cancer Res Treat*. (2019). doi: 10.4143/crt.2019.050. [Epub ahead of print].
28. Ozkan E, West A, Dedelow JA, Chu BF, Zhao W, Yildiz VO, et al. CT gray-level texture analysis as a quantitative imaging biomarker of epidermal growth factor receptor mutation status in adenocarcinoma of the lung. *AJR Am J Roentgenol*. (2015) 205:1016–25. doi: 10.2214/AJR.14.14147
29. Xu F, Ma X, Wang Y, Tian Y, Tang W, Wang M, et al. CT texture analysis can be a potential tool to differentiate gastrointestinal stromal tumors without KIT exon 11 mutation. *Eur J Radiol*. (2018) 107:90–7. doi: 10.1016/j.ejrad.2018.07.025
30. Gierach GL, Li H, Loud JT, Greene MH, Chow CK, Lan L, et al. Relationships between computer-extracted mammographic texture pattern features and BRCA1/2 mutation status: a cross-sectional study. *Breast Cancer Res*. (2014) 16:424. doi: 10.1186/s13058-014-0424-8
31. Ditmer A, Zhang B, Shujaat T, Pavlina A, Luibrand N, Gaskill-Shipley M, et al. Diagnostic accuracy of MRI texture analysis for grading gliomas. *J Neurooncol*. (2018) 140:583–9. doi: 10.1007/s11060-018-2984-4
32. Skogen K, Schulz A, Dormagen JB, Ganeshan B, Helseth E, Server A. Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. *Eur J Radiol*. (2016) 85:824–9. doi: 10.1016/j.ejrad.2016.01.013
33. Nowosielski M, Galldiks N, Iglseder S, Kickingereder P, von Deimling A, Bendszus M, et al. Diagnostic challenges in meningioma. *Neuro Oncol*. (2017) 19:1588–98. doi: 10.1093/neuonc/nox101
34. Lin BJ, Chou KN, Kao HW, Lin C, Tsai WC, Feng SW, et al. Correlation between magnetic resonance imaging grading and pathological grading in meningioma. *J Neurosurg*. (2014) 121:1201–8. doi: 10.3171/2014.7.JNS132359

35. Yan PF, Yan L, Hu TT, Xiao DD, Zhang Z, Zhao HY, et al. The potential value of preoperative MRI texture and shape analysis in grading meningiomas: a preliminary investigation. *Transl Oncol*. (2017) 10:570–7. doi: 10.1016/j.tranon.2017.04.006
36. Dellacasa Bellingegni A, Gruppioni E, Colazzo G, Davalli A, Sacchetti R, Guglielmelli E, et al. NLR, MLP, SVM, and LDA: a comparative analysis on EMG data from people with trans-radial amputation. *J Neuroeng Rehabil*. (2017) 14:82. doi: 10.1186/s12984-017-0290-6
37. Scheme E, Englehart K. Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use. *J Rehabil Res Dev*. (2011) 48:643–59. doi: 10.1682/JRRD.2010.09.0177

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 11 December 2019
doi: 10.3389/fonc.2019.01393

Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data

Darcie A. P. Delzell[1]*, *Sara Magnuson*[1], *Tabitha Peter*[1], *Michelle Smith*[1] and *Brian J. Smith*[2]

[1] Department of Mathematics and Computer Science, Wheaton College, Wheaton, IL, United States, [2] Department of Biostatistics, University of Iowa, Iowa City, IA, United States

As awareness of the habits and risks associated with lung cancer has increased, so has the interest in promoting and improving upon lung cancer screening procedures. Recent research demonstrates the benefits of lung cancer screening; the National Lung Screening Trial (NLST) found as its primary result that preventative screening significantly decreases the death rate for patients battling lung cancer. However, it was also noted that the false positive rate was very high (>94%).In this work, we investigated the ability of various machine learning classifiers to accurately predict lung cancer nodule status while also considering the associated false positive rate. We utilized 416 quantitative imaging biomarkers taken from CT scans of lung nodules from 200 patients, where the nodules had been verified as cancerous or benign. These imaging biomarkers were created from both nodule and parenchymal tissue. A variety of linear, nonlinear, and ensemble predictive classifying models, along with several feature selection methods, were used to classify the binary outcome of malignant or benign status. Elastic net and support vector machine, combined with either a linear combination or correlation feature selection method, were some of the best-performing classifiers (average cross-validation AUC near 0.72 for these models), while random forest and bagged trees were the worst performing classifiers (AUC near 0.60). For the best performing models, the false positive rate was near 30%, notably lower than that reported in the NLST.The use of radiomic biomarkers with machine learning methods are a promising diagnostic tool for tumor classification. The have the potential to provide good classification and simultaneously reduce the false positive rate.

Keywords: radiomics, machine learning, CT image, biomarkers, lung cancer

OPEN ACCESS

Edited by:
Lei Deng,
Jacobi Medical Center, United States

Reviewed by:
Seyedmehdi Payabvash,
Yale School of Medicine, United States
Jongphil Kim,
Moffitt Cancer Center, United States
Jinhua Yu,
Fudan University, China
Hui Li,
University of Chicago, United States

***Correspondence:**
Darcie A. P. Delzell
darcie.delzell@wheaton.edu

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *30 April 2019*
Accepted: *26 November 2019*
Published: *11 December 2019*

Citation:
Delzell DAP, Magnuson S, Peter T, Smith M and Smith BJ (2019) Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data. Front. Oncol. 9:1393. doi: 10.3389/fonc.2019.01393

1. INTRODUCTION

Publication of primary results from the National Lung Screening Trial (NLST) reported that lung cancer screening, especially when performed with low dose computed tomography (CT) scans, can significantly reduce the mortality rate of lung cancer. This result highlights the benefits of lung cancer screening; however, the NLST also found that screening results had a notably high rate of false positive results. Of the total number of low dose CT scans in the NLST, the false

positive rate surpassed 94% (1). The NLST researchers noted that the high false positive rate was a challenge which required further research, and that challenge persists to the present. The negative consequences associated with false positive exam results can include patient anxiety and unnecessary invasive diagnostic procedures such as biopsy (2, 3).

High-throughput extraction of features from imaging data composes the essence of radiomics, an emerging field of research which offers significant improvement to decision-support in oncology (4, 5). Current work examines the predictive power of quantitative imaging biomarkers, which are quantitative features extracted from routine medical images (4, 6, 7), as inputs within predictive classifying models. The information contained in the imaging biomarkers has the potential to improve classification accuracy in a variety of statistical models (2).

Across the literature, quantitative biomarkers taken from imaging data have been used to develop models with the intent to identify and analyze associations between radiomic/nodule features (stages or histological characteristics) and clinical outcomes (survival, recurrence, etc.). Previous work in radiomics aimed at classification of lung nodules has examined a variety of outcomes (5, 8–12). Zhu et al. used outcome categories for lung cancer type with a LASSO classification model (13). Zhang et al. examined outcomes for local/distant failure using several machine learning classifiers (5). Pamar et al. used clusters of biomarkers as predictors in models of overall survival (14). Dilger et al. used an expanded set of radiomic features that included both nodule and parenchymal tissue. They showed an increase in classification performance when the parenchymal tissue was included in feature extraction (3).

In this paper, we investigate the predictive power of biomarkers (computed from both nodule and parenchymal tissue as calculated by Dilger et al. (3)) to classify lung nodule status as malignant/benign while also considering the false positive rate. Our comprehensive approach includes multiple combinations of models and filtering techniques. In particular, combinations of twelve machine learning classifiers along with six feature selection methods were compared, using area under the receiver operating characteristic curve (AUC) as the model performance metric.

2. METHODS

2.1. Dataset

This retrospective study analyzed data originally taken from 200 CT scans of the lungs of patients at the University of Iowa Hospital. Pathology and radiology reports were reviewed to identify an analysis set of patients who met eligibility criteria of having (a) a solitary lung nodule (5–30 mm) and (b) a malignant nodule confirmed on histopathology or a benign nodule confirmed on histopathology or by size stability for at least 24 months. Manual segmentations were performed by a graduate student trained in medical image analysis in order to define a region of interest (ROI) around each nodule. The ROIs were defined to include amounts of parenchyma approximately proportional to the nodule sizes. Individual ROI voxels were labeled as belonging to either the nodule or the parenchyma, with radiomic features calculated separately for each to produce the complete set of 416 (approximately half nodule and half parenchyma) quantitative imaging biomarkers. These biomarkers measured features such as intensity, shape, and texture of the ROI (15). This study is a secondary analysis of de-identified data originally collected with approval from the University of Iowa institutional review board. Demographic information can be found in **Table 1**.

TABLE 1 | Demographics of patient cohort.

	Malignant	Benign
Number of patients	110	90
Female	51 (46.4%)	63 (70.0%)
Male	59 (53.6%)	27 (30.0%)
Age, yrs (mean ± SD)	65.7 ± 11.2	58.2 ± 13.2
Pack-years (mean ± SD)	38.4 ± 31.2	11.2 ± 16.9
Nodule size, mm (range, mean ± SD)	7 – 44, 19.1 ± 6.3	6 – 30, 15.2 ± 5.8

A strength of the dataset is its fairly balanced malignant/benign status breakdown, with 45% of the cases malignant and 55% benign. Many machine learning-based classifying algorithms assume that the outcomes of a data set are balanced, but this assumption is not met when the proportion of outcomes is highly uneven. The data set used in this work has a nearly even ratio of malignant and benign nodules (16).

2.2. Radiomic Features

The 416 radiomic features which were available for this investigation quantified nodule characteristics from CT images acquired from a variety of scanner protocols through the University of Iowa Hospital. The most common CT models used were Siemens SOMATOM Definition, Siemens Sensation 16, Sensation Biograph 40, and Toshiba Aquilion. Using these machines, several protocols were used, including Chest CT scans with and without contrast, CT Angiography scans, Extrenal CT scans, PET/CT scans, and CT: Chest, Abdomen, and Pelvis scans. Slice thickness ranged from 1.0 to 6.0 mm with an average of 3.3 mm (15). From these scans, voxels labeled as parenchyma and nodule were used in the extraction of four classes of features: intensity, shape, border, and texture. The intensity of CT images described the radiodensity of the anatomy [measured using Hounsfield units (HU)] as well as heterogeneity of the nodule. Shape features examined sphericity and the maximum diameter of the nodule. Sphericity was computed by comparing the volume of the nodule to its surface area, and maximum diameter was measured using the Response Evaluation Criteria in Solid Tumors (RECIST). The border features were measured using a rubber band straightening transform (RBST). The texture features were extracted from the nodule and parenchyma regions using Laws' Texture Energy Measures (TEM). From these TEMs, the mean, variance, kurtosis, and skewness of the nodule and parenchyma were extracted. Radiomic features were extracted using a Matlab based CAD tool, and the mathematical definitions for all of the radiomic measurements are described in full in Dilger (17).

TABLE 2 | Summary of feature selection methods.

Feature selection method	Abbreviation
Linear combination	lincom
Pairwise correlation	corr.95
PCA - 0.85 cutoff	pca.85
PCA - 0.90 cutoff	pca.90
PCA - 0.95 cutoff	pca.95
Unfiltered	nofilter

2.3. Feature Selection Methods

As is common in radiomics studies with hundreds of features, many of the biomarkers (features) used as predictors were highly correlated with one another; this challenge necessitated feature selection in order to avoid collinearity, reduce dimensionality, and minimize noise (11, 16, 18, 19). To this end, we considered three feature selection methods: a linear combinations filter, a pairwise correlation filter, and principle component analysis.

For the linear combinations filter (lincom), a QR decomposition along with an iterative procedure is used to determine if some predictors are linear combinations of others. Predictors are sequentially removed until the design matrix is full rank. The pairwise correlation filter removes those predictors whose pairwise correlation is greater than a specified cutoff. The two predictors with the largest absolute correlation are first considered. Of those two, the predictor with the highest average absolute correlation with all other variables is removed. This process continues until all the predictors left have pairwise absolute correlations less than the cutoff. After investigating multiple cutoffs, we chose a cutoff value of 0.95 for the pairwise correlation filter (corr.95) since this cutoff removed highly correlated variables but still retained a large number of features. Principal component analysis reduces dimensionality by creating new, uncorrelated predictors which explain a large proportion of the variance in the predictor space. Principal component analysis was implemented at three different cutoffs (pca.85, pca.90, pca.95), where the number of components accounted for either 85, 90, or 95% of the variance in the predictor space (**Table 2**).

2.4. Classifiers and Performance Metrics

Combinations of the six feature selection methods and twelve classifiers were investigated by implementing a 10-fold repeated cross-validation framework with five repeats, a standard validation technique (5, 13, 16, 20, 21). The feature selection methods were included in the cross-validation algorithm so that their contribution to the final model fit is reflected in the performance metrics. The classifiers are from three different families: linear, nonlinear, and ensemble (22). Of the linear classifiers, an elastic net (elasticnet), a logistic regression (logistic), a partial least squares model (pls), and a logistic regression with Step AIC were fit. The nonlinear classifiers include a K-nearest neighbors model (knn), a neural network (nnet), and three support vector machines: a linear kernel (svml), a polynomial kernel (svmpoly), and a radial kernel (svmr). The ensemble models used included bagged classification trees (bag), random forest (rf), and stochastic gradient boosting (gbm) (**Table 3**).

TABLE 3 | Summary of classifiers.

Model family	Classifier	Abbreviation
Linear	Elastic net	elasticnet
	Logistic regression	logistic
	Partial least squares	pls
	Logistic regression with step AIC	glmStepAIC
Nonlinear	K-nearest neighbors	knn
	Neural network	nnet
	Support vector machine (linear kernel)	svml
	SVM (polynomial kernel)	svmpoly
	SVM (radial kernel)	svmr
Ensemble	Bagged trees	bag
	Random forest	rf
	Stochastic gradient boosting	gbm

The quality of model performance in most machine learning algorithms is dependent upon the choice of various tuning parameters. Some tuning parameters take into account the number of predictors after feature selection. For example, the mtry tuning parameter for rf, which determines the number of candidate variables at each branch, is equal to the square root of the number of predictors. Other tuning parameters were chosen based on standard practice (22, 23). For example, the decay tuning parameter for nnet, which helps prevent overfitting, generally takes the values of 0.1, 0.01, and 0.001. All models were fit using the caret R package (24). Our R code implementing the feature selection and classification models is presented as **Supplementary Material**.

3. RESULTS

The linear combinations filter removed 217 biomarkers, leaving a set of 199 predictors. The pairwise correlation filter retained 39 predictors, while principal components analysis retained 12, 14, and 18 components at the 85, 90, and 95% levels, respectively.

Figure 1 gives the predictive performance (AUC) of each feature selection method (in rows) and classifier (in columns), averaged over the 50-folds/repeats in the cross-validation. Logistic regression models cannot be calculated when the number of predictors is larger than the number of observations, so the nofilter row is blank for this classifier. The large number of predictors also caused multiple computing issues with the neural net classifier, so training this classifier without using any feature selection was not considered. **Table 4** gives the highest average AUC for each classifier across the various feature selection methods. Principal component analysis yields lower AUC values for all of the classifying models. Using lincom, the top four classification methods perform well, with AUC $\geq$ 0.728 (we note that svmr with corr.95 also has an average AUC = 0.728). The standard deviation over the folds/repeats is also given, along with sensitivity, specificity, and false positive rate statistics. Specificity and sensitivity were computed using a 0.5 threshold from the

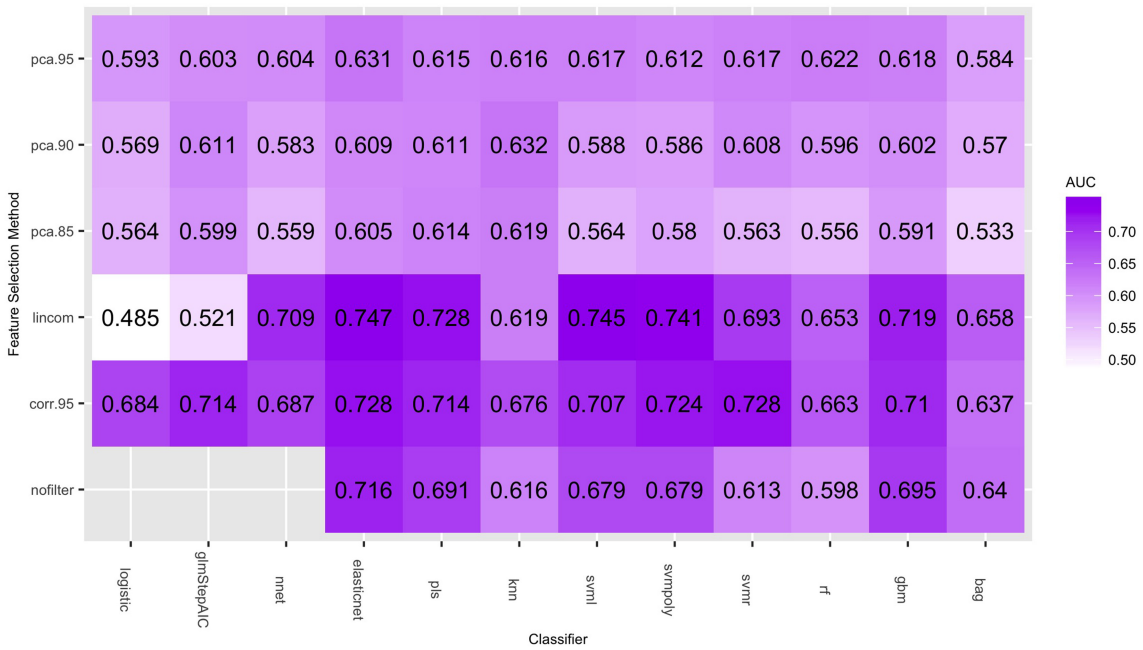


FIGURE 1 | Average AUC values (over the 50 repeated cross-validation testing sets) of each feature selection/classifier combination.

TABLE 4 | AUC values for classifiers with highest predictive performance (SD taken over the 50 cross-validation testing sets).

Classifier	Feature method	AUC	SD	Sensitivity	Specificity	False positive rate	SD
elasticnet	lincom	0.747	0.111	0.616	0.729	0.271	0.136
svml	lincom	0.745	0.112	0.549	0.765	0.235	0.126
svmpoly	lincom	0.741	0.113	0.569	0.781	0.219	0.132
pls	lincom	0.728	0.111	0.627	0.707	0.293	0.126
svmr	corr.95	0.728	0.106	0.542	0.780	0.220	0.148
gbm	lincom	0.714	0.106	0.596	0.733	0.267	0.140
glmStepAIC	corr.95	0.714	0.110	0.636	0.684	0.316	0.130
nnet	lincom	0.709	0.113	0.620	0.707	0.293	0.143
logistic	corr.95	0.684	0.108	0.600	0.689	0.311	0.116
knn	corr.95	0.676	0.109	0.482	0.738	0.262	0.117
rf	corr.95	0.663	0.124	0.473	0.730	0.270	0.127
bag	lincom	0.658	0.106	0.529	0.702	0.298	0.146

model predicted class probabilities. The AUC standard deviations are fairly similar, while sensitivity and specificity have larger variation. The false positive rates are more variable than the AUC values, and the mean false positive rates are all notably lower (all less than 32%) than the 94% found in the results of the NLST.

Figure 2 shows the distribution of the AUC scores for the four best performing classifiers: elasticnet, svml, svmpoly, and pls. Among all feature selection methods, corr.95 and lincom yielded the highest AUC values on average across these four classifiers. The lincom feature selection with the elasticnet classifier has the best overall predictive performance (AUC = 0.747), followed by the svml classifier with the lincom feature selection (AUC = 0.745). As has been observed in other radiomic studies, support vector machines perform well with respect to predictive performance (21).

The boxplots in **Figure 3** show the distribution of the false positive rates for the four best performing classifiers. These distributions show that the lowest false positive rates were achieved in combination with either the lincom or corr.95 feature selection methods for all four of these classifiers. These two feature selection methods result in both the highest average AUC values and the lowest false positive rates.

Figure 4 gives the ROC curve for the best performing classifier/feature selection combination (elasticnet/lincom). Although the NLST did not report false negative rates, the ROC curve displays the tradeoff between specificity and sensitivity. While the classifiers have reduced the false positive rate, the tradeoff is an increase in the false negative rate, which would be estimated to be near 0.38 for this particular classifier. This natural tradeoff between specificity and sensitivity for classifiers would suggest that radiomic methods should not be the sole diagnostic tool in lung cancer diagnosis. However, the reduction of the false positive rate for a non-invasive procedure is a substantial improvement and supports the inclusion of these methods in clinical practice.

4. DISCUSSION

While awareness of the benefits of preventative screening for lung cancer has increased in recent years, there is still a need

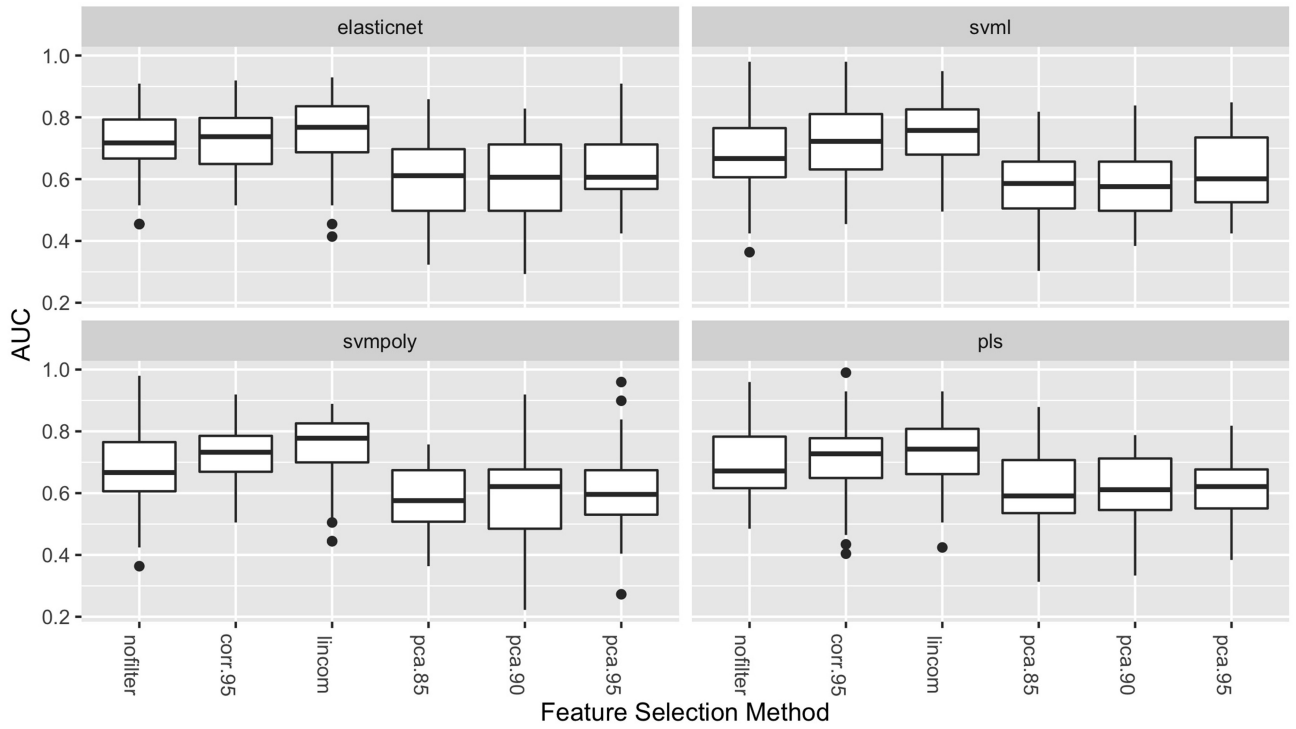


FIGURE 2 | Boxplots of AUC values (over the 50 repeated cross-validation testing sets) for each feature selection method for the four best-performing classifiers.

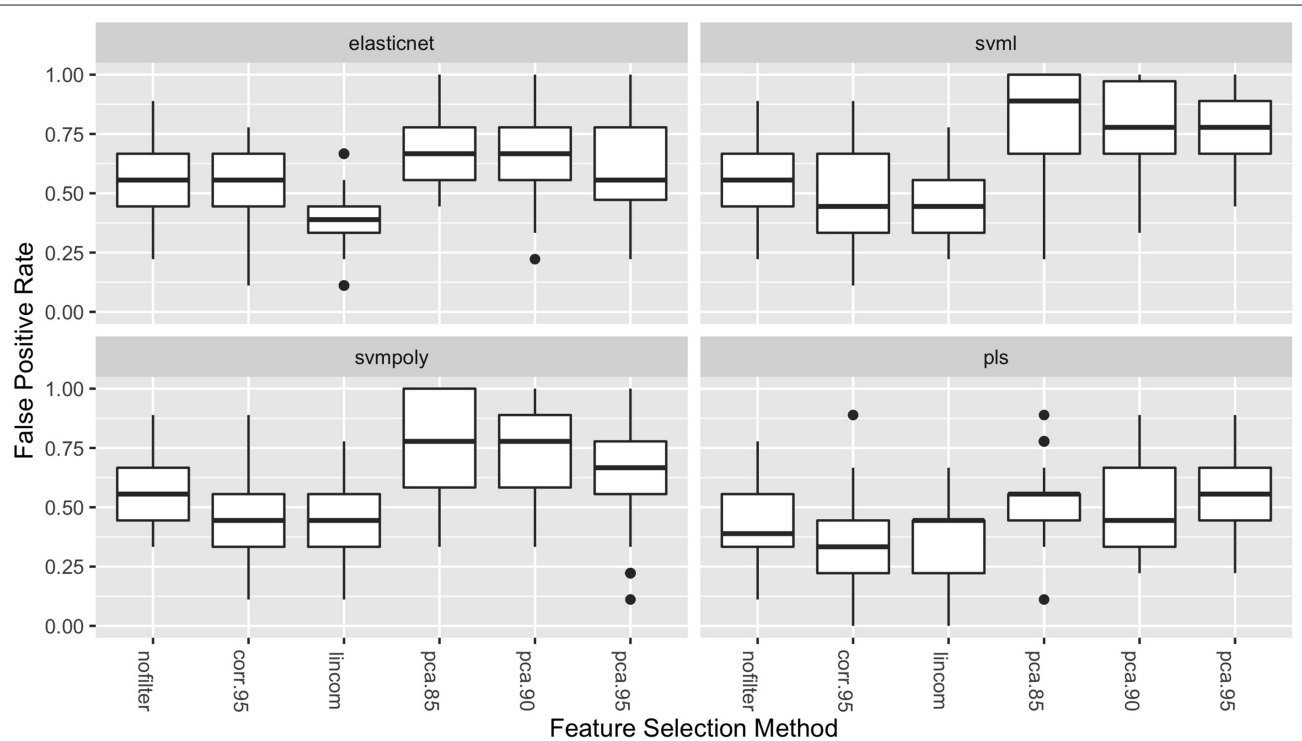


FIGURE 3 | Boxplots of the false positive rates (over the 50 repeated cross-validation testing sets) for each feature selection method for the four best-performing classifiers.

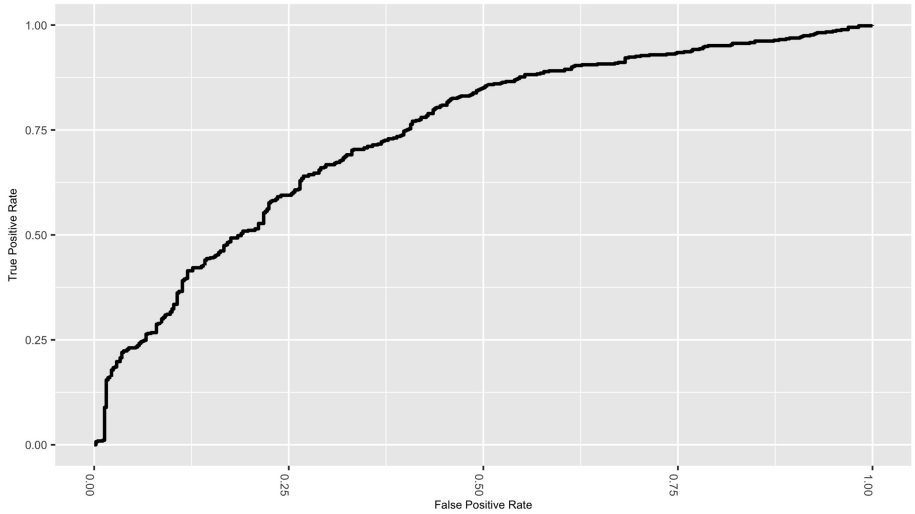


FIGURE 4 | ROC curve for the elastic net classifier with the linear combinations filter.

for improved accuracy in nodule classification. Moreover, a high false positive rate for the diagnostic outcome of lung cancer screening remains a major challenge. Nodule characteristics (biomarkers) calculated from CT scans offer the possibility of improved nodule classification through various modeling techniques. Machine learning algorithms have the potential to harness the predictive power in nodule characteristics. However, little work has been done to compare the performance of various machine learning methods used in conjunction with different feature selection methods, especially as they relate to lung cancer tumor diagnosis.

However, models to predict pulmonary nodule status have been developed and evaluated in other studies. Chen et al. extracted 750 imaging features and compared the performance of a support vector machine (SVM) trained with all to an SVM trained with a sequential forward selection of 4 features (2). Leave-one-out cross-validation demonstrated superior accuracy of 84% for the 4-feature model vs. 56% for all features. Alahmari et al. studied the prognostic performance of radiomics features and found the addition of feature changes over time (delta radiomics) to improve AUC performance from 0.773 to 0.822 (25). SVM and random forest models as well as different feature selection algorithms were considered in their analysis. Final results are presented for random forest models and ReliefF feature selection, suggesting that these were the optimal choices, although comparisons to the others were not presented. A computer-aided lung nodule detection system was proposed by Ma et al. (26). In their approach, multiscale nodule and vessel enhancement filters were applied to patient images prior to extracting 979 radiomics features for training of a random forest classifier. Comparisons to other modeling approaches were not made. Uthoff et al. used a set of 922 radiomics features that is an extension of ours with both nodule features and parenchyma features calculated in 25, 50, 75, and 100% bands around the maximal in-plane diameter of the nodule (27). They used k-medoids clustering to select features for training of an artificial neural network. K-medoids feature selection is similar in spirit to the high correlation selection approach we used in that both reduce the number of features by selecting representative ones from those that are similar. Comparisons to other modeling approaches are not presented in their publication.

In this study, we considered the ability of nodule biomarkers to accurately predict malignant/benign status. The elastic net, support vector machines with polynomial and linear kernels, and partial least squares were the most predictive classifiers. When combined with the linear combination and correlation feature selection methods, these four classifiers had AUC values comparable in accuracy to the most predictive models studied in previous radiomic analyses (14, 16, 21). Furthermore, we observed that these classifiers greatly reduced the false positive rate from that given in the NLST results.

The observations from this investigation suggest that classifiers such as support vector machines and elastic net perform well with quantitative imaging biomarkers as their predictors. We also show that the chosen feature selection method will impact model performance, and we recommend using linear combination or a correlation-based reduction method over principal components. Different CT modalities and/or different patient population characteristics may yield different results. In order to recommend a particular model for application in a clinical setting, these results would need to be externally validated.

As as comparison, the two best classifier/feature selection combinations were fit with both the 416 biomarkers, as well as the demographic variables of sex, age, and pack-years (the number of packs smoked per day multiplied by the number of years smoked). Elastic Net with the Linear Combination filter had an average AUC of 0.747 (see **Table 4**) without the demographic variables included. This number was increased to 0.854 when these variables were added. The Linear Support Vector Machine with the Linear Combination filter had an average AUC of

0.745 without the demographic variables included. This number was increased to 0.820 when these variables were added. This suggests that radiomic features, while having good predictive performance, can be enhanced when other patient characteristics are included in the model.

Taken together, a number of common themes emerge from our present work and the past work of others. First, methods that reduce the number of features prior to model training appear to improve predictive performance. We believe this is especially true in the field of radiomics where large numbers of features tend to be highly correlated. Oftentimes, there are many features that do not provide additional information because they are linear combinations of others and may be removed with a linear combination filter. In addition, radiomics features tend to exhibit strong clustering for which high correlation or k-medoid selection seems to improve prediction even when in the cases of models, like random forests and gradient boosting, that perform automatic feature selection. Second, our work suggests that SVM performs well in the radiomics setting and supports its use by others. Furthermore, we found the commonly used random forest model to have poor performance; whereas, the less commonly used in radiomics—but commonly used in genomics—elastic net model was our top performer. Thus, we encourage consideration and reporting of more than one modeling approach in radiomics research. Finally, there is strong evidence that pulmonary features derived from the parenchyma and that reflect changes over time help with prediction. Likewise, as is the case in many fields, improvements in prediction are often achieved when utilizing subject matter expertise in the development of features and modeling approaches.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

Authors acknowledge financial support from the National Institute of Health (NIH R25HL131467) and the National Cancer Institute (NCI P30CA086862). This research was also supported by the G. W. Aldeen Fund at Wheaton College.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01393/full#supplementary-material

REFERENCES

1. The National Lung Screening Trial Research Team. Reduced lung-cancer mortality rate with low-dose computed tomographic screening. *N Engl J Med*. (2011) 365:395–409. doi: 10.1056/NEJMoa1102873
2. Chen CH, Chang CK, Tu CY, Liao WC, Wu BR, Chou KT, et al. Radiomic features analysis in computed tomography images of lung nodule classification. *PLoS ONE*. (2018) 13:e0192002. doi: 10.1371/journal.pone.0192002
3. Dilger SK, Uthoff J, Judisch Aea. Improved pulmonary nodule classification utilizing quantitative lung parenchyma features. *J Med Imaging*. (2015) 2:041004. doi: 10.1117/1.JMI.2.4.041004
4. Gillies R, Kinahan P, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169
5. Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett*. (2017) 403:21–7. doi: 10.1016/j.canlet.2017.06.004
6. Kumar V, Gu Y, Basu Sea. Radiomics: the process and the challenges . *Magn Reson Imaging*. (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
7. Ortiz-Ramón R, Larroza A, Ruiz-España S, Arana E, Moratal D. Classifying brain metastases by their primary site of origin using a radiomics approach based on texture analysis: a feasibility study. *Eur Radiol*. (2018) 28:4514–23. doi: 10.1007/s00330-018-5463-6
8. Kuruvilla J, Gunavathi K. Lung cancer classification using neural networks for CT images. *Comput Methods Prog Biomed*. (2014) 113:202–9. doi: 10.1016/j.cmpb.2013.10.011
9. Lambin P, Rios-Velazquez E, Leijenaar Rea. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
10. Lin Y, Leng Q, Jiang Z, Guarnera MA, Zhou Y, Chen X, et al. A classifier integrating plasma biomarkers and radiological characteristics for distinguishing malignant from benign pulmonary nodules. *Int J Cancer*. (2017) 141:1240–8. doi: 10.1002/ijc.30822
11. Parmar C, Leijenaar RTH, Grossmann P, Velazquez ER, Bussink J, Rietveld D, et al. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head &neck cancer. *Sci Rep*. (2015) 5:1–10. doi: 10.1038/srep11044
12. Parmar C, Grossmann P, Rietveld Dea. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *Front Oncol*. (2015) 5:272. doi: 10.3389/fonc.2015.00272
13. Zhu X, Dong D, Chen Z, Fang M, Zhang L, Song J, et al. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *Eur Radiol*. (2018) 28:2772–8. doi: 10.1007/s00330-017-5221-1
14. Parmar C, Grossmann P, Bussink Jea. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep*. (2015) 5:13087. doi: 10.1038/srep13087
15. Dilger SKN. *Pushing the Boundaries: Feature Extraction From the Lung Improves Pulmonary Nodule Classification*. Iowa City, IA: University of Iowa (2016).
16. Zhang Y, Oikonomou A, Wong Aea. Radiomics-based prognosis analysis for non small cell lung cancer. *Sci Rep*. (2017) 7:46349. doi: 10.1038/srep46349
17. Dilger SKN. *The Use of Surrounding Lung Parenchyma for the Automated Classification of Pulmonary Nodules*. Iowa City, IA: University of Iowa (2013).
18. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non small cell lung cancer. *Radiology*. (2016) 281:947–57. doi: 10.1148/radiol.2016152234
19. Lambin P, Leijenaar R, Deist Tea. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. (2017) 14:749. doi: 10.1038/nrclinonc.2017.141
20. Krafft SP, Briere TM, Court LE, Martel MK. The utility of quantitative ct radiomics features for improved prediction of radiation pneumonitis. *Med Phys*. (2018) 45:5317–24. doi: 10.1002/mp.13150
21. Sun T, Wang J, Li Xea. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. *Comput Methods Prog Biomed*. (2013) 111:519–24. doi: 10.1016/j.cmpb.2013.04.016

22. Kuhn M, Johnson K. *Applied Predictive Modeling*. New York, NY: Springer (2013).
23. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw Articles*. (2008) 28:1–26. doi: 10.18637/jss.v028.i05
24. Kuhn M, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, et al. *caret: Classification and Regression Training*. R package version 6.0-80 (2018). Available online at: https://CRAN.R-project.org/package=caret
25. Alahmari SS, Cherezov D, Goldgof DB, Hall LO, Gillies RJ, Schabath MB. Delta radiomics improves pulmonary nodule malignancy prediction in lung cancer screening. *IEEE Access*. (2018) 6:77796–806. doi: 10.1109/ACCESS.2018.2884126
26. Ma J, Zhou Z, Ren Y, Xiong J, Fu L, Wang Q, et al. Computerized detection of lung nodules through radiomics. *Med Phys*. (2017) 44:4148–58. doi: 10.1002/mp.12331
27. Uthoff J, Stephens MJ, Newell Jr JD, Hoffman EA, Larson J, Koehn N, et al. Machine learning approach for distinguishing malignant and benign lung nodules utilizing standardized perinodular parenchymal features from CT. *Med Phys*. (2019) 46:3207–16. doi: 10.1002/mp.13592

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

CORRECTION
published: 05 June 2020
doi: 10.3389/fonc.2020.00866

Corrigendum: Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data

Darcie A. P. Delzell[1]*, *Sara Magnuson*[1], *Tabitha Peter*[1], *Michelle Smith*[1] and *Brian J. Smith*[2]

[1] Department of Mathematics and Computer Science, Wheaton College, Wheaton, IL, United States, [2] Department of Biostatistics, University of Iowa, Iowa City, IA, United States

Keywords: radiomics, machine learning, CT image, biomarkers, lung cancer

A Corrigendum on

Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data

by Delzell, D. A. P., Magnuson, S., Peter, T., Smith, M., and Smith, B. J. (2019). Front. Oncol. 9:1393. doi: 10.3389/fonc.2019.01393

OPEN ACCESS

Edited and reviewed by:
Rong Tian,
Sichuan University, China

***Correspondence:**
Darcie A. P. Delzell
darcie.delzell@wheaton.edu

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *03 April 2020*
Accepted: *01 May 2020*
Published: *05 June 2020*

Citation:
Delzell DAP, Magnuson S, Peter T, Smith M and Smith BJ (2020) Corrigendum: Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data. Front. Oncol. 10:866. doi: 10.3389/fonc.2020.00866

The data analyzed for this study were generated by Samantha Dilger, Ph.D and Jessica Sieren, Ph.D (Departments of Radiology and Biomedical Engineering, University of Iowa, Iowa City, IA, United States) who control the rights to the data and do not intend for the data to be shared publicly. Accordingly, this data which was included as Supplementary Material in the original article is being removed. In addition, the data were taken from a mix of low and high-dose CT scans, which were incorrectly referred to in the original article as low-dose scans.

The corrections below have been made to the **Methods**, subsection **Dataset**, paragraph 1.

"This retrospective study analyzed data originally taken from 200 CT scans of the lungs of patients at the University of Iowa Hospital. Pathology and radiology reports were reviewed to identify an analysis set of patients who met eligibility criteria of having (a) a solitary lung nodule (5–30 mm) and (b) a malignant nodule confirmed on histopathology or a benign nodule confirmed on histopathology or by size stability for at least 24 months. Manual segmentations were performed by a graduate student trained in medical image analysis in order to define a region of interest (ROI) around each nodule. The ROIs were defined to include amounts of parenchyma approximately proportional to the nodule sizes. Individual ROI voxels were labeled as belonging to either the nodule or the parenchyma, with radiomic features calculated separately for each to produce the complete set of 416 (approximately half nodule and half parenchyma) quantitative imaging biomarkers. These biomarkers measured features such as intensity, shape, and texture of the ROI (15). This study is a secondary analysis of de-identified data originally collected with approval from the University of Iowa institutional review board. Demographic information can be found in Table 1."

The dataset has been removed from the online Supplementary Material and replaced with R code implementing the feature selection and classification models described in Methods Sections 2.3 and 2.4 of the article. The **Methods** section, subsection **Classifiers and Performance Metrics**, paragraph 2 has been updated to include a reference to the supplementary code as follows:

"The quality of model performance in most machine learning algorithms is dependent upon the choice of various tuning parameters. Some tuning parameters take into account the number of predictors after feature selection. For example, the mtry tuning parameter for rf, which determines the number of candidate variables at each branch, is equal to the square root of the number of predictors. Other tuning parameters were chosen based on standard practice (22, 23). For example, the decay tuning parameter for nnet, which helps prevent overfitting, generally takes the values of 0.1, 0.01, and 0.001. All models were fit using the caret R package (24). Our R code implementing the feature selection and classification models is presented as Supplementary Material."

The authors apologize for the inclusion of the data in the Supplementary Material and misstatement of "low-dose" CT. We state that these do not change the scientific conclusions of the article in any way. The original article has been updated.

frontiers in Oncology

ORIGINAL RESEARCH
published: 17 December 2019
doi: 10.3389/fonc.2019.01371

Ability of Radiomics in Differentiation of Anaplastic Oligodendroglioma From Atypical Low-Grade Oligodendroglioma Using Machine-Learning Approach

Yang Zhang[1,2†], Chaoyue Chen[1†], Yangfan Cheng[2], Yuen Teng[2], Wen Guo[2], Hui Xu[3], Xuejin Ou[2], Jian Wang[4], Hui Li[2], Xuelei Ma[5,6] and Jianguo Xu[1*]*

[1] Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu, China, [2] West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China, [3] Radiology Department, West China Hospital, Sichuan University, Chengdu, China, [4] School of Computer Science, Nanjing University of Science and Technology, Nanjing, China, [5] Department of Biotherapy, Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [6] State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University and Collaborative Innovation Center for Biotherapy, Chengdu, China

Objectives: To investigate the ability of radiomics features from MRI in differentiating anaplastic oligodendroglioma (AO) from atypical low-grade oligodendroglioma using machine-learning algorithms.

Methods: A total number of 101 qualified patients (50 participants with AO and 51 with atypical low-grade oligodendroglioma) were enrolled in this retrospective, single-center study. Forty radiomics features of tumor images derived from six matrices were extracted from contrast-enhanced T1-weighted (T1C) images and fluid-attenuation inversion recovery (FLAIR) images. Three selection methods were performed to select the optimal features for classifiers, including distance correlation, least absolute shrinkage and selection operator (LASSO), and gradient boosting decision tree (GBDT). Then three machine-learning classifiers were adopted to generate discriminative models, including linear discriminant analysis, support vector machine, and random forest (RF). Receiver operating characteristic analysis was conducted to evaluate the discriminative performance of each model.

Results: Nine predictive models were established based on radiomics features from T1C images and FLAIR images. All of the classifiers represented feasible ability in differentiation, with AUC more than 0.840 when combined with suitable selection method. For models based on T1C images, the combination of LASSO and RF classifier represented the highest AUC of 0.904 in the validation group. For models based on FLAIR images, the combination of GBDT and RF classifier showed the highest AUC of 0.861 in the validation group.

Conclusion: Radiomics-based machine-learning approach could potentially serve as a feasible method in distinguishing AO from atypical low-grade oligodendroglioma.

Keywords: radiomics, machine learning, oligodendroglioma, anaplastic oligodendroglioma, magnetic resonance imaging, grading

OPEN ACCESS

Edited by:
Meiyappan Solaiyappan, Johns Hopkins University, United States

Reviewed by:
Xudong Shen, Guizhou Medical University (GMU), China
Jiawen Zhang, Huashan Hospital, Fudan University, China
Juan Antonio Hernandez-Tamames, Erasmus Medical Center, Netherlands

***Correspondence:**
Xuelei Ma
drmaxuelei@gmail.com
Jianguo Xu
drjianguoxu@gmail.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *16 June 2019*
Accepted: *21 November 2019*
Published: *17 December 2019*

Citation:
Zhang Y, Chen C, Cheng Y, Teng Y, Guo W, Xu H, Ou X, Wang J, Li H, Ma X and Xu J (2019) Ability of Radiomics in Differentiation of Anaplastic Oligodendroglioma From Atypical Low-Grade Oligodendroglioma Using Machine-Learning Approach. Front. Oncol. 9:1371. doi: 10.3389/fonc.2019.01371

INTRODUCTION

Oligodendroglial tumors, one of the most common subtypes of gliomas, are classified into oligodendroglioma (grade II) and anaplastic oligodendroglioma (AO) (grade III) according to the 2016 World Health Organization (WHO) classification system (1). The clinical management and prognosis of oligodendrogliomas are closely relevant to the histopathological grade. AO is considered as the malignant tumor with aggressive behavior and requires radiotherapy and chemotherapy after the maximum safe resection, whereas patients with low-grade oligodendroglioma usually undergo less postoperative treatment and have better survival outcomes (2, 3). Therefore, the accurate preoperative assessment of tumor grade is clinically important for treatment planning and prognosis prediction. Magnetic resonance (MR) scan is recommended in pre-surgical evaluation of oligodendroglioma grade, as the contrast enhancement pattern is typically considered as the characteristics of high-grade glioma (2, 4). However, up to 50% of low-grade oligodendroglioma showed similar patterns with enhancement on MR imaging (MRI), making the discrimination from AO challenging in these cases (5).

Radiomics is an emerging field that can extract quantitative parameters from medical images to provide non-visual information calculated with mathematical formulas (6). Previous studies suggested that the combination of radiomics and machine-learning algorithms showed promising potential in differential diagnosis, pre-surgical grading, and prognosis prediction of intracranial tumors (7–10). However, it has never been applied in the grade prediction of oligodendrogliomas. Because radiomics could potentially reflect the underlying pathophysiology of lesions, we hypothesized that it might detect the differences that were difficult to obtain by visual inspection between AO and atypical oligodendroglioma (6, 11). Therefore, the purpose of the present study was to investigate the ability of radiomics-based machine learning technology in distinguishing AO from atypical low-grade oligodendroglioma. A set of radiomics parameters was extracted from MR images, and a series of discriminative models were established using different combinations of selection methods and machine-learning algorithms.

MATERIALS AND METHODS

Patient Selection

In this retrospective study, we screened our institutional database to review the patients who were diagnosed and treated at the neurosurgery department of our institution from January 2015 to December 2018. According to the 2016 WHO Classification of Tumors of the Central Nervous System, the presence of isocitrate dehydrogenase (IDH) mutation and 1p/19q codeletion is necessary for diagnosis of both oligodendroglioma and AO. Therefore, we carefully viewed the pathological reports and genetic testing results of all participants, ensuring that enrolled patients histopathologically and genetically met the 2016 WHO criteria. We initially selected 241 potentially eligible patients who were: (1) with pathological confirmation of oligodendroglioma ($N = 182$) or AO ($N = 59$); (2) with conclusive genetic testing results (presence of IDH mutation and 1p/19q codeletion); (3) with pre-therapeutic MR images. Among 182 patients with low-grade oligodendroglioma, 68 of them were selected as atypical oligodendroglioma defined as a low-grade oligodendroglioma with enhancement patterns on MRI. The exclusion criteria were as follows: (1) incomplete medical records ($N = 11$); (2) recorded history of receiving radiosurgery, chemotherapy, or radiotherapy before MR scans ($N = 9$); (3) previous history of any other cerebral diseases, such as stroke, subarachnoid hemorrhage ($N = 6$). The process of patient enrollment was shown in **Figure 1**. The clinical parameters, such as gender, age, Ki-67 labeling index of tumor, and days between MR scan and surgery were also recorded. This study was approved by the Ethics Committee of Sichuan University. The written informed consent was obtained from all participants enrolled in this study (written informed consent for patients under the age of 16 was obtained from parents or guardians).

MRI Acquisition

All patients enrolled took MR scan *via* 3.0T GE SIGNA MRI scanner in our institution. In this study, contrast-enhanced T1-weighted (T1C) and fluid-attenuation inversion recovery (FLAIR) images were selected to perform texture analysis for the following reasons: first, they were the most important sequences in the diagnosis of oligodendrogliomas; second, the boundary of tumor and normal brain tissue should be clear and recognizable on images for precise delineation (**Figure 2**). The parameters of T1C image were as follows: TR/TE = 1,540/2.4 ms, slice thickness = 1 mm, axial FOV = 24 × 24 cm^2, and data matrix = 256 × 256. The parameters of FLAIR image were as follows: TR/TE = 4,000/393 ms, slice thickness = 1 mm, axial FOV = 24 × 24 cm^2, and data matrix = 516 × 516. Gadopentetate dimeglumine (0.1 mmol/kg) was used as the contrast agent for T1C sequence. MR images of all participants were collected with uniform standards through Picture Archiving and Communication Systems from our institutional radiology department.

Texture Features Extraction

Texture features were extracted from MR images by two researchers together under the guidance of senior radiologists using LIFEx software (http://www.lifexsoft.org) (12). Following the instructions of the software, we manually contoured the regions of interest (ROI) on axial image slice by slice (obvious cystic area was not included in ROI considering the interference of cystic fluid). Disagreements between researchers on tumor boundary were addressed by consulting the senior radiologists. The edema band and adjacent structure invasion were carefully separated from the tumor tissue through the difference in contrast enhancement patterns in T1C images. Anatomic structures around the tumor were also recorded to help with delineation in FLAIR images. To ensure the accuracy of texture parameters, ROI was only drawn on the biggest one for tumors with clear boundary and on tumor-confirmed area for tumors with vague boundary. Even following this strategy, 12 FLAIR

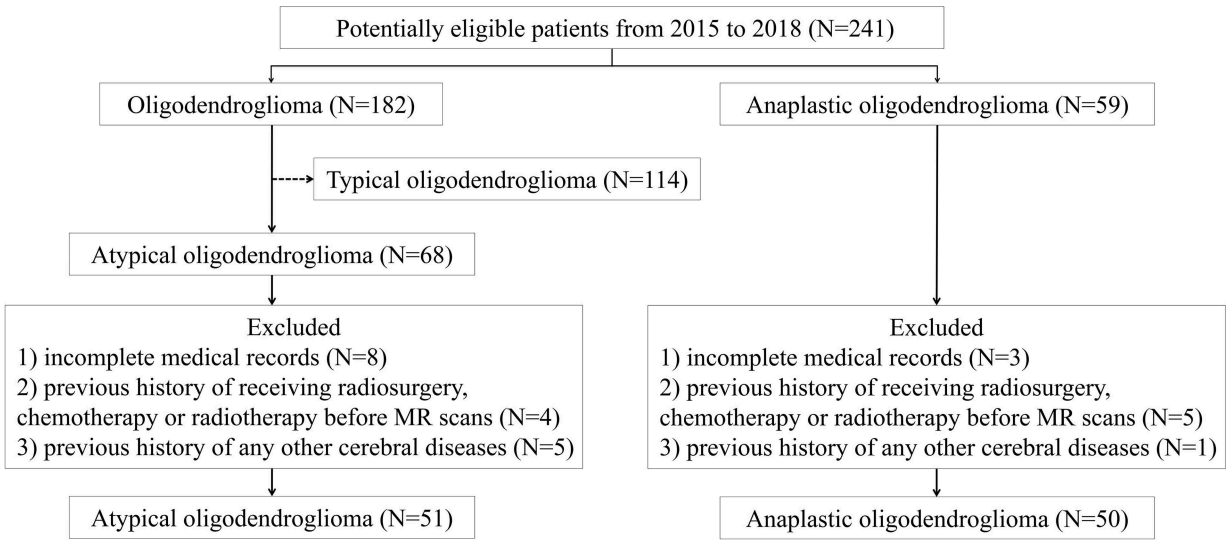


FIGURE 1 | The flowchart of patient enrollment process. MR, magnetic resonance.

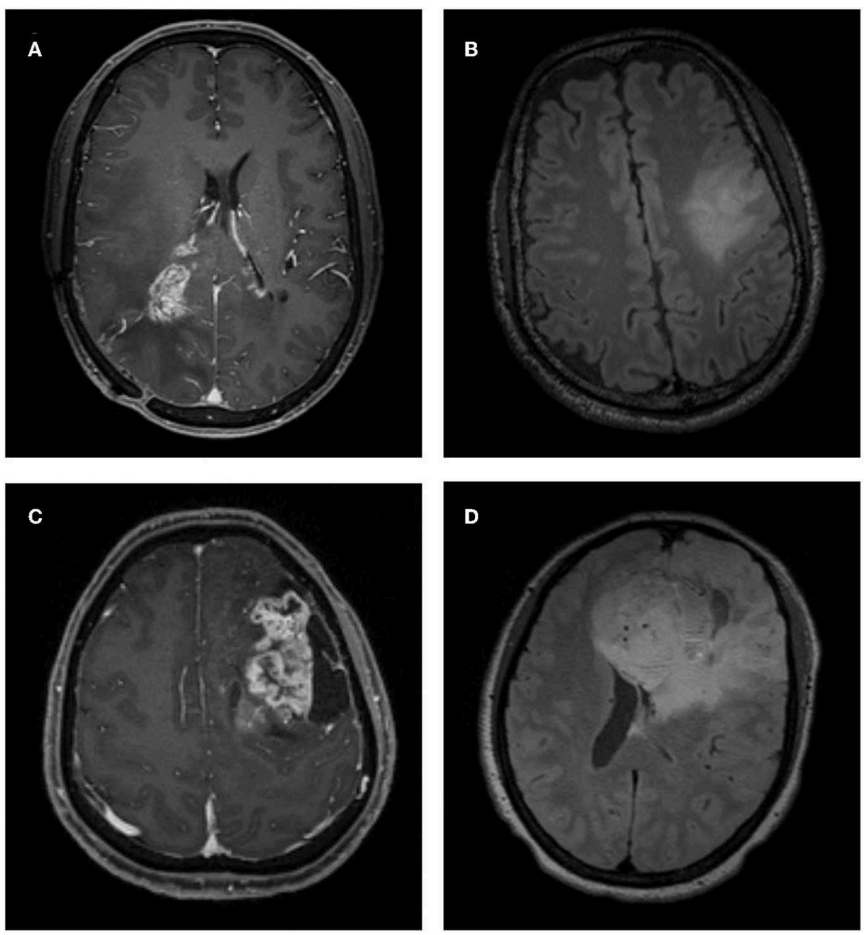


FIGURE 2 | Examples of atypical low-grade oligodendroglioma and anaplastic oligodendroglioma on MRI. **(A)** A patient with atypical low-grade oligodendroglioma in contrast-enhanced T1-weighted (T1C) image. **(B)** A patient with atypical low-grade oligodendroglioma in fluid-attenuation inversion recovery (FLAIR) image. **(C)** A patient with anaplastic oligodendroglioma in T1C image. **(D)** A patient with anaplastic oligodendroglioma in FLAIR image.

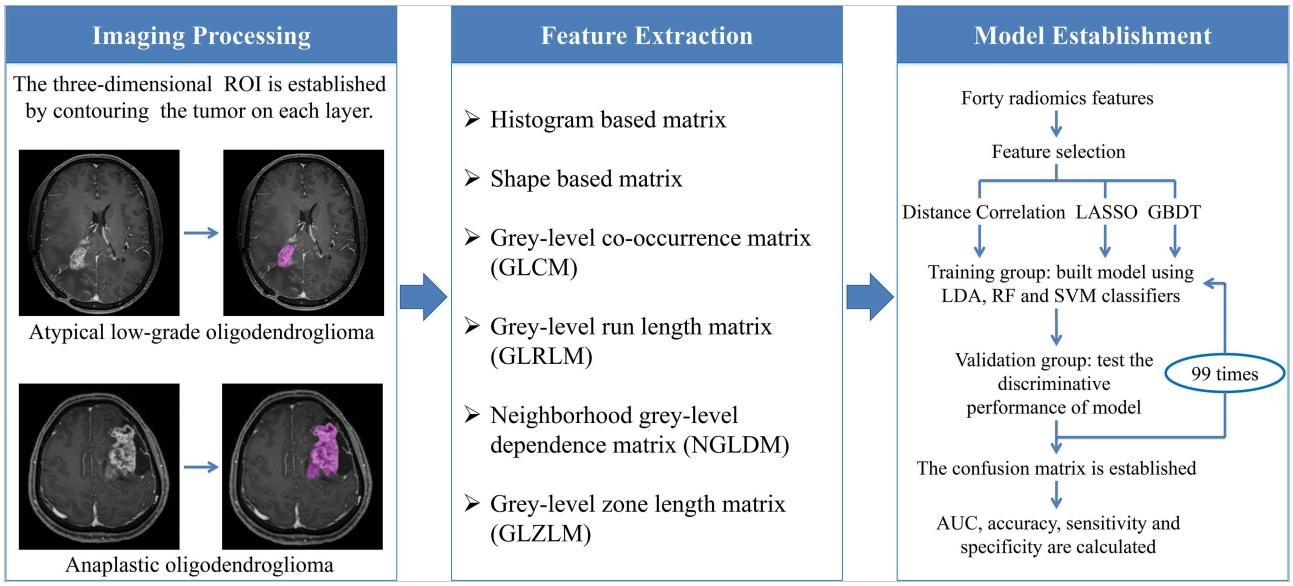


FIGURE 3 | The workflow chart from image processing to model establishment. ROI, regions of interest; LASSO, least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree; LDA, linear discriminant analysis; RF, random forest; SVM, support vector machine; AUC, area under the curve.

images were excluded because we were unable to delineate the tumor due to the interference of edema.

A total of 40 texture features were extracted from six matrices in the first or the second orders, including Histogram-based matrix, Shape-based matrix, Gray-level co-occurrence matrix (GLCM), Gray-level run length matrix (GLRLM), neighborhood gray-level dependence matrix (NGLDM), and Gray-level zone length matrix (GLZLM) (**Supplementary Material 1**). The association between features was assessed with Pearson's correlations (**Supplementary Material 2**).

Model Establishment

The optimal features needed to be selected first because the number of radiomics features was too large and not all of them were statistically significant. Features were chosen using three selection methods, namely, distance correlation, least absolute shrinkage and selection operator (LASSO), and gradient boosting decision tree (GBDT). Then three machine-learning classifiers were adopted to generate discriminative models, including random forest (RF), linear discriminant analysis (LDA), and support vector machine (SVM). LDA and SVM classifiers were chosen because they were representatives of linear and non-linear classification algorithms, respectively (13). Different from LDA or SVM, RF was considered the hybrid model of linear and non-linear classifiers by some researchers (14, 15). The patients were randomly divided into the training group and the validation group with the ratio of 4:1. The models were first trained with the training group and then applied to the independent validation group to test their discriminative performance, and this procedure was repeated for 100 cycles. A confusion matrix was established combining the histopathological results and predictions of models based on which the sensitivity, specificity, and accuracy were calculated. Area under the receiver operating characteristic curve (AUC) for both training group and validation group was also recorded to evaluate the discriminative ability of different models. The workflow from imaging processing to model establishment was shown in **Figure 3**.

RESULTS

Patient Characteristics

A total number of 101 qualified patients (50 individuals with AO and 51 with atypical low-grade oligodendroglioma) were enrolled in the present study. The gender ratio of participants was 54:47 (male:female). The average ages of patients with AO and atypical low-grade oligodendroglioma were 47.1 and 38.7 years, respectively. The detailed characteristics of patients and lesions were summarized in **Table 1**.

Model Assessment

A total of nine predictive models were built through the combination of three selection methods (distance correlation, LASSO, and GBDT) and three machine-learning classifiers (RF, LDA, and SVM). Radiomics features from T1C images and FLAIR images were introduced into models, respectively. All of the classifiers represented feasible discriminative ability with AUC more than 0.840 in the validation group when combined with the suitable selection method.

Among models using parameters from T1C images, the combination of LASSO and RF classifier (LASSO + RF) was proven to show the highest AUC of 0.904 in the validation group. Moreover, RF classifier seemed to be the optimal classification algorithm in differentiation for the reason that all RF-based models showed excellent performance with AUC

TABLE 1 | Characteristics of patients and lesions.

Characteristics	Atypical low-grade oligodendroglioma (*n* = 51)	Anaplastic oligodendroglioma (*n* = 50)
Age, *n* (%)		
0–20 years	5 (9.8)	1 (2.0)
21–40 years	22 (43.1)	15 (30.0)
41–60 years	19 (37.3)	23 (46.0)
61–80 years	5 (9.8)	11 (22.0)
Mean age (range) (year)	38.7 (7–71)	47.1 (16–76)
Gender, *n* (%)		
Male	29 (56.9)	25 (50.0)
Female	22 (43.1)	25 (50.0)
Ki-67 labeling index, *n* (%)		
<10%	35 (68.6)	9 (18.0)
≥10%	16 (31.4)	41 (82.0)
Average days between MR scan and surgery	9.4	7.9

MR, magnetic resonance.

over 0.920 in training group and 0.870 in the validation group (**Table 2**). For three models using the LDA classifier, receiver operating characteristic (ROC) analysis suggested that they all represented feasible discriminative ability, with AUC of 0.880 (distance correlation + LDA), 0.835 (LASSO + LDA), and 0.879 (GBDT + LDA) in the validation group (**Table 3**). For SVM-based models, only distance correlation + SVM showed feasible performance, with AUC of 0.866 in the validation group. Inadequate discriminative ability was observed in LASSO + SVM (AUC = 0.702 in the validation group) compared to other models, and overfitting was observed in GBDT + SVM (**Table 4**).

Among models using parameters from FLAIR images, GBDT + RF was found to represent the highest AUC of 0.861 in the validation group. Besides, other two models using RF classifier also displayed feasible discriminative ability, with AUC of 0.836 (distance correlation + RF) and 0.855 (LASSO + RF) in the validation group (**Table 2**). For three models using LDA classifier, ROC analysis demonstrated that the AUC in the validation group were 0.843, 0.819, and 0.848, respectively (**Table 3**). Among SVM-based models, distance correlation + SVM represented the best performance in differentiation with AUC of 0.860 in the validation group. Overfitting was observed in GBDT + SVM again, indicating that this model might be unsuitable for the grade prediction (**Table 4**).

DISCUSSION

Accurate preoperative evaluation of tumor grade is important for treatment facilitation and prognosis prediction. Lacking specific blood biomarkers, MR scan is commonly performed to evaluate oligodendroglioma grade pre-surgically with high spatial resolution and tissue resolution. However, atypical low-grade oligodendroglioma with contrast enhancement could complicate the differentiation from AO (16). Searching for accurate diagnosis, the value of advanced MRI techniques in oligodendroglioma grading had been investigated in previous studies (17, 18). Nevertheless, these advanced imaging techniques require additional expense and platforms and are not routinely conducted for every patient in clinical work. In the current study, a series of radiomics parameters were extracted from conventional MR sequences and fed into machine-learning classifiers to differentiate AO from atypical low-grade oligodendroglioma. Several predictive models with suitable combination were proven to represent feasible ability in grade prediction. Given that both T1C and FLAIR sequences are routinely performed in clinical examination, machine learning-based radiomics could potentially serve as the imaging biomarkers to aid preoperative diagnosis.

Radiomics has been investigated in recent studies, implying that the parameters are associated with tumor histopathology and abnormal microenvironment. The texture features calculate the image characteristics from different aspects, statistically reflecting intratumoral heterogeneity, cellular density, and level of vascularization (19–21). This theory has been verified by previous researches that the shift of texture parameters was associated with irregularity in blood vessel distribution and intratumoral hypoxia (22, 23). Given that these biological procedures were regulated by DNA, texture parameters were also related to molecular pathologic characteristics of tumors, such as mutation status of IDH and Kirsten Ras (KRAS) (24, 25). As for oligodendrogliomas, AO is histologically characterized by high cellular density, nuclear atypia, and microvascular proliferation, which might contribute to its radiological characteristics, such as contrast enhancement. Thus, we hypothesized that texture parameters might help discriminate between grade II and III oligodendrogliomas.

Moreover, with analyzable statistics converted from images, the novel computer technology could be employed. Similar researches suggested that radiomics combined with machine-learning algorithms displayed promising potential in various fields, including differential diagnosis of glioblastoma, pre-surgical grading of glioma, and prediction of patient survival outcomes (8, 26–28). It is worth noting that previous studies primarily focused the value of radiomics in distinguishing low-grade glioma vs. high-grade glioma, whereas the possible different characteristics among the histological subtypes of glioma were not taken into consideration (29–31). However, the heterogeneity of different glioma subtypes might interfere with the accuracy of the models. Therefore, our study first applied radiomics in grade prediction of oligodendrogliomas, a specific, common subtype of gliomas. More importantly, we focused on the situation where visual inspection was not sufficient in discrimination, aiming to explore the ability of radiomics-based machine learning in differentiating AO from atypical low-grade oligodendroglioma.

Compared with previous studies on glioma grading, we employed more selection methods (distance correlation, LASSO, and GBDT) and machine learning classifiers (RF, LDA, and SVM), wishing to identify the optimal model with the best discriminative performance. The results indicated that all of

TABLE 2 | Discriminative performance of models using RF classifier and different selection methods in distinguishing anaplastic oligodendroglioma from atypical low-grade oligodendroglioma in the training group and the validation group.

Selection method	Training group				Validation group			
	AUC	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity
T1C image								
Distance correlation	0.927	0.928	0.959	0.901	0.874	0.876	0.925	0.825
LASSO	0.945	0.946	0.976	0.921	0.904	0.900	0.971	0.833
GBDT	0.959	0.960	0.984	0.939	0.896	0.895	0.952	0.838
FLAIR image								
Distance correlation	0.911	0.835	0.775	0.915	0.836	0.833	0.813	0.868
LASSO	0.946	0.863	0.844	0.882	0.855	0.756	0.780	0.725
GBDT	0.957	0.882	0.839	0.931	0.861	0.783	0.770	0.806

RF, random forest; AUC, area under the curve; LASSO, least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree; T1C, contrast-enhanced T1-weighted; FLAIR, fluid-attenuation inversion recovery.

TABLE 3 | Discriminative performance of models using LDA classifier and different selection methods in distinguishing anaplastic oligodendroglioma from atypical low-grade oligodendroglioma in the training group and the validation group.

Selection method	Training group				Validation group			
	AUC	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity
T1C image								
Distance correlation	0.896	0.898	0.919	0.879	0.880	0.886	0.935	0.835
LASSO	0.928	0.929	0.949	0.911	0.835	0.829	0.926	0.748
GBDT	0.918	0.918	0.918	0.917	0.879	0.881	0.904	0.854
FLAIR image								
Distance correlation	0.866	0.796	0.727	0.900	0.843	0.783	0.740	0.887
LASSO	0.891	0.807	0.752	0.879	0.819	0.739	0.735	0.746
GBDT	0.943	0.862	0.836	0.889	0.848	0.817	0.802	0.841

LDA, linear discriminant analysis; AUC, area under the curve; LASSO, least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree; T1C, contrast-enhanced T1-weighted; FLAIR, fluid-attenuation inversion recovery.

TABLE 4 | Discriminative performance of models using SVM classifier and different selection methods in distinguishing anaplastic oligodendroglioma from atypical low-grade oligodendroglioma in the training group and the validation group.

Selection method	Training group				Validation group			
	AUC	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity
T1C								
Distance correlation	0.885	0.889	0.981	0.829	0.866	0.857	0.989	0.760
LASSO	0.759	0.770	0.930	0.700	0.702	0.657	0.881	0.570
GBDT	1.000	1.000	1.000	1.000	/	/	/	/
FLAIR								
Distance correlation	0.904	0.738	0.650	0.965	0.860	0.772	0.715	0.953
LASSO	0.712	0.689	0.616	0.878	0.606	0.678	0.664	0.716
GBDT	1.000	1.000	1.000	1.000	/	/	/	/

SVM, support vector machine; AUC, area under the curve; LASSO, least absolute shrinkage and selection operator; GBDT, gradient boosting decision tree; T1C, contrast-enhanced T1-weighted; FLAIR, fluid-attenuation inversion recovery; /, overfitting.

the classifiers represented feasible discriminative ability when combined with suitable selection method, and RF-based models showed the best performance with highest AUC in the validation group. RF classifier is a robust classification algorithm that has represented high discriminative performance in many studies (24, 32, 33). The mechanism of RF classification algorithm is to build subtrees by using the training bootstrap samples and choose the classification with the most votes over all trees in the forest (34). On the other hand, the results also indicated that the selection method with different mechanisms may have effects on the performance of the models. Distance correlation is the representative of filter models that rank features

based on certain characteristics and remove irrelevant features without classification algorithms, whereas LASSO and GBDT were representatives of embedded models that embed feature selection with classifier construction (35). However, we must note that most models represented similar diagnostic performance, and the differences in AUC may be partly attributed to the relatively small study cohort. Future studies with larger sample sizes are required to validate our results and further investigate the optimal model for grade prediction.

There were some limitations in the present study. First, this was a retrospective study; the selection bias was inevitable. Second, radiomics features were extracted from T1C and FLAIR sequences, whereas the value of features from other sequences like diffusion-weighted imaging (DWI) was unclear. Future studies are required to explore whether the features from other sequences could help improve the discriminative ability. Third, our models were not externally validated because this study was conducted in a single institution. However, the image processing and model establishment were conducted using open-source packages, providing the potential for researchers to verify our results in the future. Fourth, considering IDH and 1p/19q status could be reflected in texture parameters, it is reasonable to think that other molecular biomarkers may be associated with parameters. However, this point was not considered in the current study because of the relatively small sample size and single subtype of gliomas. Future larger studies are required to validate our results and to rectify the defects.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Sichuan University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

YZ participated in study conception, image collection, feature extraction, and drafted the manuscript. CC participated in image collection, feature extraction, and manuscript revision. YC and YT participated in feature extraction and manuscript revision. WG extracted texture feature and performed statistical analysis. HX and XO collected image data and extracted texture feature. JW participated in model construction and statistical analysis. HL participated in data processing and statistical analysis. XM and JX participated in study conception and manuscript revision. All authors read and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01371/full#supplementary-material

Supplementary Material 1 | Original data of texture features extracted from T1C and FLAIR sequences.

Supplementary Material 2 | Heat map showing Pearson's correlations between all pairs of features.

REFERENCES

1. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* (2016) 131:803–20. doi: 10.1007/s00401-016-1545-1
2. Weller M, van den Bent M, Tonn JC, Stupp R, Preusser M, Cohen-Jonathan-Moyal E, et al. European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. *Lancet Oncol.* (2017) 18:e315–29. doi: 10.1016/S1470-2045(17)30194-8
3. Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2011–2015. *Neuro Oncol.* (2018) 20:iv1–86. doi: 10.1093/neuonc/noy131
4. van den Bent MJ, Smits M, Kros JM, Chang SM. Diffuse infiltrating oligodendroglioma and astrocytoma. *J Clin Oncol.* (2017) 35:2394–401. doi: 10.1200/JCO.2017.72.6737
5. van den Bent MJ, Chang SM. Grade II and III oligodendroglioma and astrocytoma. *Neurol Clin.* (2018) 36:467–84. doi: 10.1016/j.ncl.2018.04.005
6. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
7. Alcaide-Leon P, Dufort P, Geraldo AF, Alshafai L, Maralani PJ, Spears J, et al. Differentiation of enhancing glioma and primary central nervous system lymphoma by texture-based machine learning. *AJNR Am J Neuroradiol.* (2017) 38:1145–50. doi: 10.3174/ajnr.A5173
8. Papp L, Pötsch N, Grahovac M, Schmidbauer V, Woehrer A, Preusser M, et al. Glioma survival prediction with combined analysis of *in vivo* C-MET PET features, *ex vivo* features, and patient features by supervised machine learning. *J Nucl Med.* (2018) 59:892–9. doi: 10.2967/jnumed.117.202267
9. Lu Y, Liu L, Luan S, Xiong J, Geng D, Yin B. The diagnostic value of texture analysis in predicting WHO grades of meningiomas based on ADC maps: an attempt using decision tree and decision forest. *Eur Radiol.* (2019) 29:1318–28. doi: 10.1007/s00330-018-5632-7
10. Artzi M, Bressler I, Ben Bashat D. Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. *J Magn Reson Imaging.* (2019) 50:519–28. doi: 10.1002/jmri.26643
11. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging.* (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
12. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: A freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* (2018) 78:4786–9. doi: 10.1158/0008-5472.CAN-18-0125
13. Dellacasa Bellingegni A, Gruppioni E, Colazzo G, Davalli A, Sacchetti R, Guglielmelli E, et al. NLR, MLP, SVM, and LDA: a comparative analysis on

EMG data from people with trans-radial amputation. *J Neuroeng Rehabil.* (2017) 14:82. doi: 10.1186/s12984-017-0290-6
14. Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324
15. Jovic A, Bogunovic N. Random forest-based classification of heart rate variability signals by using combinations of linear and nonlinear features. (2010) 29:29–32. doi: 10.1007/978-3-642-13039-7_8
16. White ML, Zhang Y, Kirby P, Ryken TC. Can tumor contrast enhancement be used as a criterion for differentiating tumor grades of oligodendrogliomas? *AJNR Am J Neuroradiol.* (2005) 26:784–90. Available online at: http://www.ajnr.org/content/26/4/784.long
17. Arevalo-Perez J, Kebede AA, Peck KK, Diamond E, Holodny AI, Rosenblum M, et al. Dynamic contrast-enhanced MRI in low-grade versus anaplastic oligodendrogliomas. *J Neuroimaging.* (2016) 26:366–71. doi: 10.1111/jon.12320
18. Lin Y, Xing Z, She D, Yang X, Zheng Y, Xiao Z, et al. IDH mutant and 1p/19q co-deleted oligodendrogliomas: tumor grade stratification using diffusion-, susceptibility-, and perfusion-weighted MRI. *Neuroradiology.* (2017) 59:555–62. doi: 10.1007/s00234-017-1839-6
19. Kim JH, Ko ES, Lim Y, Lee KS, Han BK, Ko EY, et al. Breast cancer heterogeneity: MR imaging texture analysis and survival outcomes. *Radiology.* (2017) 282:665–75. doi: 10.1148/radiol.2016160261
20. Fujima N, Homma A, Harada T, Shimizu Y, Tha KK, Kano S, et al. The utility of MRI histogram and texture analysis for the prediction of histological diagnosis in head and neck malignancies. *Cancer Imaging.* (2019) 19:5. doi: 10.1186/s40644-019-0193-9
21. Nardone V, Tini P, Nioche C, Mazzei MA, Carfagno T, Battaglia G, et al. Texture analysis as a predictor of radiation-induced xerostomia in head and neck patients undergoing IMRT. *Radiol Med.* (2018) 123:415–23. doi: 10.1007/s11547-017-0850-7
22. Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA. Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. *Radiology.* (2013) 266:326–36. doi: 10.1148/radiol.12112428
23. Ganeshan B, Miles KA. Quantifying tumour heterogeneity with CT. *Cancer Imaging.* (2013) 13:140–9. doi: 10.1102/1470-7330.2013.0015
24. Zhou H, Chang K, Bai HX, Xiao B, Su C, Bi WL, et al. Machine learning reveals multimodal MRI patterns predictive of isocitrate dehydrogenase and 1p/19q status in diffuse low- and high-grade gliomas. *J Neurooncol.* (2019) 142:299–307. doi: 10.1007/s11060-019-03096-0
25. Oh JE, Kim MJ, Lee J, Hur BY, Kim B, Kim DY, et al. Magnetic resonance-based texture analysis differentiating KRAS mutation status in rectal cancer. *Cancer Res Treat.* (2019). doi: 10.4143/crt.2019.050. [Epub ahead of print].
26. Nakagawa M, Nakaura T, Namimoto T, Kitajima M, Uetani H, Tateishi M, et al. Machine learning based on multi-parametric magnetic resonance imaging to differentiate glioblastoma multiforme from primary cerebral nervous system lymphoma. *Eur J Radiol.* (2018) 108:147–54. doi: 10.1016/j.ejrad.2018.09.017
27. Yang Y, Yan LF, Zhang X, Nan HY, Hu YC, Han Y, et al. Optimizing texture retrieving model for multimodal MR image-based support vector machine for classifying glioma. *J Magn Reson Imaging.* (2019) 49:1263–74. doi: 10.1002/jmri.26524
28. Tian Q, Yan LF, Zhang X, Zhang X, Hu YC, Han Y, et al. Radiomics strategy for glioma grading using texture features from multiparametric MRI. *J Magn Reson Imaging.* (2018) 48:1518–28. doi: 10.1002/jmri.26010
29. Ditmer A, Zhang B, Shujaat T, Pavlina A, Luibrand N, Gaskill-Shipley M, et al. Diagnostic accuracy of MRI texture analysis for grading gliomas. *J Neurooncol.* (2018) 140:583–9. doi: 10.1007/s11060-018-2984-4
30. Xie T, Chen X, Fang J, Kang H, Xue W, Tong H, et al. Textural features of dynamic contrast-enhanced MRI derived model-free and model-based parameter maps in glioma grading. *J Magn Reson Imaging.* (2018) 47:1099–111. doi: 10.1002/jmri.25835
31. Wang Q, Li Q, Mi R, Ye H, Zhang H, Chen B, et al. Radiomics nomogram building from multiparametric MRI to predict grade in patients with glioma: a cohort study. *J Magn Reson Imaging.* (2019) 49:825–33. doi: 10.1002/jmri.26265
32. Park YW, Oh J, You SC, Han K, Ahn SS, Choi YS, et al. Radiomics and machine learning may accurately predict the grade and histological subtype in meningiomas using conventional and diffusion tensor imaging. *Eur Radiol.* (2018) 29:4068–76. doi: 10.1007/s00330-018-5830-3
33. Suh HB, Choi YS, Bae S, Ahn SS, Chang JH, Kang SG, et al. Primary central nervous system lymphoma and atypical glioblastoma: differentiation using radiomics approach. *Eur Radiol.* (2018) 28:3832–9. doi: 10.1007/s00330-018-5368-4
34. Azar AT, Elshazly HI, Hassanien AE, Elkorany AM. A random forest classifier for lymph diseases. *Comput Methods Programs Biomed.* (2014) 113:465–73. doi: 10.1016/j.cmpb.2013.11.004
35. Tang J, Alelyani S, Liu H. Feature selection for classification: a review. In Aggarwal CC, editor. *Data Classification: Algorithms and Applications*. Boca Raton, FL: CRC Press (2014). p. 37–64. doi: 10.1201/b17320

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 10 January 2020
doi: 10.3389/fonc.2019.01382

Differentiation of Small Hepatocellular Carcinoma From Dysplastic Nodules in Cirrhotic Liver: Texture Analysis Based on MRI Improved Performance in Comparison Over Gadoxetic Acid-Enhanced MR and Diffusion-Weighted Imaging

Xi Zhong[1†], Hongsheng Tang[2†], Bingui Lu[1], Jia You[1], Jinsong Piao[3], Peiyu Yang[1] and Jiansheng Li[1]*

[1] Department of Radiology, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, China, [2] Department of Abdominal Surgery, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, China, [3] Department of Pathology, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, China

OPEN ACCESS

Edited by:
Xuelei Ma,
Sichuan University, China

Reviewed by:
Andre Bongers,
University of New South Wales, Australia
Guolin Ma,
China-Japan Friendship Hospital, China

***Correspondence:**
Jiansheng Li
lijiansheng@gzhmu.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 23 May 2019
Accepted: 22 November 2019
Published: 10 January 2020

Citation:
Zhong X, Tang H, Lu B, You J, Piao J, Yang P and Li J (2020) Differentiation of Small Hepatocellular Carcinoma From Dysplastic Nodules in Cirrhotic Liver: Texture Analysis Based on MRI Improved Performance in Comparison Over Gadoxetic Acid-Enhanced MR and Diffusion-Weighted Imaging. Front. Oncol. 9:1382. doi: 10.3389/fonc.2019.01382

Background: Accurate characterization of small (3 cm) hepatocellular carcinoma (sHCC) and dysplastic nodules (DNs) in cirrhotic liver is challenging. We aimed to investigate whether texture analysis (TA) based on T2-weighted images (T2WI) is superior to qualitative diagnosis using gadoxetic acid-enhanced MR imaging (Gd-EOB-MRI) and diffusion-weighted imaging (DWI) for distinguishing sHCC from DNs in cirrhosis.

Materials and methods: Sixty-eight patients with 73 liver nodules (46 HCCs, 27 DNs) pathologically confirmed by operation were included. For imaging diagnosis, three sets of images were reviewed by two experienced radiologists in consensus: a Gd-EOB-MRI set, a DWI set, and a combined set (combination of Gd-EOB-MRI and DWI). For TA, 279 texture features resulting from T2WI were extracted for each lesion. The performance of each approach was evaluated by a receiver operating characteristic analysis. The area under the receiver operating characteristic curve (A_z), sensitivity, specificity, and accuracy were determined.

Results: The performance of TA ($A_z = 0.96$) was significantly higher than that of imaging diagnosis using Gd-EOB-MRI set ($A_z = 0.86$) or DWI set ($A_z = 0.80$) alone in differentiation of sHCC from DNs ($P = 0.008$ and 0.025, respectively). The combination of Gd-EOB-MRI and DWI showed a greater sensitivity (95.6%) but reduced specificity (66.7%). The specificity of TA (92.6%) was significantly higher than that of the combined set ($P < 0.001$), but no significant difference was observed in sensitivity (97.8 vs. 95.6%, $P = 0.559$).

Conclusion: TA-based T2WI showed a better classification performance than that of qualitative diagnosis using Gd-EOB-MRI and DW imaging in differentiation of sHCCs from DNs in cirrhotic liver. TA-based MRI may become a potential imaging biomarker for the early differentiation HCCs from DNs in cirrhosis.

Keywords: hepatocellular carcinoma, liver cirrhosis, diffusion magnetic resonance imaging, gadoxetic acid, texture analysis

INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the most common malignancies; almost 80% of HCC occurs in patients with cirrhosis (1, 2). Hepatocarcinogenesis in cirrhosis usually shows a multistep progression from benign nodules, early HCC, and progressive HCC (3). Early detection of HCC, differentiation from benign cirrhotic nodules, provides the greatest chance for long-term survival (4). However, a complete characterization of these nodules still remains a difficult diagnostic dilemma due to the overlap of imaging features (5, 6).

Based on the criteria of the American Association for the Study of Liver Diseases, arterial enhancement followed by later (portal or equilibrium phase) washout is defined as a conclusive diagnosis for HCC (7). However, this typical enhancement pattern is not always presented, especially for some well-differentiated or small HCCs (8, 9). Diffusion-weighted imaging (DWI) can provide additional value to routine dynamic MRI by improving the diagnostic sensitivity (10, 11). The restricted diffusion facilitates HCC diagnosis by reflecting tissue hypercellularity (12). Nevertheless, some small HCCs may not show restricted diffusion (13, 14).

Recently, as a hepatocyte-specific intake agent MR imaging, gadoxetic acid-enhanced MR imaging (Gd-EOB-MRI) provides both early dynamic vascular phase and delayed hepatobiliary phase (HBP) information, which has been increasingly applied in the characterization of liver nodules. Gd-EOB-MRI has been demonstrated a higher sensitivity for detecting HCCs than conventional dynamic MRI due to hypointensity on HBP images (5, 15, 16). However, some small HCCs may not show hypointensity on HBP images; in contrast, some DNs are hypointensity (5, 17, 18).

Texture analysis (TA) based on medical images is a postprocessing approach in differential diagnosis of benign and malignant diseases (19). TA based on MRI has been used for distinguishing breast cancer from normal tissue and classifying histological types (20), e.g., differentiating prostate cancer from normal tissue and classifying prostate cancers with different Gleason scores (21). In liver assessments, texture-based MRI can be used to differentiate different single liver lesions (22, 23), evaluate hepatic fibrosis and cirrhosis grades (24), and predict the HCC histological grade (25).

The value of TA-based MRI for discriminating cirrhotic nodules remains unclear; we hypothesized that MRI-based TA may be helpful to distinguish HCCs from DNs. Thus, we performed this study to estimate the feasibility of TA-based T2-weighted images in the differentiation of sHCC from DNs in cirrhotic liver.

MATERIALS AND METHODS

Patient Samples

This retrospective study was approved by the institutional review board of our hospital, and patient's informed consent was waived. We reviewed 455 consecutive patients with cirrhosis who underwent liver MRI to exclude HCC between January 2015 and October 2018. The inclusion criteria were as follows: (1) pathologically proven HCCs or DNs by surgical resection, (2) nodule diameter smaller than 3 cm and larger than 1 cm, (3) underwent both DW and Gd-EOB-MRI, and (4) received no treatment before MRI. Based on the inclusion criteria, ultimately a total of 68 patients [42 men, 26 women; median, 56 years (range, 30–73 years)] with 73 liver nodules (46 HCCs, 27 DNs) were enrolled. The patient inclusion flowchart is shown in **Figure 1**.

MRI Acquisition

MR imaging was performed in a 3.0-T whole-body MR system (Achieva; Philips Healthcare, Best, Netherlands) with a 16-channel phased-array coil. The MRI protocol consisted of a respiratory-triggered T1-weighted turbo field-echo in-phase and opposed sequence [TR/first echo TE and second echo TE, 10/2.5 ms (in-phase) and 3.55 ms (opposed-phase); flip angle, 10°; matrix, 256 × 224; bandwidth, 434.3 Hz/pixel] and a breath-hold fat-saturated T2-weighted fast spin-echo sequence (TR/TE, 2,096/72 ms; flip angle, 90°; matrix size, 324 × 256; bandwidth, 258.4 Hz/pixel) with a 5-mm section thickness and a field of view (FOV) of 30–38 cm. DWI was performed using a respiratory triggering single-shot echo planar imaging sequence with b values of 0 and 800 s/mm^2, spectral presaturation with inversion recovery for fat suppression, using a TR/TE of 1,600/70, matrix size of 100 × 100, acceleration factor of SENSE of 4.0, FOV of 30–35 cm, slice thickness of 5 mm, slice gap of 1 mm, and 33 axial slices. For Gd-EOB-MRI, unenhanced, arterial-phase (20–35 s), portal-phase (60 s), late-phase (3 min), and 20-min delayed HBP images were obtained using a T1-weighted three-dimensional (3D) turbo-field-echo sequence (T1 high-resolution isotropic volume examination; Philips Healthcare) (3.1/1.5; flip angle, 10°; matrix size, 228 × 211; bandwidth, 724.1 Hz/pixel) with a 2-mm section thickness and an FOV of 32–38 cm. The contrast agent was automatically administered intravenously at a rate of 1 ml/s for a dose of 0.025 mmol/kg body weight using a power injector, followed by a 20-ml saline flush.

Image Analysis

All images were analyzed separately and independently reviewed by two radiologists (B.G.L and P.Y.Y, with 15 and 10 years' experience of liver MR imaging, respectively) who were

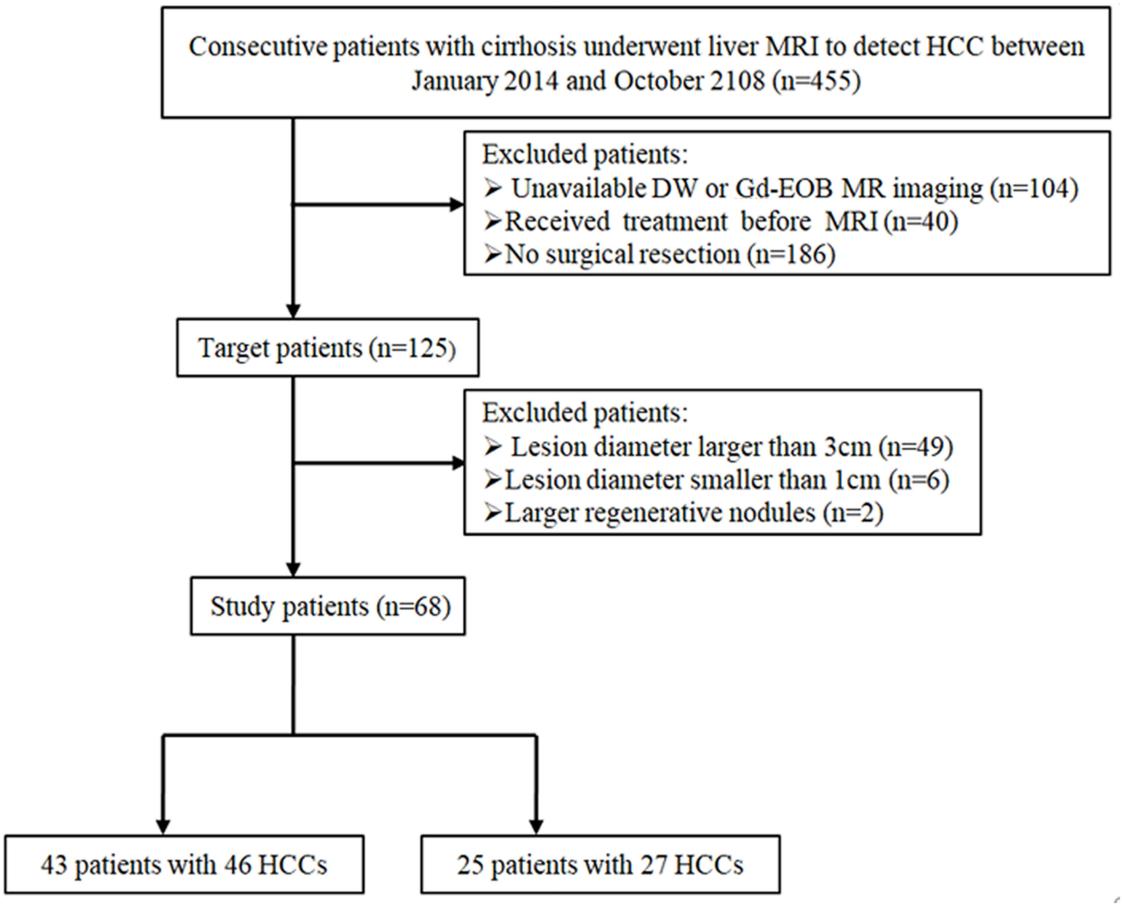


FIGURE 1 | Flowchart of the study population.

blinded to the patients' clinical data and the pathological diagnosis. Three image sets were reviewed, respectively: a Gd-EOB-MRI set (precontrast T1- and T2-weighted images and arterial, portal, equilibrium, and HBP images), a DWI set (precontrast T1- and T2-weighted images and DW images), and combined sets. Four-week interval between image reviews was performed for the three reviewing sessions to avert any recall bias. The signal intensity (SI) of each lesion was evaluated on Gd-EOB-MR and DW images. The SI features were classified into three groups: hypointensity, isointensity, or hyperintensity compared to the surrounding liver parenchyma.

As described in previous studies (5, 15, 26), in Gd-EOB-MRI set, the diagnostic criteria for HCC were defined as follows: (a) a nodule showed typical enhancement pattern (arterial enhancement and late portal or equilibrium washout); (b) a nodule with arterial enhancement without later washout, but hypointensity on HBP images, or peripheral rim enhancement on the late dynamic phase images (capsular appearance); and (c) a nodule without arterial enhancement, but larger than 1.5 cm and showed hypointensity on HBP images. In the DW set, if a lesion showed hyperintensity on DW images, it was interpreted as an HCC (14). In combined sets, if a lesion satisfied the HCC criteria of Gd-EOB-MRI or DWI, it would be identified as an HCC.

Texture Analysis

Texture Calculation

The axial FS T2-weighted images were exported in ".dicom" format from the PACS for texture analysis. One of the radiologists (X.Z) manually segmented images for each lesion using a free open-source software package MaZda 4.6 (URL: http://www.eletel.p.lodz.pl/programy/mazda/), and a single region of interest (ROI) was defined and delineated on the image section depicting the maximum lesion diameter (**Figure 2A**). Seven lesions (two HCCs and five DNs) were isointense on FS-T2-WI, in this case, T1-weighted or gadoxetic acid-enhanced images were used for accurate ROI placement. Refer to previous studies (20, 23), ROI gray-level normalization was performed by adjusting image intensities in the range of $u \pm 3\sigma$ (where u is the gray-level mean and σ is the gray-level standard deviation). A total of 279 texture parameters that derived from six statistical image descriptors were computed for each ROI (**Table 1**).

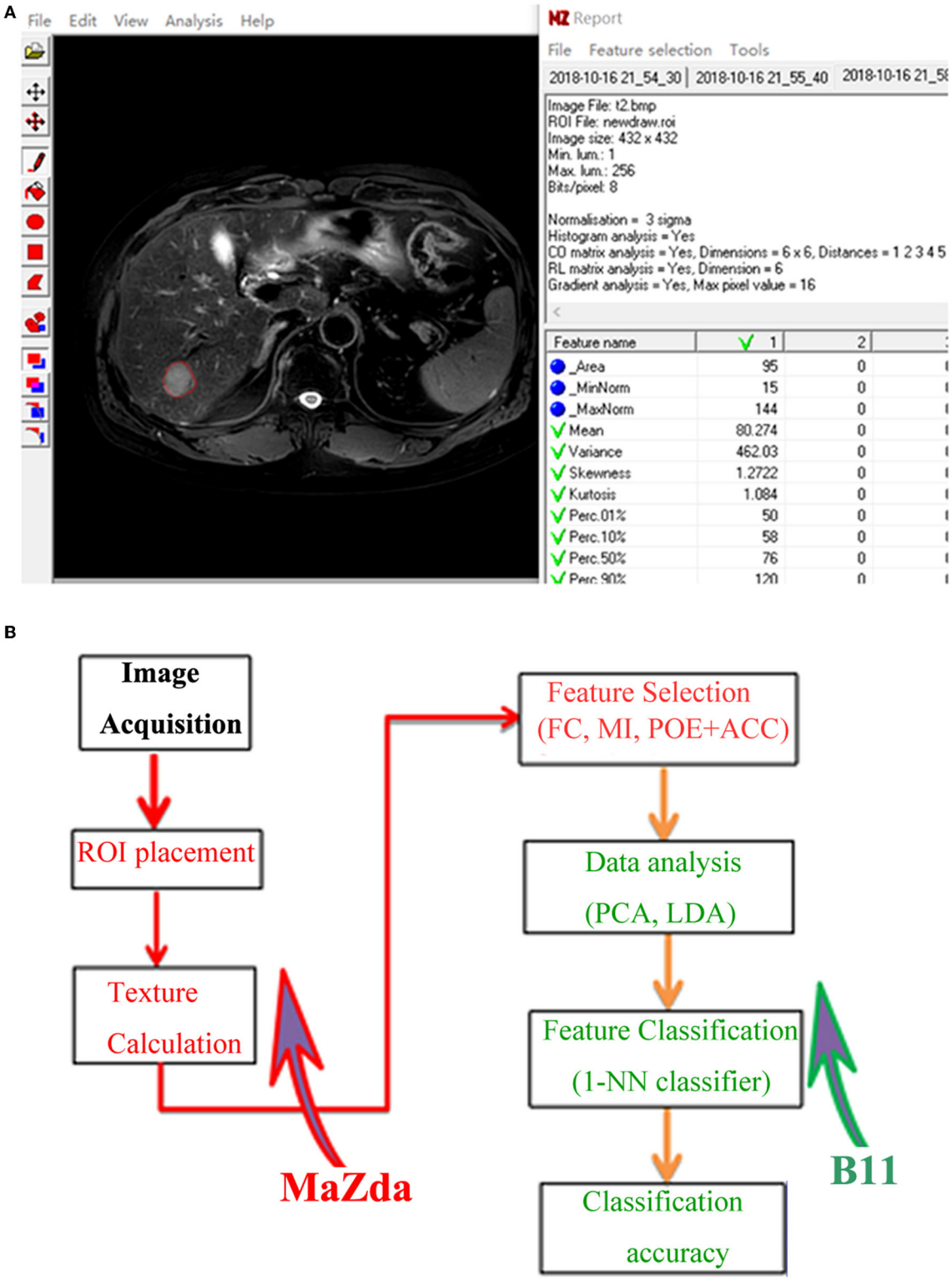


FIGURE 2 | (A) Texture calculation for a 63-year-old man with a pathologically proven hepatocellular carcinoma (HCC) on fat-saturated T2-weighted images by using MaZda. A single region of interest (ROI) was defined and delineated on the image section depicting the maximum lesion diameter. **(B)** Procedure of texture analysis: texture calculation and feature selection were performed using MaZda; data analysis and classification were performed using program B11. ROI, region of interest; FC, Fisher coefficients; MI, mutual information; POE + ACC, minimization of both classification error probability and average correlation coefficients; PCA, principal component analysis; LDA, linear discriminant analysis; 1-NN, *first* nearest neighbor.

TABLE 1 | Texture parameters calculated in MaZda.

Computational methods	Number	Texture parameters
Histogram	9	Mean, variance, skewness, kurtosis, percentiles 1, 10, 50, 90, and 99%
Co-occurrence matrix	220	Angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, and difference entropy. Features are computed for 5 between-pixels distances (1, 2, 3, 4, and 5) and for four various directions (vertical, horizontal, 0, and 135).
Run-length matrix	20	Run-length non-uniformity, gray-level non-uniformity, long-run emphasis, short run emphasis, and fraction of image in runs. Features are computed for four various directions (vertical, horizontal, 0, and 135).
Absolute gradient	5	Mean, variance, skewness, kurtosis, and percentage of pixels with nonzero gradient
Autoregressive model	5	Teta1 ($\theta1$), Teta2 ($\theta2$), Teta3 ($\theta3$), Teta4 ($\theta4$), and Sigma (σ)
Wavelet ($n = 20$)	20	WavEn (wavelet energy). Feature is computed at five scales within four frequency bands LL, LH, HL, and HH.

TABLE 2 | Patients' characteristics.

Parameters	Value
Patient number	68
Age median [range] (years)	56 (30–73)
Male/female	42/26
Child-pugh	
A	50
B	12
C	6
AFP serum >200 ng/ml	11
AFP serum<200 ng/ml	51
AFP unobtainable	6
Etiology of liver cirrhosis*	
HBV	51
HCV	13
Ethanol	8
Others	2

*AFP, Alpha-fetoprotein; *A patient could have multiple etiologies.*

TABLE 3 | Signal features of hepatocellular nodules.

SI features	HCCs ($n = 46$)	DNs ($n = 27$)	HGDN ($n = 13$)
DWI			
Hypointensity	0 (0%)	4 (14.8%)	2 (15.4%)
Isointensity	5 (10.9%)	15 (55.6%)	4 (30.8%)
Hyperintensity	41 (89.1%)	8 (29.6%)	7 (53.8%)
Arterial phase			
Hypointensity	6 (13.1%)	4 (14.8%)	1 (7.7%)
Isointensity	10 (21.7%)	10 (37.0%)	4 (30.8%)
Hyperintensity	30 (65.2%)	13 (48.1%)	8 (61.5%)
Portal phase			
Hypointensity	30 (65.2%)	4 (14.8%)	1 (7.7%)
Isointensity	4 (8.7%)	8 (29.6%)	2 (15.4%)
Hyperintensity	12 (26.1%)	15 (55.6%)	10 (76.9%)
Equilibrium phase			
Hypointensity	36 (78.3%)	6 (22.2%)	3 (23.0%)
Isointensity	6 (13.0%)	16 (59.2%)	5 (38.5%)
Hyperintensity	4 (8.7%)	5 (18.6%)	5 (38.5%)
HBPI			
Hypointensity	40 (87.0%)	3 (11.1%)	2 (15.4%)
Isointensity	5 (10.9%)	19 (70.3%)	7 (53.8%)
Hyperintensity	1 (2.1%)	5 (18.6%)	4 (30.8%)

SI, signal intensity; HCC, Hepatocellular carcinoma; DNs, Dysplastic nodules; HGDN, high-grade dysplastic nodules; DW, Diffusion-weighted imaging; HBPI, Hepatobiliary phase imaging.

Feature Selection

To determine the most discriminative texture features for differentiating sHCCs from DNs, as mentioned previously (23), we used three texture feature selection methods, including Fisher coefficients, minimization of both classification error probability and average correlation coefficients (POE + ACC), and mutual information (MI), respectively.

Feature Classification

Feature classification was performed in a statistical program B11 (version 4.6). As described in a previous study (27), principal component analysis (PCA) and linear discriminant analysis (LDA) were used for reducing the feature vector dimension and increasing the discriminative power. Then, the first nearest neighbor (1-NN) classifier with feature vector standardization was applied to determine classification accuracy (23, 24). The procedure of TA is shown in **Figure 2B**.

Histopathology Evaluation

International Working Party criteria was used for the evaluation of hepatocellular nodular (28). DNs were defined as a lesion with hepatocytes dysplasia but no definite histological features of malignancy, which were classified as low- or high-grade based on the cytological and architectural atypia (29).

Statistical Analysis

The sensitivity and specificity for differentiation of sHCCs from DNs were calculated for qualitative diagnosis and TA. The overall diagnostic efficiency was evaluated by calculating

area under the receiver operating characteristic (ROC) curve (A_z), and the ROC curves were plotted based on the dichotomous classification results of each diagnostic approach, and the diagonal segments are produced by ties. Mann–Whitney and chi-square test (or Fisher test) were performed for quantitative and categorical variables, respectively. All the statistical tests were performed using SPSS 16.0 (SPSS Inc., Chicago, IL, USA) package, and statistical significance was accepted for $P < 0.05$.

RESULT

Patient Characteristics

Patients' characteristics are summarized in **Table 2**. Of the 68 patients, 63.2% (43/68) with 46 lesions were diagnosed with HCCs (diameter range, 1.2–3.0 cm; mean, 1.9 cm), and 36.8% (25/68) of patients with 27 nodules were diagnosed with DNs (diameter range, 1.0–2.9 cm; mean, 1.7 cm). Of the 27 DNs, 13 were high-grade DNs (HGDNs) and 14 were low-grade DNs (LGDNs).

TABLE 4 | Diagnostic performance of DW and gadoxetic acid-enhanced imaging.

Imaging sets	A_z [95% CI]	Sensitivity	Specificity	Accuracy
Gd-EOB-MRI set	0.86 [0.76, 0.95]	82.6% (38/46)	88.9% (24/27)	84.9% (62/73)
DWI set	0.80 [0.68, 0.91]	89.1% (41/46)	70.3% (19/27)	82.2% (60/73)
Combined sets	0.81 [0.69, 0.93]	95.6% (44/46)	66.7% (18/27)	84.9% (62/73)

Diagnostic Performance of Qualitative MRI Diagnosis

The SI features of HCCs and DNs on Gd-EOB-MRI and DWI are shown in **Table 3.** The diagnostic performance of each imaging set for differentiating sHCCs from DNs are shown in **Table 4**.

In Gd-EOB-MRI set, among the 46 sHCCs, 50% (23/46) of lesions showed typical enhancement patterns (**Figure 3**): 13 lesions showed arterial enhancement without late washout but hypointensity on HBP images (**Supplementary Figure 1**), 2 nodules showed hypovascular on dynamic study, but larger than 1.5 cm and showed hypointensity on HBP images (**Figure 4**), and the other 8 nodules satisfied none of the Gd-EOB-MRI criteria for HCC (**Supplementary Figure 2**). Of the 27 DNs, no nodule showed atypical enhancement patterns, 24 nodules (11 HGDNs, 13 LGDNs) showed iso/hyperintensity on HBP images (**Figure 5**), and 3 lesions (2 HGDNs, 1 LGDN) showed arterial enhancement and hypointensity on HBP images. The sensitivity and specificity for differentiating sHCCs from DNs in Gd-EOB-MRI set were 82.6% (38/46) and 88.9% (24/27), respectively.

In DWI set, 41 nodules among the 46 sHCCs showed hyperintensity on DWI (**Figure 3**), and 5 nodules showed iso/hypointensity on DWI (**Figure 4**). Of the 27 DNs, 19 nodules (6 HGDNs, 13 LGDNs) showed iso/hypointensity on DWI (**Supplementary Figure 3**), and 8 nodules (6 HGDNs, 2 LGDNs)

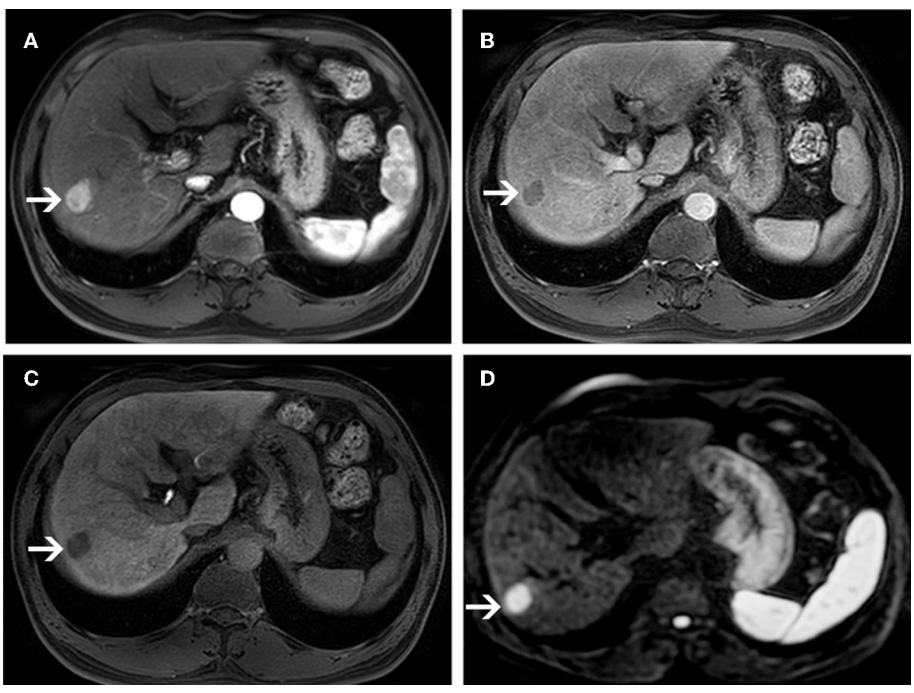


FIGURE 3 | MR images of a 58-year-old man with a pathologically proven HCC (white arrows) and a history of hepatitis B virus infection. An arterial-phase image **(A)** shows an enhancing nodule in segment VI of the liver. An equilibrium phase MR image **(B)** shows a nodule demonstrating washout of contrast material, and showing capsular appearance. At the hepatobiliary phase **(C)**, the lesion is hypointensity compared to the surrounding liver parenchyma. On diffusion-weighted image **(D)**, the lesion is hyperintensity compared to the surrounding liver parenchyma.

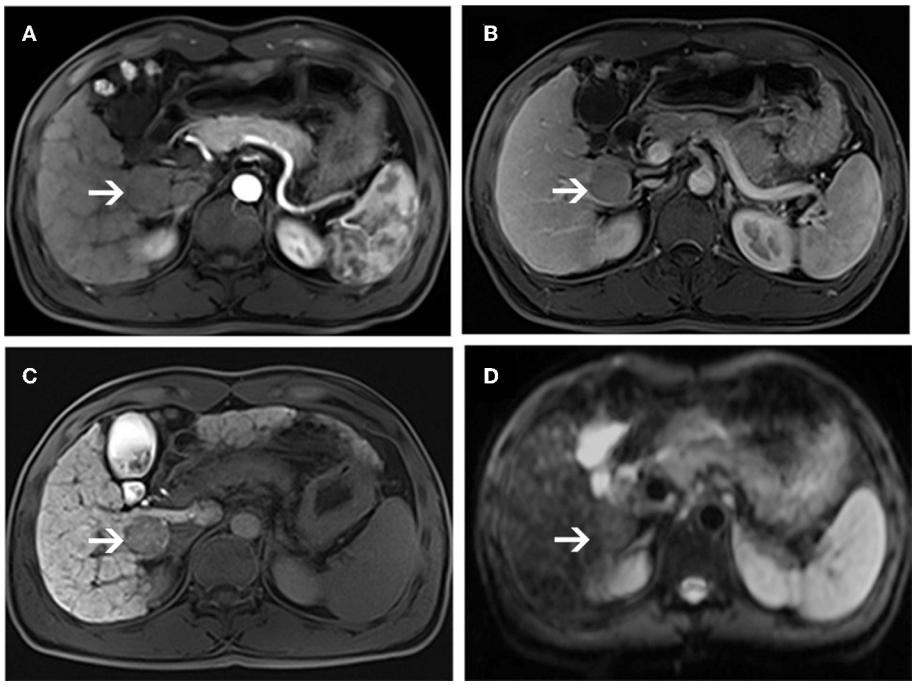


FIGURE 4 | MR images of a 66-year-old woman with a pathologically proven HCC (white arrows) and a history of hepatitis C virus infection. An arterial-phase image **(A)** shows a hypovascular nodule in segment I of the liver. An equilibrium phase MR image **(B)** shows a slightly hypointensity nodule compared to the surrounding liver parenchyma. A hepatobiliary phase image **(C)** shows a hypointensity lesion compared to the surrounding liver parenchyma. On diffusion-weighted image **(D)**, the lesion is nearly isointensity compared to the surrounding liver parenchyma.

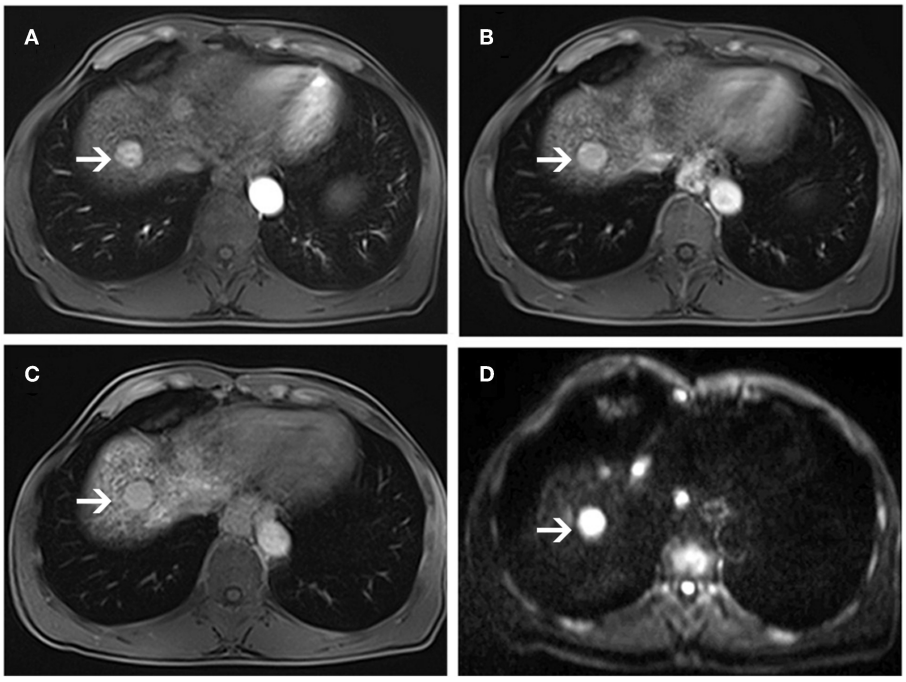


FIGURE 5 | MR images of a 56-year-old man with a pathologically proven high-grade dysplastic nodule (white arrows) and a history of hepatitis B virus infection. Arterial-phase image **(A)** shows an enhancing nodule in segment VIII of the liver. Equilibrium-phase MR image **(B)** shows a nodule not demonstrating washout of the contrast material. A hepatobiliary phase image **(C)** shows nearly isointensity compared to the surrounding liver parenchyma. On diffusion-weighted image **(D)**, the lesion shows hyperintensity compared to the surrounding liver parenchyma.

TABLE 5 | Texture feature subsets best-suited for the discrimination of w-HCCs and DNs on T2-W images, according to Fisher coefficient, the PEO+ACC, and Mutual information.

Feature rank	Fisher coefficient	POE + ACC	Mutual information
1	WavEnLL_s-1	WavEnHH_s-3	WavEnLL_s-2
2	WavEnLL_s-2	WavEnLH_s-3	WavEnLL_s-1
3	S(0,1)SumOfSqs	WavEnLL_s-3	S(0,5)SumAverg
4	S(0,1)SumAverg	WavEnHH_s-2	S(0,3)SumOfSqs
5	S(0,1)SumVarnc	WavEnLH_s-2	S(0,2)SumOfSqs
6	S(1,0)SumOfSqs	WavEnLL_s-2	S(1,-1)SumVarnc
7	S(1,0)SumVarnc	WavEnLH_s-1	S(1,-1)SumAverg
8	S(0,2)SumAverg	WavEnLL_s-1	S(0,1)SumAverg
9	S(1,-1)SumAverg	Vertl_LngREmph	S(0,1)SumOfSqs
10	S(0,3)SumAverg	S(0,1)SumAverg	S(1,0)SumOfSqs

POE + ACC, minimization of both classification error probability and average correlation coefficients; WavEn, Wavelet energy; SumOfSqs, Sum of squares; SumAverg, Sum average; SumVarnc, Sum variance; Vertl_LngREmph, Vertical long-run emphasis.

showed hyperintensity on DWI (**Figure 5**). The sensitivity and specificity for identifying sHCCs from DNs in DWI set were 89.1% (41/46) and 70.3% (19/27), respectively.

In combined sets, of the 46 sHCCs, only 2 nodules were mistaken for DNs. Among the 27 DNs, 18 nodules (7 HGDNs, 11 LGDNs) were diagnosed accurately. Consequently, the sensitivity and specificity for differentiating sHCCs from DNs in combined sets was 95.6% (44/46) and 66.7% (18/27), respectively.

TA Results

Texture subsets based on MI and Fisher coefficients were frequently derived from the co-occurrence matrix, whereas texture features created using the POE + ACC method were frequently derived from wavelet (**Table 5**).

Fisher coefficients, POE + ACC, and MI methods resulted in a similar misclassification rate of 4.1–6.8, 4.1–6.8, and 4.1–5.5%, respectively (**Table 6**). In terms of feature classification, PCA and LDA resulted in an equivalent misclassification rate of 4.1–6.8 and 4.1–5.5%, respectively (**Table 6**).

Both LDA combining Fisher coefficients and PCA combining MI Fisher showed the lowest misclassification rate of 4.1% (3/73). Only one HCC was misclassified as a DN, and two DNs were misclassified as HCCs (**Figures 6A,B**). With regard to the ROC analysis, TA demonstrated an A_z, sensitivity, specificity, and accuracy of 0.96 (95% CI: 0.91, 1), 97.8% (45/46), 92.6% (25/27), and 95.9% (70/73), respectively (**Table 5**).

Comparison of Diagnostic Performance

The ROC curve of each diagnostic method is shown in **Figure 6C**. The diagnostic performance of TA ($A_z = 0.96$, 95% CI: 0.91, 1) was significantly higher than that of imaging diagnosis with DWI ($A_z = 0.80$, 95% CI: 0.68, 0.91) or Gd-EOB-MRI ($A_z = 0.86$, 95% CI: 0.76, 0.95) alone ($P = 0.008$ and 0.025, respectively). The specificity of TA (92.6%) was significantly higher than that of DWI and Gd-EOB-MRI combined (66.7%) ($P < 0.001$), but no significant difference was observed in sensitivity (97.8 vs. 95.6%; $P = 0.559$).

DISCUSSION

This study aimed to identify whether MRI-based TA can be used to distinguish sHCC from DNs in cirrhotic liver. We also compared the performance of TA with DWI and Gd-EOB-MRI. The findings showed that TA-based T2WI had a satisfactory diagnostic value. The diagnostic efficacy of TA was significantly higher than that of qualitative diagnosis with DWI or Gd-EOB-MRI alone. Although the combination of DWI and Gd-EOB-MRI showed sensitivity equivalent to that of TA, TA showed significantly higher specificity than that of the combination qualitative diagnosis.

In the present study, we found that only 50% of sHCCs fit the American Association for the Study of Liver Diseases criteria, but HBP imaging improved the detection of sHCC; up to 15 sHCCs with atypical enhancement were detected by hypointensity on HBP images. In Gd-EOB-MRI set, the sensitivity and specificity were 82.6 and 88.9%, respectively, which is similar to the 85% sensitivity but significantly higher than the 42% specificity on imaging using gadoxetic acid disodium (14), and is similar to the 92% specificity but higher than the 71% sensitivity on imaging using gadobenate dimeglumine (30). Furthermore, Gd-EOB-MRI yields a better specificity than that of DWI set, with a specificity 18.6% greater than that of DWI, which is inconsistent with one study (14) in which Gd-EOB-MRI showed lower specificity than DWI for differentiating HCC from benign hepatic nodules. Nevertheless, we still found that 13.4% of sHCCs did not show hypointensity on 20-mine HBP images, which may be related to the overexpression of organic anionic transporting polypeptide 8 (OATP-8) in tumors; about 5–12% of small HCCs overexpress organic anionic transporting polypeptide 8 (18).

In DWI set, we found that the overall sensitivity in the identification of sHCCs and DNs was almost 89.1%, which was supported by previous reports showing that 81–88% of sHCCs showed hyperintensity on DW images (14, 31). Nevertheless, our study showed a relatively low specificity of 70.3% compared with some previous studies that reported specificity values of 79.0–94.4% (32, 33), which may be attributed to six HGDNs that showed some imaging features supporting HCC, such as hyperintensity on arterial phase without washout, hypointensity on HBP images, and/or hyperintensity on DWI.

In our study, the combination DW and Gd-EOB-MRI demonstrated an increase in the sensitivity for diagnosing sHCC compared with each imaging modality alone. Thus, the results are concordant with the previous data reported by Park et al. (34). However, our study resulted in a lower specificity of only 66.7%. Other than the expected variation between observers and institutions, the difference might be attributable to the fact that the benign hepatic nodules included in our study were all DNs.

The feasibility of TA in the classification of liver lesions has been widely discussed in CT and MRI (22, 23, 25, 35, 36). It is well-known that ROI placement plays a key role for TA; if a lesion shows isointensity, it may be difficult to place an ROI accurately. In this study, relatively large number of liver nodules showed isointensity on both DWI and dynamic Gd-EOB-MRI. Thus, this study analyzed the value of TA based on T2W-MRI images in discrimination of sHCC from DNs in cirrhosis. To our

TABLE 6 | Diagnostic performance of texture analysis.

Feature selection method	TA Method	Misclassified rates	A_z [95% CI]	Sensitivity	Specificity	Accuracy
Fisher coefficient	PCA	5/73 (6.8%)	0.94 [0.87, 1]	97.8% (45/46)	85.2% (23/27)	93.2% (68/73)
	LDA	3/73 (4.1%)	0.96 [0.91, 1]	97.8% (45/46)	92.6% (25/27)	95.9% (70/73)
POE + ACC	PCA	4/73 (5.5%)	0.94 [0.87, 1]	95.7% (44/46)	92.6% (25/27)	94.5% (69/73)
MI	LDA	5/73 (6.8%)	0.93 [0.85, 1]	95.7% (44/46)	88.9% (24/27)	93.2% (68/73)
	PCA	3/73 (4.1%)	0.96 [0.91, 1]	97.8% (45/46)	92.6% (25/27)	95.9% (70/73)
	LDA	4/73 (5.5%)	0.94 [0.87, 1]	95.7% (44/46)	92.6% (25/27)	94.5% (69/73)

POE + ACC, Minimization of both classification error probability and average correlation coefficients; MI, Mutual information; PCA, Principle component analysis; LDA, Linear discriminant analysis.

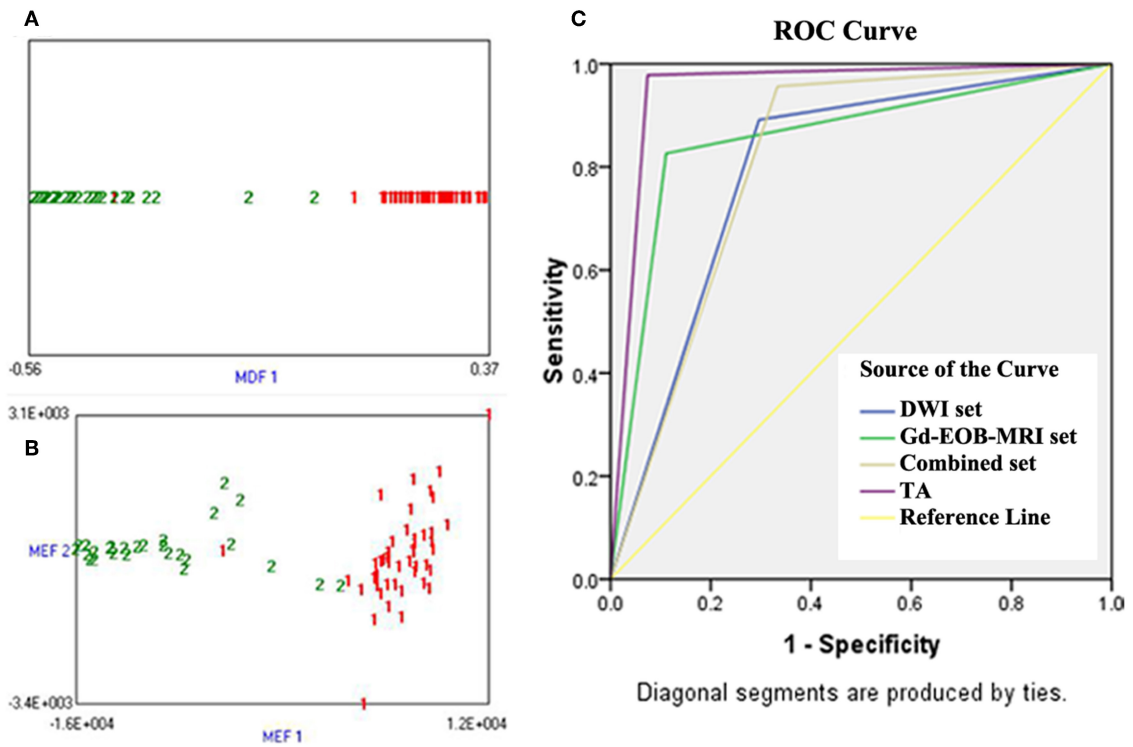


FIGURE 6 | (A) Linear discriminant analysis (LDA) combining Fisher coefficient for classifying small hepatocellular carcinoma (sHCC) and dysplastic nodules (DNs), 1 (red) represents sHCC and 2 (green) represents DNs; it shows that one sHCC was misclassified as a DN, and two DNs were misclassified as sHCC. **(B)** Principal component analysis (PCA) combining MI for classifying sHCC and DNs, 1 (red) represents sHCC and 2 (green) represents DNs; it shows that one sHCC was misclassified as a DN, and two DNs were misclassified as sHCC. **(C)** Receiver operating characteristic (ROC) curves of imaging sets and TA for the differentiation of sHCC from DNs, the ROC curves were plotted based on the dichotomous classification results of each diagnostic approach, and the diagonal segments are produced by ties. The position of the "kink" of curves represented diagnostic efficacy; *Y*-axis represented sensitivity, and *X*-axis represented 1-specificity.

knowledge, this is the first study to assess whether HCCs and DNs in cirrhosis can be fully classified using TA. In comparison with previous studies that used T2WI-based TA for classification of liver lesions, the misclassification rate by TA was 4.1%, which was lower than the 9.7% misclassification rate for distinguishing HCC from hepatic hemangioma and metastases (22) and the 12% misclassification rate for distinguishing liver cysts and hemangiomas (23). Furthermore, the primary advantage of our study was that we compared the diagnostic efficacy of TA with imaging diagnosis and found that TA showed better performance than that shown by imaging diagnosis with DW and gadoxetic acid-enhanced imaging alone. Although combined imaging strategy showed similar sensitivity as TA (95.6 vs. 97.8%) for identification of sHCC and DNs, the specificity of TA (92.6%) was significantly higher than that of the combined approach (63.0%).

In terms of misclassification rates, the performances of these feature selection methods (Fisher, POE + ACC, or MI)

showed no clear superiority, supporting the results of previous studies (20, 23). For the MI and Fisher methods, the texture parameters resulting from the co-occurrence matrix were more frequently assigned to the feature subsets than parameters of any other category, thus supporting findings of previous studies. Interestingly, for POE + ACC selection, texture parameters resulting from the wavelet were more frequently assigned to the feature subsets than parameters of any other category.

This study had some limitations. First, we could not divide the study population into training and test datasets due to the relatively small size and because we were primarily interested in the feasibility of texture-based classification for identification of sHCCs and DNs in cirrhosis. Second, we did not assess the lesions that were not detected on MRI because the TA and imaging diagnosis is quite difficult to perform in those lesions, so it might cause the possibility of a bias at inclusion. Third, the performance of TA combined with qualitative diagnosis was not assessed because TA has showed significantly higher performance than qualitative imaging diagnosis. Actually, the combination of TA and qualitative diagnosis may improve performance, and it needs further researches to confirm the additional value of combination diagnosis.

In conclusion, this preliminary study demonstrates that MRI-based TA shows better classification performance than imaging diagnosis for discriminating sHCC from DNs in cirrhotic liver. Although promising, these results are preliminary and require verification using a larger and independent dataset to appraise their potential for clinical translation. After validation, texture-based MRI may become a potential imaging biomarker for early differentiating HCCs from DNs in cirrhosis.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

This retrospective study was approved by the institutional review board at Affiliated Cancer Hospital & Institute of Guangzhou Medical University, and the requirement of patients' informed consent was waived.

AUTHOR CONTRIBUTIONS

XZ and JL: conception and design. XZ and HT: manuscript writing. JY, JP, and PY: provision of study materials or patients. JY and JP: collection and assembly of data. BL and JL: MRI analysis and interpretation. XZ and HT: statistical analysis. XZ and JL: final approval of manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01382/full#supplementary-material

REFERENCES

1. El-Serag HB, Davila JA, Petersen NJ, McGlynn KA. The continuing increase in the incidence of hepatocellular carcinoma in the United States: an update. *Ann Intern Med*. (2003) 139:817–23. doi: 10.7326/0003-4819-139-10-200311180-00009
2. McGlynn KA, London WT. Epidemiology and natural history of hepatocellular carcinoma. *Best Pract Res Clin Gastroenterol*. (2005) 19:3–23. doi: 10.1016/j.bpg.2004.10.004
3. Choi BI, Takayasu K, Han MC. Small hepatocellular carcinomas and associated nodular lesions of the liver: pathology, pathogenesis, and imaging findings. *AJR Am J Roentgenol*. (1993) 160:1177–87. doi: 10.2214/ajr.160.6.8388618
4. Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirenti A, Bozzetti F, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med*. (1996) 334:693–9. doi: 10.1056/NEJM199603143341104
5. Di Martino M, Anzidei M, Zaccagna F, Saba L, Bosco S, Rossi M, et al. Qualitative analysis of small (</=2 cm) regenerative nodules, dysplastic nodules and well-differentiated HCCs with gadoxetic acid MRI. *BMC Med Imaging*. (2016) 16:62. doi: 10.1186/s12880-016-0165-5
6. Park HJ, Choi BI, Lee ES, Park SB, Lee JB. How to differentiate borderline hepatic nodules in hepatocarcinogenesis: emphasis on imaging diagnosis. *Liver Cancer*. (2017) 6:189–203. doi: 10.1159/000455949
7. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology*. (2011) 53:1020–2. doi: 10.1002/hep.24199
8. Jang HJ, Kim TK, Burns PN, Wilson SR. Enhancement patterns of hepatocellular carcinoma at contrast-enhanced US: comparison with histologic differentiation. *Radiology*. (2007) 244:898–906. doi: 10.1148/radiol.2443061520
9. Li CS, Chen RC, Tu HY, Shih LS, Zhang TA, Lii JM, et al. Imaging well-differentiated hepatocellular carcinoma with dynamic triple-phase helical computed tomography. *Br J Radiol*. (2006) 79:659–65. doi: 10.1259/bjr/12699987
10. Park MS, Kim S, Patel J, Hajdu CH, Do RK, Mannelli L, et al. Hepatocellular carcinoma: detection with diffusion-weighted versus contrast-enhanced magnetic resonance imaging in pretransplant patients. *Hepatology*. (2012) 56:140–8. doi: 10.1002/hep.25681
11. Vandecaveye V, De Keyzer F, Verslype C, Op de Beeck K, Komuta M, Topal B, et al. Diffusion-weighted MRI provides additional value to conventional dynamic contrast-enhanced MRI for detection of hepatocellular carcinoma. *Eur Radiol*. (2009) 19:2456–66. doi: 10.1007/s00330-009-1431-5
12. Taouli B, Koh DM. Diffusion-weighted MR imaging of the liver. *Radiology*. (2010) 254:47–66. doi: 10.1148/radiol.09090021
13. Kim YK, Kim CS, Han YM, Lee YH. Detection of liver malignancy with gadoxetic acid-enhanced MRI: is addition of diffusion-weighted MRI beneficial? *Clin Radiol*. (2011) 66:489–96. doi: 10.1016/j.crad.2010.09.007
14. Lee MH, Kim SH, Park MJ, Park CK, Rhim H. Gadoxetic acid-enhanced hepatobiliary phase MRI and high-b-value diffusion-weighted imaging to distinguish well-differentiated hepatocellular carcinomas from benign nodules in patients with chronic liver disease. *AJR Am J Roentgenol*. (2011) 197:W868–75. doi: 10.2214/AJR.10.6237
15. Ahn SS, Kim MJ, Lim JS, Hong HS, Chung YE, Choi JY. Added value of gadoxetic acid-enhanced hepatobiliary phase MR imaging in the diagnosis of hepatocellular carcinoma. *Radiology*. (2010) 255:459–66. doi: 10.1148/radiol.10091388
16. Kawada N, Ohkawa K, Tanaka S, Matsunaga T, Uehara H, Ioka T, et al. Improved diagnosis of well-differentiated hepatocellular carcinoma with gadolinium ethoxybenzyl diethylene triamine pentaacetic acid-enhanced magnetic resonance imaging and Sonazoid

contrast-enhanced ultrasonography. *Hepatol Res*. (2010) 40:930–6. doi: 10.1111/j.1872-034X.2010.00697.x
17. Kogita S, Imai Y, Okada M, Kim T, Onishi H, Takamura M, et al. Gd-EOB-DTPA-enhanced magnetic resonance images of hepatocellular carcinoma: correlation with histological grading and portal blood flow. *Eur Radiol*. (2010) 20:2405–13. doi: 10.1007/s00330-010-1812-9
18. Kitao A, Matsui O, Yoneda N, Kozaka K, Shinmura R, Koda W, et al. The uptake transporter OATP8 expression decreases during multistep hepatocarcinogenesis: correlation with gadoxetic acid enhanced MR imaging. *Eur Radiol*. (2011) 21:2056–66. doi: 10.1007/s00330-011-2165-8
19. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clin Radiol*. (2004) 59:1061–9. doi: 10.1016/j.crad.2004.07.008
20. Holli K, Laaperi AL, Harrison L, Luukkaala T, Toivonen T, Ryymin P, et al. Characterization of breast cancer types by texture analysis of magnetic resonance images. *Acad Radiol*. (2010) 17:135–41. doi: 10.1016/j.acra.2009.08.012
21. Wibmer A, Hricak H, Gondo T, Matsumoto K, Veeraraghavan H, Fehr D, et al. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. *Eur Radiol*. (2015) 25:2840–50. doi: 10.1007/s00330-015-3701-8
22. Li Z, Mao Y, Huang W, Li H, Zhu J, Li W, et al. Texture-based classification of different single liver lesion based on SPAIR T2W MRI images. *BMC Med Imaging*. (2017) 17:42. doi: 10.1186/s12880-017-0212-x
23. Mayerhoefer ME, Schima W, Trattnig S, Pinker K, Berger-Kulemann V, Ba-Ssalamah A. Texture-based classification of focal liver lesions on MRI at 3.0 Tesla: a feasibility study in cysts and hemangiomas. *J Magn Reson Imaging*. (2010) 32:352–9. doi: 10.1002/jmri.22268
24. Jirak D, Dezortova M, Taimr P, Hajek M. Texture analysis of human liver. *J Magn Reson Imaging*. (2002) 15:68–74. doi: 10.1002/jmri.10042
25. Zhou W, Zhang L, Wang K, Chen S, Wang G, Liu Z, et al. Malignancy characterization of hepatocellular carcinomas based on texture analysis of contrast-enhanced MR images. *J Magn Reson Imaging*. (2017) 45:1476–84. doi: 10.1002/jmri.25454
26. Choi SH, Lee JM, Yu NC, Suh KS, Jang JJ, Kim SH, et al. Hepatocellular carcinoma in liver transplantation candidates: detection with gadobenate dimeglumine-enhanced MRI. *AJR Am J Roentgenol*. (2008) 191:529–36. doi: 10.2214/AJR.07.2565
27. Abbasian Ardakani A, Gharbali A, Saniei Y, Mosarrezaii A, Nazarbaghi S. Application of texture analysis in diagnosis of multiple sclerosis by magnetic resonance imaging. *Glob J Health Sci*. (2015) 7:68–78. doi: 10.5539/gjhs.v7n6p68
28. International Working Party. Terminology of nodular hepatocellular lesions. *Hepatology*. (1995) 22:983–93 doi: 10.1016/0270-9139(95)90324-0
29. International Consensus Group for Hepatocellular Neoplasia The International Consensus Group for Hepatocellular Neoplasia. Pathologic diagnosis of early hepatocellular carcinoma: a report of the international consensus group for hepatocellular neoplasia. *Hepatology*. (2009) 49:658–64. doi: 10.1002/hep.22709
30. Kim JI, Lee JM, Choi JY, Kim YK, Kim SH, Lee JY, et al. The value of gadobenate dimeglumine-enhanced delayed phase MR imaging for characterization of hepatocellular nodules in the cirrhotic liver. *Invest Radiol*. (2008) 43:202–10. doi: 10.1097/RLI.0b013e31815d6929
31. Nasu K, Kuroki Y, Tsukamoto T, Nakajima H, Mori K, Minami M. Diffusion-weighted imaging of surgically resected hepatocellular carcinoma: imaging characteristics and relationship among signal intensity, apparent diffusion coefficient, and histopathologic grade. *AJR Am J Roentgenol*. (2009) 193:438–44. doi: 10.2214/AJR.08.1424
32. Xu PJ, Yan FH, Wang JH, Shan Y, Ji Y, Chen CZ. Contribution of diffusion-weighted magnetic resonance imaging in the characterization of hepatocellular carcinomas and dysplastic nodules in cirrhotic liver. *J Comput Assist Tomogr*. (2010) 34:506–12. doi: 10.1097/RCT.0b013e3181da3671
33. Shin SK, Kim YS, Choi SJ, Shim YS, Jung DH, Kwon OS, et al. Characterization of small (</=3 cm) hepatic lesions with atypical enhancement feature and hypointensity in hepatobiliary phase of gadoxetic acid-enhanced MRI in cirrhosis: a STARD-compliant article. *Medicine*. (2017) 96:e7278. doi: 10.1097/MD.0000000000007278
34. Park MJ, Kim YK, Lee MW, Lee WJ, Kim YS, Kim SH, et al. Small hepatocellular carcinomas: improved sensitivity by combining gadoxetic acid-enhanced and diffusion-weighted MR imaging patterns. *Radiology*. (2012) 264:761–70. doi: 10.1148/radiol.12112517
35. Raman SP, Schroeder JL, Huang P, Chen Y, Coquia SF, Kawamoto S, et al. Preliminary data using computed tomography texture analysis for the classification of hypervascular liver lesions: generation of a predictive model on the basis of quantitative spatial frequency measurements–a work in progress. *J Comput Assist Tomogr*. (2015) 39:383–95. doi: 10.1097/RCT.0000000000000217
36. Ganeshan B, Miles KA, Young RC, Chatwin CR. Texture analysis in non-contrast enhanced CT: impact of malignancy on texture in apparently disease-free areas of the liver. *Eur J Radiol*. (2009) 70:101–10. doi: 10.1016/j.ejrad.2007.12.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 31 January 2020
doi: 10.3389/fonc.2020.00053

Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region

Qiuchang Sun[1†], Xiaona Lin[2†], Yuanshen Zhao[1], Ling Li[3], Kai Yan[1,4], Dong Liang[1], Desheng Sun[2] and Zhi-Cheng Li[1*]*

[1] Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [2] Department of Ultrasonic Imaging, Peking University Shenzhen Hospital, Shenzhen, China, [3] Ultimage Lab, Suzhou, China, [4] Peng Cheng Laboratory, Shenzhen, China

OPEN ACCESS

Edited by:
Rong Tian,
Sichuan University, China

Reviewed by:
Laurence Gluch,
The Strathfield Breast Centre, Australia
Zhongxiang Ding,
Hangzhou First People's Hospital, China

***Correspondence:**
Zhi-Cheng Li
zc.li@siat.ac.cn
Desheng Sun
szdssun@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 01 September 2019
Accepted: 13 January 2020
Published: 31 January 2020

Citation:
Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, Sun D and Li Z-C (2020) Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region. Front. Oncol. 10:53. doi: 10.3389/fonc.2020.00053

Objective: Axillary lymph node (ALN) metastasis status is important in guiding treatment in breast cancer. The aims were to assess how deep convolutional neural network (CNN) performed compared with radiomics analysis in predicting ALN metastasis using breast ultrasound, and to investigate the value of both intratumoral and peritumoral regions in ALN metastasis prediction.

Methods: We retrospectively enrolled 479 breast cancer patients with 2,395 breast ultrasound images. Based on the intratumoral, peritumoral, and combined intra- and peritumoral regions, three CNNs were built using DenseNet, and three radiomics models were built using random forest, respectively. By combining the molecular subtype, another three CNNs and three radiomics models were built. All models were built on training cohort (343 patients 1,715 images) and evaluated on testing cohort (136 patients 680 images) with ROC analysis. Another prospective cohort of 16 patients was enrolled to further test the models.

Results: AUCs of image-only CNNs in both training/testing cohorts were 0.957/0.912 for combined region, 0.944/0.775 for peritumoral region, and 0.937/0.748 for intratumoral region, which were numerically higher than their corresponding radiomics models with AUCs of 0.940/0.886, 0.920/0.724, and 0.913/0.693. The overall performance of image-molecular CNNs in terms of AUCs on training/testing cohorts slightly increased to 0.962/0.933, 0.951/0.813, and 0.931/0.794, respectively. AUCs of both CNNs and radiomics models built on combined region were significantly better than those on either intratumoral or peritumoral region on the testing cohort ($p < 0.05$). In the prospective study, the CNN model built on combined region achieved the highest AUC of 0.95 among all image-only models.

Conclusions: CNNs showed numerically better overall performance compared with radiomics models in predicting ALN metastasis in breast cancer. For both CNNs and radiomics models, combining intratumoral, and peritumoral regions achieved significantly better performance.

Keywords: breast cancer, deep learning, radiomics, axillary lymph node metastasis, breast ultrasound, peritumoral region

INTRODUCTION

Breast cancer is the leading malignancy in females (1). Axillary lymph node (ALN) metastasis status is one of the most important factors in guiding treatment decision making in breast cancer (2). Traditionally, the nodal status was assessed by surgical methods such as sentinel lymph node biopsy (SLNB) and axillary lymph node dissection (ALND) (3). According to the guideline from American Society of Clinical Oncology, SLNB is considered to have a high overall accuracy ranging from 93 to 97.6% with a relatively low false negative rate (FNR) ranging from 4.6 to 16.7% in detecting axillary metastasis (4). However, these surgical approaches have been considered controversial due to the invasiveness, potential complications, and possible overtreatment (3–6).

Ultrasound is a widely-used tool in breast cancer assessment as it is non-invasive, radiation-free, real-time and well-tolerated in women. Previous studies have shown that axillary ultrasound (AUS) may provide useful information relevant to ALN status in breast cancer (7). However, AUS alone has moderate sensitivity and may not be a reliable predictor for nodal metastasis (7, 8). Recently, imaging-based machine learning approaches have been demonstrated promising in cancer diagnosis. There are two most popular machine learning approaches: radiomics analysis and convolutional neural networks (CNN). Radiomics analysis relies on a pipeline including extraction of numerous handcrafted imaging features, followed by feature selection and machine learning-based classification. Handcrafted radiomics features extracted from the breast tumor area have been demonstrated predictive in ALN metastasis, with FNRs ranging from 13.9 to 25% (9, 10). However, handcrafted features are limited to the current knowledge of medical imaging, which may limit the potential of the predictive model. Deep learning improves this handcrafted pipeline by automatically learning discriminative features directly from images. Recent studies have shown that deep CNN-based approaches can achieve state-of-the-art performance in lesion detection and cancer diagnosis (11–13). To our knowledge, no studies have assessed breast ultrasound-based CNN in predicting ALN status for breast tumor.

Most studies have focused on mining predictive imaging features within the tumor, while the surrounding tissues were ignored. Previous evidence has shown that the peritumoral region—the tumor-adjacent parenchyma immediately surrounding the tumor mass—may offer valuable outcome-associated information (14–16). Two recent studies have demonstrated that handcrafted imaging features from peritumoral region in Dynamic Contrast-Enhanced MRI (DCE-MRI) are associated with sentinel lymph node metastasis (9) and pathological complete response to neoadjuvant chemotherapy (17) in breast cancer. Here, we hypothesize that deep CNN built based on intra- and peritumoral regions in breast ultrasound could provide relevant information in predicting ALN status. We are interested in comparing the performance of deep CNNs and radiomics models. Additionally, breast cancer can be classified into different molecular subtypes with distinct prognosis and respond differently to specific therapies (18). Therefore, we further assessed if deep CNNs or radiomics models combining imaging features and molecular subtypes could offer improved accuracy.

In this hypothesis-driven study, we first developed deep CNNs and radiomics models based on intratumoral, peritumoral, and combined regions in breast ultrasound images for predicting ALN metastasis. We then aimed to find out how on each region deep CNNs performed compared with radiomics models.

MATERIALS AND METHODS

Study Population

The study was approved by the Ethics Committee of Peking University Shenzhen Hospital (PUSH). Informed consent was waved from all patients by the ethics committee of PUSH. From the pathology and radiology databases in PUSH, a retrospective search was performed to recruit female patients with breast cancer between January 2016 and December 2018. The inclusion criteria were patients (1) with histologically-confirmed primary breast cancer, (2) with pretreatment breast ultrasound images, (3) with known ALN metastasis status determined by final histopathology, (4) with known molecular subtypes, and (5) without neoadjuvant chemotherapy prior to SLNB or ALND. The exclusion rules were that patients (1) with very small region of interest in the ultrasound images (<100 pixels) and (2) without SLNB or ALND. Finally, 479 patients with 479 breast tumors (136 positive and 343 negative ALNs) were included in this study. This cohort was randomly divided into a training cohort of 359 patients and a testing cohort of 120 patients at a ratio 3:1. The patient recruitment pathway was shown in **Figure S1**.

The baseline clinical and histopathological data were derived from patient medical records, including age, histological grade, immunohistochemistry (IHC) results and ALN status (positive or negative). According to the 2017 St Gallen International Expert Consensus, each patient was classified into one of four molecular subtypes: human epidermal growth factor receptor-2 (HER2) positive, triple-negative, Luminal A, and Luminal B (18). The status of HER2, ER, progesterone receptors (PR) and Ki-67 was

assessed by IHC. Based on the IHC results, the subtype can be determined.

Ultrasound Image Acquisition

The breast ultrasound examinations were performed by breast radiologists in our center using the Hitachi Ascendus ultrasound system equipped with 13–3 MHz linear array transducers. The examinations and assessments were conducted according to the 5th edition of Breast Imaging Reporting and Data System (BI-RADS) presented by American College of Radiology (ACR) (19). The parameters were set as follows: depth, 4–5 cm; brightness gain, 10–25 dB; dynamic range, 70 dB; frame rate, 26 frame *per second*. Patients were placed in supine or lateral position. The field of view was set to have the pectoralis muscle at the deepest aspect of the image. The focal zone was adjusted to be centered at the lesion. Ultrasound images were acquired and documented into the Picture Archiving and Communication Systems (PACS). For each lesion, five images were selected from PACS by a breast radiologist (XL with 5 years' experience in breast radiology) and used in our study according to the following scheme: (1) an image along the longest axis of lesion. (2) an image orthogonal to the first image. (3) three images at other angles where the lesion was clearly presented. The five images together represented the ultrasonographic features of a 3D lesion from different angels. For all 479 patients, we finally obtained 2395 images in total, including 1715 images (343 patients) in the training cohort and 680 images (136 patients) in the testing cohort.

ROI Delineation

The tumor region in each ultrasound image was manually delineated using the ITK-SNAP software (http://www.itksnap.org) by one radiologist (XL) who were blinded to the clinical and histopathological data of patients. A second breast radiologist (DS with 12 years' experience in breast radiology) reviewed all the delineations. Any disagreement between the two raters was resolved by discussion and consensus. The peritumoral regions were obtained by dilating the delineated tumor contour by ∼5 mm based on a standard morphological dilation operation using an inhouse program implemented in Matlab 2016b (MathWorks, Natick, MA). For each ultrasound slice, three region of interest (ROI) images were finally obtained: the intratumor ROI, the peritumor ROI, and the combined ROI that merged the intratumor and the peritumoral regions. Examples of ultrasound slices overlapped with intratumoral and peritumoral ROIs for two patients were shown in **Figure 1**.

Deep Learning With DenseNet

Deep CNN can automatically learn discriminative features from imaging data by stacking multiple convolutional layers. Among different CNN variants, densely connected convolutional network (DenseNet) has shown superior classification performance as it strengthens feature propagation while reduces parameter number (20). This is accomplished by connecting each layer to every other layer in a feed-forward fashion with less computational complexity. Here, our model was built based on the standard DenseNet-121 (20). All ROI images were resized into 224 × 224. The resized ROI images were used as input and transformed through the chained convolutional layers, yielding a class probability vector as the prediction results. The network was trained from scratch with cross entropy loss function and Adam optimizer with a learning rate of 0.0001, a batch size of 16, and a regularization weight of 0.0001. In the training cohort, data augmentation approaches including random rotation, random shear and random zoom were employed before the training procedure to avoid possible overfitting. The network was implemented on Keras (https://keras.io/) with the TensorFlow library as the backend (https://www.tensorflow.org/). The architecture of the image-only CNN network was shown in **Figure 2**. The details of the convolutional network implementation can be found in **Table S1**.

Deep Learning-Based Predictive Model Building

For predicting the nodal status, three image-only CNN models, including the intratumoral CNN, the peritumoral CNN and the combined-region CNN, were built with the DenseNet based on the intratumor ROI images, the peritumor ROI images, and the combined ROI images, respectively. Furthermore, three corresponding image-molecular models were also built based on the DenseNet by using both ROI images and molecular subtype information as the network input. Specifically, the molecular subtype information was incorporated as input to the fully-connected layers of the DenseNet, as shown in **Figure 2**.

Radiomics Feature Extraction and Selection

For each ultrasound slice, 104 radiomics features were extracted from each of the three ROI areas by using an open-source toolbox named Pyradiomics (https://pyradiomics.readthedocs.io) (21). Three groups of features were extracted, including shape features, intensity features, and texture features, as summarized in **Table S2**. Eleven shape features describing the geometric characteristics of the ROI were extracted. Eighteen intensity features describing the first-order distribution of the ROI intensities were extracted. Seventy-five texture features were computed to describe the patterns, or the high-order distributions of the ROI intensities with five different methods, including the gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray level size zone matrix (GLSZM), gray level dependence matrix (GLDM), and neighborhood gray-tone difference matrix (NGTDM). The detailed definitions of the radiomics features used can be found in two articles (22, 23). Having high-dimensional radiomics features, feature selection is required to reduce the dimension and avoid overfitting. Here an efficient machine learning-based wrapper algorithm, Boruta, was used to select a subset of features that were relevant to the prediction outcome (24). Boruta evaluated feature relevance iteratively by comparing the importance of original features with that achieved by artificially added random features, yielding an all-relevant subset of features that was considered optimal for the classification task. Here we used the R package Boruta for Boruta feature selection.

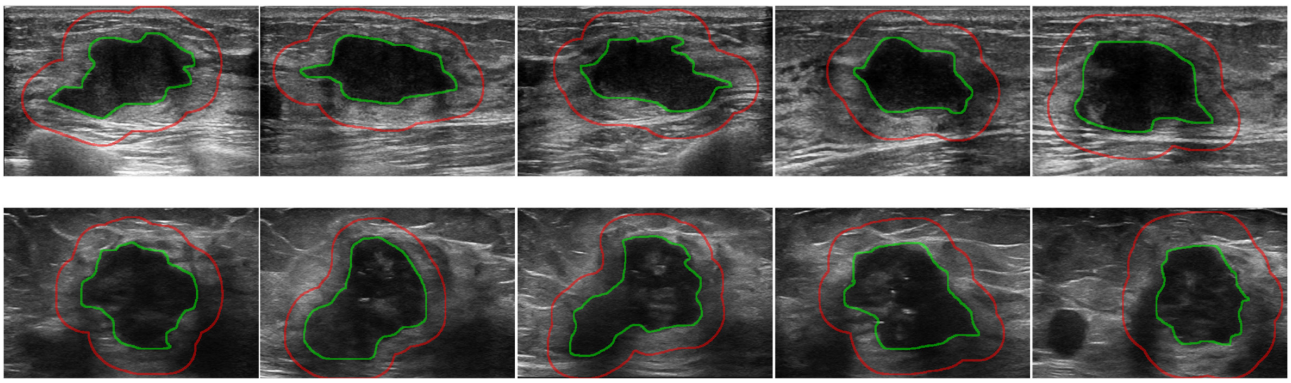


FIGURE 1 | Examples of ultrasound slices overlapped with intratumoral regions (green) and peritumoral regions (red) from two patients. **(Top)** A patient with positive ALN. **(Bottom)** A patient with negative ALN.

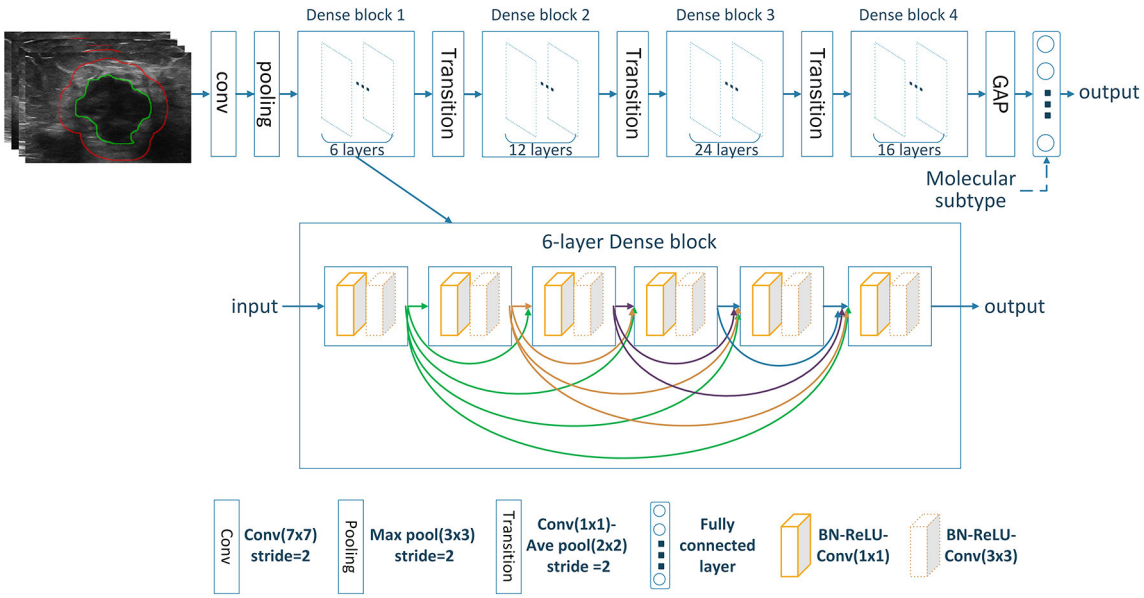


FIGURE 2 | The architecture of the deep CNN used in our study.

Radiomics-Based Predictive Model Building

Based on the selected radiomics features, three image-only radiomics models were built using random forest algorithm (25) based on the intratumor ROI, the peritumor ROI, and the combined ROI, respectively. Correspondingly, three image-molecular radiomics models were also built using random forest by integrating ROI images and molecular subtype information as the input. After testing different settings, the tree number of all random forest classifiers was set to 300. Gini index was used as importance measure (26). The R package randomForest was used for random forest classification.

Statistical Analysis

The difference in age, histological grades and molecular subtypes between training and testing cohorts was assessed with χ^2 test or Wilcoxon rank-sum test, where appropriate. All 12 prediction models (3 image-only CNNs, 3 image-only radiomics models, 3 image-molecular CNNs and 3 image-molecular radiomics models) were trained on the training cohort and evaluated on the testing cohort. Because each tumor had five ultrasound images, there were five corresponding prediction outcomes in the form of class probabilities. Among them, the median probability was chosen as the final prediction of each tumor and was used for statistical analysis. The prediction performance was assessed by the area under the receiver operating characteristic (ROC) curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV). The AUCs between two models were statistically compared using a DeLong test (27). All statistical analyses were performed with R software, version 3.5.1 (https://www.r-project.org/). All statistical tests were two sided, and $p < 0.05$ indicated significant.

RESULTS

Patient and tumor characteristics are summarized in **Table 1**. No significant difference was found in patient age, histological grades, molecular subtypes and ALN status between the training and testing cohorts ($p = 0.457$ to 0.844).

Image-Only Deep CNNs vs. Radiomics Models

The predictive performance of the three image-only deep CNNs and the three image-only radiomics models in both training and testing cohorts is summarized in **Table 2**. Their ROC curves in both training and testing cohorts are shown in **Figure 3**, respectively. The radiomics feature selection results can be found in **Table S3**. Among all six image-only models, the combined-region CNN achieved the best performance with a highest AUC of 0.912 and a highest accuracy of 89.3% in the testing cohort. In the testing cohort, the CNN built on each region performed better than the corresponding radiomics model built on the same region in terms of AUC and accuracy, but the differences of AUCs between the CNNs and their corresponding radiomics models were not statistically significant (Image-only CNN vs. Radiomics: Intratumoral: AUC 0.748 vs. 0.693, $p = 0.534$; Peritumoral: AUC 0.775 vs. 0.724, $p = 0.531$; Combined-region: AUC 0.912 vs. 0.886, $p = 0.601$).

Image-Molecular Deep CNNs vs. Radiomics Models

The performance of the three image-molecular CNNs and the three image-molecular radiomics models is summarized in **Table 3**. Their ROC curves in both training and testing cohorts are shown in **Figure 4**. From **Tables 2**, **3**, it can be found that the overall performance of the image-molecular models was slightly higher than those of their corresponding image-only models in the testing cohort, but no significant AUC differences were found between them. Among all 12 predictive models built in our study, the image-molecular CNN model built based on the combined-region achieved the best performance with a highest AUC of 0.933, a highest accuracy of 90.3% and a highest NPV of 0.958 in the testing cohort. All image-molecular CNNs achieved higher AUCs and higher accuracy than their corresponding radiomics models built based on the same tumoral region, but there were no significant differences between their AUCs (Image-molecular CNN vs. Radiomics: Intratumoral: AUC 0.794 vs. 0.706, $p = 0.308$; Peritumoral: AUC 0.813 vs. 0.743, $p = 0.334$; Combined-region: AUC 0.933 vs. 0.905, $p = 0.531$).

Assessment of Peritumoral and Intratumoral Regions

The predictive value of different tumoral regions were assessed by comparing the models built with the same machine learning methods (CNN or radiomics). It was observed that for the image-only CNNs and image-only radiomics models, the AUCs of the peritumoral models were slightly higher than those of the intratumoral models in the testing cohort, and their AUC differences were not significant (Image-only Peritumoral vs. Intratumoral: CNN: AUC 0.775 vs. 0.748, $p = 0.746$; Radiomics: AUC 0.724 vs. 0.693, $p = 0.707$). Similar results have been observed for the image-molecular models (Image-molecular Peritumoral vs. Intratumoral: CNN: AUC 0.813 vs. 0.794, $p = 0.806$; Radiomics: AUC 0.743 vs. 0.706, $p = 0.647$).

The image-only CNNs and image-only radiomics models built based on combined-region achieved higher AUCs than their corresponding models built based on either the intratumoral or peritumoral region in the testing cohort, where the AUC differences between them were significant (Image-only Combined-region vs. [Peritumoral, Intratumoral]: CNN: AUC 0.912 vs. [0.775, 0.748], [$p = 0.049$, $p = 0.031$]; Radiomics: AUC 0.886 vs. [0.724, 0.693], [$p = 0.014$, $p = 0.004$]). The image-molecular CNNs and image-molecular radiomics models built based on combined-region also achieved higher AUCs. For image-molecular models, the difference between AUCs of the combined-region CNN and either the intratumoral CNN or peritumoral CNN was significant (Image-molecular Combined-region vs. [Peritumoral, Intratumoral]: CNN: AUC 0.933 vs. [0.813, 0.794], [$p = 0.048$, $p = 0.046$]; Radiomics: AUC 0.905 vs. [0.743, 0.706], [$p = 0.006$, $p = 0.003$]).

Prospective Validation

To further validate the CNNs and radiomics models, we performed a validation study using a relatively small prospective cohort. From November 18 2019 to December 12 2019, 16 breast cancer patients (6 node positive and 10 node negative) with 80 breast ultrasound images (each had 5 images as described in section Ultrasound Image Acquisition) were finally enrolled for analysis. Age, grade, and node status were obtained for the 16 patients and were summarized in **Table 1**. All six image-only prediction models were tested. As we did not obtain IHC results, the image-molecular models were not tested. The model performance in this prospective cohort was summarized in **Table 4**. The ROC curves of all tested models were shown in **Figure S2**. We observed that the CNN built on the combined region achieved the highest AUC of 0.95 and the highest accuracy of 81.3%, where two patients with positive node and one patient with negative node were misclassified. In general, CNNs outperformed radiomics models; prediction models built on combined region outperformed those built on either intratumor region or peritumor region only. The results were consistent with previous observation on the retrospective cohort.

DISCUSSION

The major findings of this study were that deep CNN, built by combining intratumoral and peritumoral regions in breast ultrasound images, outperformed radiomics models in predicting ALN metastasis. Although imaging-based machine learning approaches have been demonstrated useful in assessing breast cancers, few studies have been done on evaluating the value of intra- and peritumoral regions in metastasis prediction (9), and no studies have investigated how breast ultrasound-based deep CNNs performed compared with radiomics models. In this study, we first developed three types of CNN models based on intratumoral, peritumoral, and combined regions, respectively in ultrasound images for assessing the nodal

TABLE 1 | A summary of patient and tumor characteristics of the study population.

Variables	All retrospective patients ($n = 479$)	Training cohort ($n = 359$)	Testing cohort ($n = 120$)	*P*-value	Prospective cohort ($n = 16$)	*P*-value
Age (mean ± SD)	48.7 ± 11.1	48.9 ± 10.9	47.9 ± 11.9	0.844	49.8 ± 11.3	0.680
Histological grade				0.755		0.556
I	187 (39.0%)	140 (39.0%)	47 (39.2%)		8 (50.0%)	
II	249 (52.0%)	190 (52.9%)	59 (49.2%)		7 (43.7%)	
III	43 (9.0%)	29 (8.1%)	14 (11.6%)		1 (6.3%)	
Molecular subtype				0.457	-	-
Luminal A	45 (9.4%)	33 (9.2%)	12 (10.0%)		-	
Luminal B	322 (67.2%)	239 (66.6%)	83 (69.2%)		-	
HER2 positive	57 (11.9%)	44 (12.3%)	13 (10.8%)		-	
Triple negative	55 (11.5%)	43 (11.9%)	12 (10.0%)		-	
ALN				0.829		0.418
Positive	136 (28.4%)	101 (28.1%)	35 (29.2%)		6 (37.5%)	
Negative	343 (71.6%)	258 (71.9%)	85 (70.8%)		10 (62.5%)	

P-values were calculated by using χ^2 test or Wilcoxon rank-sum test. P-values in the fifth column were calculated between training and testing cohorts. P-values in the seventh column were calculated between training and prospective cohorts.

TABLE 2 | A performance summary of the image−only CNNs and image−only radiomics models in training and testing cohorts in predicting ALN metastasis of breast cancer.

Model		Dataset	AUC	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)
Image−only CNN	Intra	Training	0.937 (0.908, 0.968)	84.6 –	95.7 (80.4, 98.9)	80.3 (76.6, 95.4)	65.2 (60.7, 87.8)	98.0 (92.5, 99.5)
		Testing	0.748 (0.622, 0.874)	71.8	76.0 (48.0, 92.0)	70.5 (60.3, 94.9)	45.2 (39.2, 77.8)	90.2 (84.0, 96.7)
	Peri	Training	0.944 (0.920, 0.969)	87.0 –	95.7 (89.1, 100.0)	83.7 (78.7, 91.6)	69.3 (63.5, 81.1)	98.0 (95.3, 100.0)
		Testing	0.775 (0.665, 0.886)	72.8 –	80.0 (56.0, 92.0)	70.5 (64.1, 92.3)	46.5 (40.9, 72.7)	91.7 (85.3, 96.8)
	Cmb	Training	0.957 (0.926, 0.989)	93.7 –	92.6 (86.2, 97.9)	94.1 (90.8, 99.6)	86.1 (80.0, 98.8)	97.0 (94.5, 99.1)
		Testing	0.912 (0.834, 99.0)	89.3 –	85.7 (67.9, 96.4)	90.7 (84.0, 100.0)	77.4 (66.7, 100.0)	94.4 (89.0, 98.6)
Image−only radiomics	Intra	Training	0.913 (0.870, 0.956)	87.9 –	84.8 (75.0, 91.3)	89.1 (86.6, 95.8)	75.0 (70.9, 88.6)	93.8 (90.6, 96.4)
		Testing	0.693 (0.573, 0.812)	68.9 –	56.0 (32.0, 100.0)	73.1 (28.2, 100.0)	40.0 (30.0, 100.0)	83.8 (81.0, 100.0)
	Peri	Training	0.920 (0.882, 0.958)	87.3 –	82.6 (72.8, 93.5)	89.1 (80.3, 96.7)	74.5 (63.6, 89.9)	93.0 (90.2, 97.0)
		Testing	0.724 (0.609, 0.839)	70.9 –	64.0 (48.0, 100.0)	73.1 (38.5, 91.0)	43.2 (32.4, 66.7)	86.4 (83.5, 100.0)
	Cmb	Training	0.940 (0.908, 0.973)	87.1 –	92.3 (81.3, 96.7)	85.2 (82.3, 95.9)	70.0 (66.1, 88.1)	96.7 (92.8, 98.6)
		Testing	0.886 (0.831, 0.942)	83.3 –	87.5 (72.5, 97.5)	81.8 (75.5, 91.8)	63.6 (55.7, 78.7)	94.7 (90.0, 98.9)

ACC, AUC, SEN, SPE, PPT, and NPV are short for accuracy, area under the receiver operating characteristic curve, sensitivity, specificity, positive prediction value, and negative prediction value, respectively. Intra, Peri and Cmb indicate the intratumoral model, the peritumoral model and the combined−region model, respectively. Statistical quantifications were demonstrated with 95% confidential interval (CI), when applicable.

metastasis, and further compared the performance of the three CNNs with three radiomics models built based on the same regions in nodal metastasis prediction. Moreover, we evaluated if further benefit can be obtained by integrating ultrasound images and molecular subtype information into the predictive models. Note that besides a high AUC, a high NPV is also important as accurately identifying patients with negative nodes [∼65% in all breast cancer patients (28)]

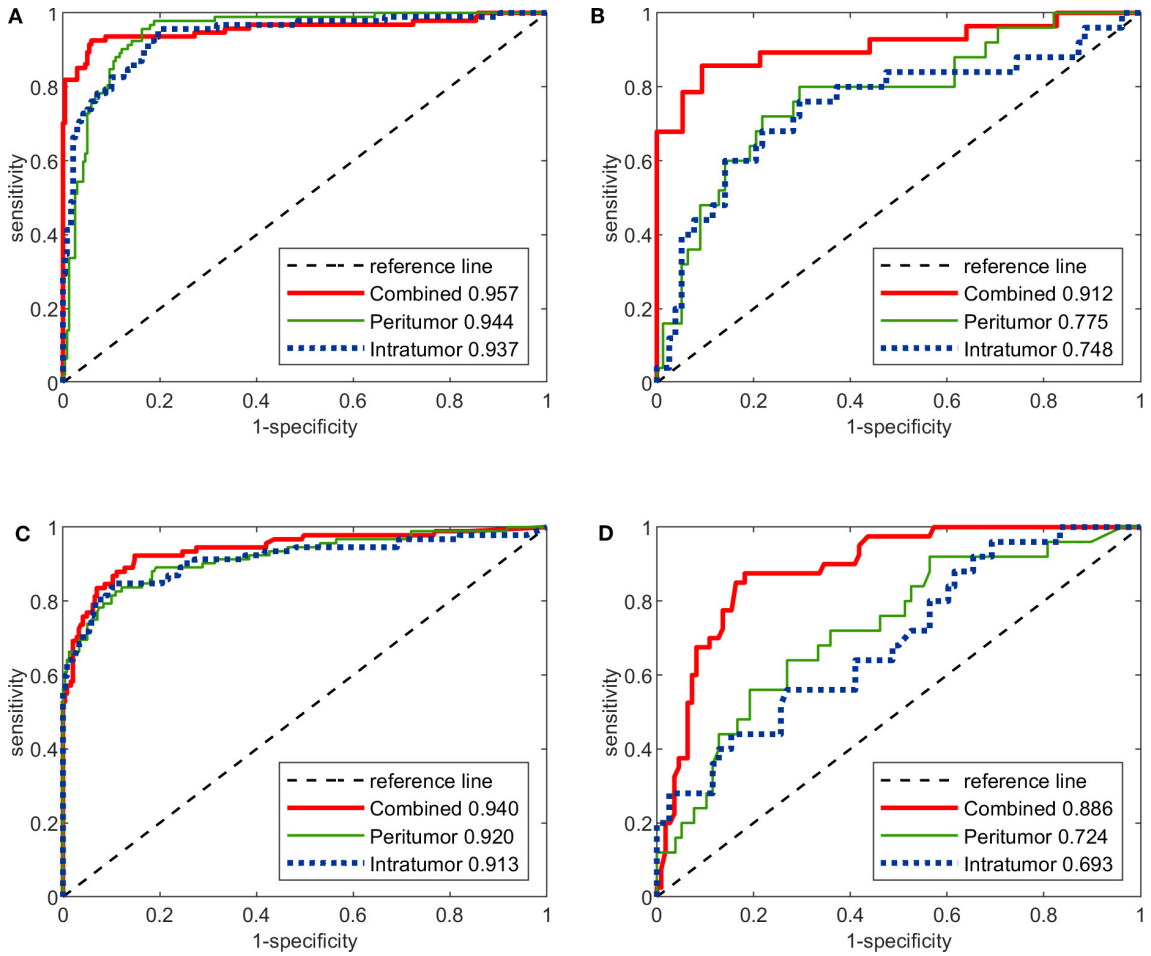


FIGURE 3 | The ROC curves of the three image-only deep CNNs and the three image-only radiomics models in both training and testing cohorts. **(A)** ROC curves of image-only CNNs in training cohort. **(B)** ROC curves of image-only CNNs in testing cohort. **(C)** ROC curves of image-only radiomics models in training cohort. **(D)** ROC curves of image-only radiomic models in testing cohort.

helps to avoid axillary overtreatment and reduce associated serious complications.

Identification of possible association between breast ultrasound features and ALN status has undoubtful clinical benefit. In clinical routine, the axilla can be staged clinically by palpation or surgically by SLNB or ALND. Although SLNB has less severe complications compared with ALND, it is not risk-free and SLNB- associated complications have been reported in large prospective trials (6). As palpation is inaccurate (29), AUS is performed to provide more relevant information. AUS alone has a reported sensitivity of 39–60%, specificity of 90–96%, PPV of 80–91%, and NPV of 75–83% (6, 30, 31). This implied that despite of an acceptable specificity above 90%, prior to surgery about 40–60% of nodal metastases cannot be found by AUS and about 20–25% of patients with a negative AUS have been assessed as modal metastases after surgery. In case of suspicious ALN, AUS alone or combined with ultrasound-guided needle biopsy is performed for axillary staging to select patients who would benefit from ALND. A recent meta-analysis has shown that the use of AUS in stratifying patients directly to fast-track ALND without SLNB leads to overtreatment in up to two-thirds of patients (32). These data indicated that AUS alone is not sufficiently accurate for axillary staging.

Recent studies have shown the value of radiomics features from primary lesion in predicting the lymph node metastasis for different cancer sites, e.g., CT radiomics features in colorectal cancer (33), MRI/CT radiomics features in bladder cancer (34, 35) and esophageal cancer (36). For breast cancer, two recent studies have assessed the value of radiomics features extracted from the primary tumor region at DCE-MRI and diffusion-weighted MRI (DWI) in predicting sentinel lymph node metastasis, where the reported AUC, sensitivity and specificity ranging from 0.805 to 0.869, 0.700–0.778, and 0.747–861 respectively (9, 10). In our study, we built three image-only radiomics models by using both peri- and intratumoral regions in multiple ultrasound slices per lesion. The combined-region radiomics model achieved an AUC of 0.886, a sensitivity of 87.5%

TABLE 3 | A performance summary of the image−molecular CNNs and image−molecular radiomics models in training and testing cohorts in predicting ALN metastasis of breast cancer.

Model		Dataset	AUC	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)
Image−molecular CNN	Intra	Training	0.931 (0.898, 0.964)	84.9 –	93.4 (78.2, 98.9)	81.7 (76.3, 96.3)	65.9 (60.3, 88.8)	97.0 (92.0, 99.5)
		Testing	0.794 (0.677, 0.911)	72.8 –	80.0 (48.0, 96.0)	70.5 (60.3, 100.0)	46.5 (40.3, 100.0)	91.7 (84.5, 98.0)
	Peri	Training	0.951 (0.928, 0.973)	88.5 –	95.7 (90.2, 100.0)	85.8 (79.9, 91.6)	72.1 (65.2, 81.1)	98.1 (96.1, 100.0)
		Testing	0.813 (0.717, 0.909)	75.7 –	88.0 (64.0, 100.0)	71.8 (64.1, 92.3)	50.0 (43.1, 74.1)	94.9 (88.2, 100.0)
	Cmb	Training	0.962 (0.933, 0.990)	92.8 –	93.5 (83.9, 97.8)	92.5 (89.6, 100.0)	82.9 (77.9, 100.0)	97.4 (94.1, 99.1)
		Testing	0.933 (0.864, 1.000)	90.3 –	89.3 (75.0, 100.0)	90.7 (82.7, 100.0)	78.1 (67.5, 100.0)	95.8 (91.0, 100.0)
Image−molecular radiomics	Intra	Training	0.931 (0.898, 0.964)	85.8 –	89.0 (78.0, 95.6)	84.6 (78.7, 94.6)	68.6 (61.5,84.8)	95.3 (91.7, 98.1)
		Testing	0.706 (0.583, 0.828)	71.8 –	64.0 (40.0, 88.0)	74.4 (46.1, 89.7)	44.4 (32.9, 66.7)	86.6 (81.1, 93.9)
	Peri	Training	0.916 (0.877, 0.955)	88.2 –	84.8 (76.1, 91.3)	89.5 (85.8, 95.4)	75.7 (70.0, 87.2)	93.9 (90.9, 96.5)
		Testing	0.743 (0.640, 0.847)	71.8 –	72.0 (60.0, 100.0)	71.8 (39.7, 85.9)	45.0 (33.8, 60.7)	88.9 (86.1, 100.0)
	Cmb	Training	0.950 (0.921, 0.980)	90.1 –	89.0 (81.3, 96.7)	90.5 (81.5, 96.7)	77.9 (65.9, 90.4)	95.7 (93.1, 98.7)
		Testing	0.905 (0.855, 0.956)	84.0 –	90.0 (77.5, 97.5)	81.8 (74.5, 93.6)	64.3 (56.5, 82.9)	95.7 (91.2, 98.9)

ACC, AUC, SEN, SPE, PPT, and NPV are short for accuracy, area under the receiver operating characteristic curve, sensitivity, specificity, positive prediction value, and negative prediction value, respectively. Intra, Peri and Cmb indicate the intratumoral model, the peritumoral model and the combined−region model, respectively. Statistical quantifications were demonstrated with 95% confidential interval (CI), when applicable.

and a specificity of 81.8% on the testing cohort, which were comparable with the previous radiomics models built with MRI.

Although promising, an efficient radiomics analysis heavily relies on a handcrafted image processing pipeline comprising three tightly coupled steps: feature extraction, feature selection and machine learning model building. Small variations in each stage may affect the prediction accuracy and stability (37). Deep CNN improves this pipeline by automatically learning predictive features on its own and yields a class probability vector as output. Currently, CNN-based learning methods have achieved diagnostic accuracy levels in skin cancer (11) and retinal diseases (12, 13), which have been unattainable via radiomics approaches. For breast cancer, a comparative study (38) demonstrated that CNN was superior to radiomic analysis in terms of a significantly higher AUC (0.88 vs. 0.81, $p < 0.001$) for classification of enhancing lesions as benign or malignant at MRI. Another comparative study in Kooi et al. (39) also demonstrated that CNN was superior to radiomics-based software in detection of mammographic breast lesions. In our study, all six CNNs (three image-only and three image-molecular) achieved higher AUC and accuracy than corresponding radiomics models built on the same regions on both training and testing cohorts. Note that in our results the differences between their AUCs (CNN vs. radiomics) were not significant (DeLong $p > 0.05$).

Most image analysis studies on breast cancer was focused on the intratumoral region. Evidences have demonstrated that imaging features of peritumoral regions can offer outcome-related information. Several studies have demonstrated that the enhancement patterns of tumor-adjacent parenchyma in DCE-MRI were associated with chemotherapy response (14), local recurrence (15), and survival (16) in breast cancer. In a recent study (40) the grade of peritumoral edema identified in breast MRI has been independently associated with disease recurrence. In study by Zhou et al. (41), it was demonstrated that the peritumoral stiffness assessed by ultrasound elastography of malignant breast lesions was higher than that of benign lesions. A 2017 study (17) was the first attempt to extract radiomics features from both intratumoral and peritumoral regions in breast DCE-MRI, where the features successfully predicted the pathological complete response to neoadjuvant chemotherapy. A more recent 2019 study (9) for the first time demonstrated the feasibility of predicting sentinel lymph node metastasis by using intratumoral and peritumoral radiomics features in DCE-MRI, achieving an AUC of 0.806 and an NPV of 82.4% with radiomics features only. Our study has shown the value of peritumoral ultrasonographic CNN features in predicting nodal metastasis with an AUC of 0.775 and an NPV of 91.6%. By combining both intra- and peritumoral regions, the CNN achieved a significantly better AUC of 0.912 and an NPV of 94.4%. The FNRs of the image-only CNN model built by combining the intra- and peritumoral regions achieved 5.9, 9.3, and 10% in the training, testing, and prospective data sets,

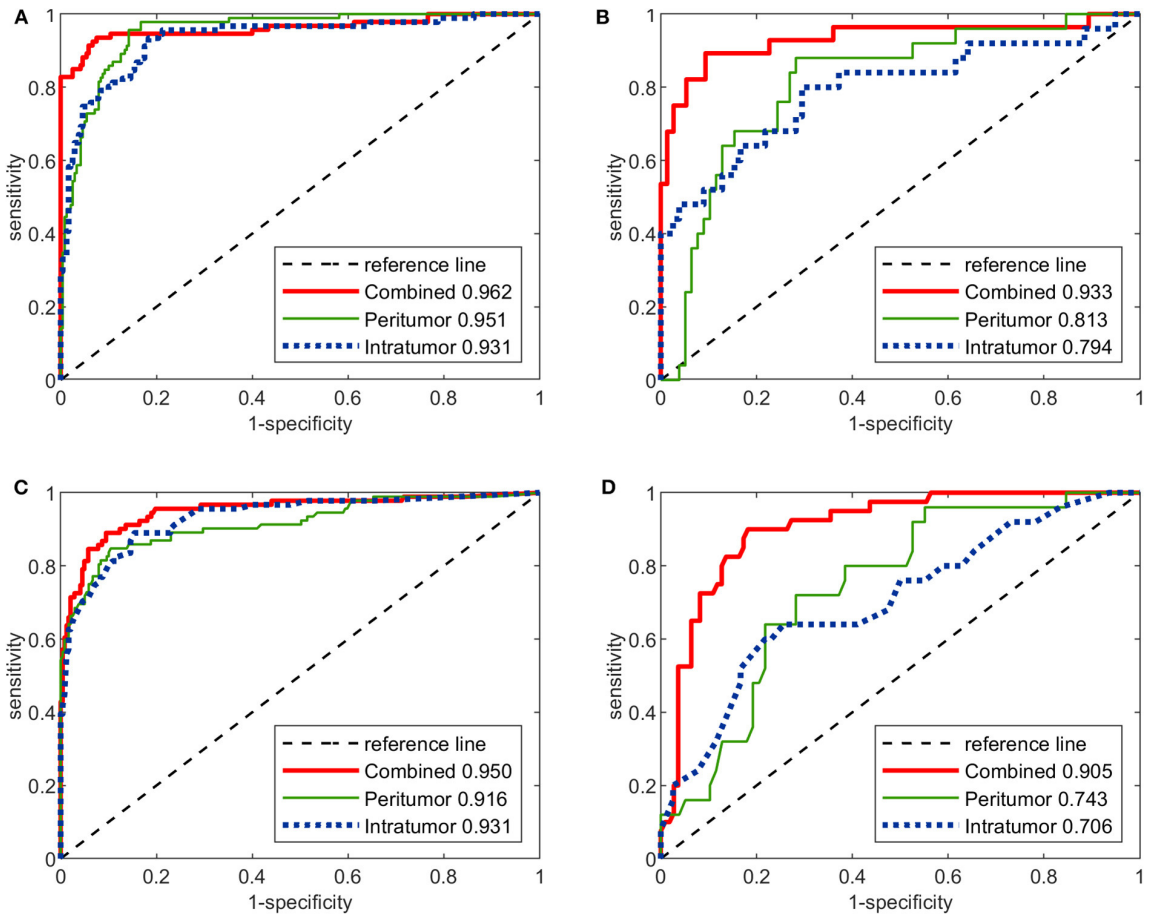


FIGURE 4 | The ROC curves of the three image-molecular deep CNNs and the three image-molecular radiomics models in both training and testing cohorts. **(A)** ROC curves of image-molecular CNNs in training cohort. **(B)** ROC curves of image-molecular CNNs in testing cohort. **(C)** ROC curves of image-molecular radiomics models in training cohort. **(D)** ROC curves of image-molecular radiomic models in testing cohort.

TABLE 4 | A performance summary of the image-only CNNs and image-only radiomics models in the prospective cohorts in predicting ALN metastasis of breast cancer.

Model		AUC	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)
Image-only CNN	Intra	0.767	75.6	50.0	90.0	75.0	75.0
	Peri	0.850	75.0	50.0	90.0	75.0	75.0
	Cmb	0.950	81.3	66.7	90.0	80.0	81.8
Image-only radiomics	Intra	0.533	68.8	33.3	90.0	66.7	69.2
	Peri	0.533	68.8	33.3	90.0	66.7	69.2
	Cmb	0.833	81.3	83.3	80.0	71.4	88.9

ACC, AUC, SEN, SPE, PPT, and NPV are short for accuracy, area under the receiver operating characteristic curve, sensitivity, specificity, positive prediction value, and negative prediction value, respectively. Intra, Peri and Cmb indicate the intratumoral model, the peritumoral model and the combined-region model, respectively.

respectively, which were superior to the image-only radiomics model with FNRs of 14.8, 18.25, and 20% in the training, testing, and prospective data sets, respectively. The FNRs of the CNN model were comparable with those of SLNB [4.6 to 16.7% (4)] and were higher than the radiomics models [13.9 to 25% (9, 10)] reported previously. By integrating the molecular subtype information, all the obtained image-molecular models, either CNN or radiomics, achieved slighter higher AUCs and NPVs.

The biological mechanism underlying the peritumoral imaging features and their association with clinical outcomes remains unclear. Many cancer studies have shown that biological changes in the tissue immediately surrounding the breast tumor mass might be potential predictive or prognostic

markers, such as peritumor lymphovascular invasion (42, 43), peritumoral lymphocytic infiltration (44), and peritumoral edema (45). In study by Zhao et al. (46) it was suggested that vascular endothelial growth factor (VEGF)-C/D induced peritumoral lymphangiogenesis may be one mechanism that leads to metastatic spread. In study by Wu et al. (16) the prognostic peritumoral features were associated with the tumor necrosis factor (TNF) signaling pathway that has been involved in oncogenic angiogenesis, invasion, and metastasis (47). Further studies are warranted to determine how the underlying biological changes were reflected by peritumor imaging features.

Our study has several limitations. The first limitation was the limited population size which may lead to bias. Larger patient population from more centers should be involved in future to improve the machined learning-based models. The population size of the prospective cohort is particularly small, where significant bias may occur. We will recruit more prospective data in future to further evaluate our methods in clinical practice. The second limitation was that all image data was obtained on the same type of ultrasound machine. In future we will evaluate our models on more heterogeneous image data acquired with different machines. Moreover, we built our CNNs and radiomics models using only ultrasound images and molecular subtypes. In future we will build more comprehensive models by incorporating more clinical and pathological data. Our future research also includes the exploring of biological mechanism underlying the association between intratumoral/peritumoral imaging features and nodal metastasis. We will also assess the possible incremental value of the tumoral ultrasonographic features over the AUS in axillary staging.

In conclusion, CNNs built on tumoral regions in ultrasound images allowed accurate prediction of ALN metastasis, which achieved higher AUC and NPV than radiomics models. Either CNNs or radiomic models built on peritumor regions performed slighter better than those built on intratumor regions, while combining both intra- and peritumoral regions achieved significantly better AUCs and higher NPVs. Further integrating the molecular subtype information into either CNNs or radiomics models can slightly benefit the performance.

DATA AVAILABILITY STATEMENT

To achieve repeatability, the data set of this study, including pretrained CNN models, imaging data of the prospective cohort, statistical analysis, and the Python implementation, was deposited into the Mendeley data library (https://data.mendeley.com/datasets/rc32mg38rb/draft?a=2333e5fd-e7b1-4603-b06e-b609d79bab11).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Peking University Shenzhen Hospital. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

DS, Z-CL, and DL conceived and designed the study. XL collected the clinical and image data and performed image pre-processing. QS, YZ, LL, and KY analyzed the image data and performed the statistical analysis. QS wrote the manuscript. All authors approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (no. 61571432) and Shenzhen Basic Research Program (JCYJ20170413162354654).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00053/full#supplementary-material

REFERENCES

1. Segal R, Miller K, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. (2018) 68:7–30. doi: 10.3322/caac.21442
2. Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ, et al. Breast cancer—major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *Cancer J Clin*. (2017) 67:290–303. doi: 10.3322/caac.21393
3. Giuliano AE, Hunt KK, Ballman KV, Beitsch PD, Whitworth PW, Blumencranz PW, et al. Axillary dissection vs. no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial. *JAMA*. (2011) 305:569–75. doi: 10.1001/jama.2011.90
4. Lyman GH, Temin S, Edge SB, Newman LA, Turner RR, Weaver DL, et al. Sentinel lymph node biopsy for patients with early-stage breast cancer: American Society of Clinical Oncology clinical practice guideline update. *J Clin Oncol*. (2014) 32:1365–83. doi: 10.1200/JCO.2013.54.1177
5. Lucci A, McCall LM, Beitsch PD, Whitworth PW, Reintgen DS, Blumencranz PW, et al. Surgical complications associated with sentinel lymph node dissection (SLND) plus axillary lymph node dissection compared with SLND alone in the American College of Surgeons Oncology Group Trial Z0011. *J Clin Oncol*. (2007) 25:3657–63. doi: 10.1200/JCO.2006.07.4062
6. Wilke LG, McCall LM, Posther KE, Whitworth PW, Reintgen DS, Leitch AM, et al. Surgical complications associated with sentinel lymph node biopsy: results from a prospective international cooperative group trial. *Ann Surg Oncol*. (2006) 13:491–500. doi: 10.1245/ASO.2006.05.013
7. Feng Y, Huang R, He Y, Lu A, Fan Z, Fan T, et al. Efficacy of physical examination, ultrasound, and ultrasound combined with fine-needle aspiration for axilla staging of primary breast cancer. *Breast Cancer Res Treat*. (2015) 149:761–5. doi: 10.1007/s10549-015-3280-z
8. Ahmed M, Douek M. Is axillary ultrasound imaging necessary for all patients with breast cancer? *Br J Surg*. (2018) 105:930–2. doi: 10.1002/bjs.10784

9. Liu C, Ding J, Spuhler K, Gao Y, Serrano Sosa M, Moriarty M, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer by radiomic signatures from dynamic contrast-enhanced MRI. *J Magnet Reson Imaging*. (2019) 49:131–40. doi: 10.1002/jmri.26224
10. Dong Y, Feng Q, Yang W, Lu Z, Deng C, Zhang L, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI. *Eur Radiol*. (2018) 28:582–91. doi: 10.1007/s00330-017-5005-7
11. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. (2017) 542:115. doi: 10.1038/nature21056
12. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. (2018) 172:1122–31. e1129. doi: 10.1016/j.cell.2018.02.010
13. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. (2018) 24:1342. doi: 10.1038/s41591-018-0107-6
14. Hattangadi J, Park C, Rembert J, Klifa C, Hwang J, Gibbs J, et al. Breast stromal enhancement on MRI is associated with response to neoadjuvant chemotherapy. *Am J Roentgenol*. (2008) 190:1630–6. doi: 10.2214/AJR.07.2533
15. Kim S-A, Cho N, Ryu EB, Seo M, Bae MS, Chang JM, et al. Background parenchymal signal enhancement ratio at preoperative MR imaging: association with subsequent local recurrence in patients with ductal carcinoma in situ after breast conservation surgery. *Radiology*. (2013) 270:699–707. doi: 10.1148/radiol.13130459
16. Wu J, Li B, Sun X, Cao G, Rubin DL, Napel S, et al. Heterogeneous enhancement patterns of tumor-adjacent parenchyma at MR imaging are associated with dysregulated signaling pathways and poor survival in breast cancer. *Radiology*. (2017) 285:401–13. doi: 10.1148/radiol.2017162823
17. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res*. (2017) 19:57. doi: 10.1186/s13058-017-0862-1
18. Curigliano G, Burstein HJP, Winer E, Gnant M, Dubsky P, Loibl S, et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Ann Oncol*. (2017) 28:1700–12. doi: 10.1093/annonc/mdx308
19. Sickles EA, D'Orsi CJ, Bassett LW, Appleton CM, Berg WA, Burnside ES. ACR BI-RADS® Atlas, Breast imaging reporting and data system. Reston, VA: American College of Radiology (2013). P. 39–48.
20. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI (2017).
21. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
22. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. (2014) 5:4006. doi: 10.1038/ncomms5006
23. Lambin P, Leijenaar RT, Deist TM, Peerlings J, De Jong EE, Van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. (2017) 14:749. doi: 10.1038/nrclinonc.2017.141
24. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw*. (2010) 36:1–13. doi: 10.18637/jss.v036.i11
25. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324
26. Louppe G, Wehenkel L, Sutera A, Geurts P. Understanding variable importances in forests of randomized trees. *Adv Neural Informat Process Syst*. (2013) 1:431–9.
27. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. (1988) 44:837–45. doi: 10.2307/2531595
28. Kuijs V, Moossdorff M, Schipper R, Beets-Tan R, Heuts E, Keymeulen K, et al. The role of MRI in axillary lymph node imaging in breast cancer patients: a systematic review. *Insights Into Imaging*. (2015) 6:203–15. doi: 10.1007/s13244-015-0404-2
29. Lanng C, Hoffmann J, Galatius H, Engel U. Assessment of clinical palpation of the axilla as a criterion for performing the sentinel node procedure in breast cancer. *Eur J Surg Oncol*. (2007) 33:281–4. doi: 10.1016/j.ejso.2006.09.032
30. Bailey A, Layne G, Shahan C, Zhang J, Wen S, Radis S, et al. Comparison between ultrasound and pathologic status of axillary lymph nodes in clinically node-negative breast cancer patients. *Am Surg*. (2015) 81:865–9.
31. Helfgott R, Mittlboeck M, Miesbauer M, Moinfar F, Haim S, Mascherbauer M, et al. The influence of breast cancer subtypes on axillary ultrasound accuracy: a retrospective single center analysis of 583 women. *Eur J Surg Oncol*. (2018) 45:538–43. doi: 10.1016/j.ejso.2018.10.001
32. Ahmed M, Jozsa F, Baker R, Rubio I, Benson J, Douek M. Meta-analysis of tumour burden in pre-operative axillary ultrasound positive and negative breast cancer patients. *Breast Cancer Res Treatment*. (2017) 166:329–36. doi: 10.1007/s10549-017-4405-3
33. Huang Y, Liang C, He L, Tian J, Liang C, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
34. Wu S, Zheng J, Li Y, Yu H, Shi S, Xie W, et al. A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer. *Clin Cancer Res*. (2017) 23:6904–11. doi: 10.1158/1078-0432.CCR-17-1510
35. Wu S, Zheng J, Li Y, Wu Z, Shi S, Huang M, et al. Development and validation of an MRI-based radiomics signature for the preoperative prediction of lymph node metastasis in bladder cancer. *EBioMedicine*. (2018) 34:76–84. doi: 10.1016/j.ebiom.2018.07.029
36. Qu J, Shen C, Qin J, Wang Z, Liu Z, Guo J, et al. The MR radiomic signature can predict preoperative lymph node metastasis in patients with esophageal cancer. *Eur Radiol*. (2019) 29:906–14. doi: 10.1007/s00330-018-5583-z
37. Li Q, Bai H, Chen Y, Sun Q, Liu L, Zhou S, et al. A fully-automatic multiparametric radiomics model: towards reproducible and prognostic imaging signature for prediction of overall survival in glioblastoma multiforme. *Sci Rep*. (2017) 7:14331. doi: 10.1038/s41598-017-14753-7
38. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C. Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast, MRI. *Radiology*. (2018) 290:290–7. doi: 10.1148/radiol.2018181352
39. Kooi T, Litjens G, Van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal*. (2017) 35:303–12. doi: 10.1016/j.media.2016.07.007
40. Cheon H, Kim HJ, Kim TH, Ryeom H-K, Lee J, Kim GC, et al. Invasive breast cancer: Prognostic value of peritumoral edema identified at preoperative MR imaging. *Radiology*. (2018) 287:68–75. doi: 10.1148/radiol.2017171157
41. Zhou J, Zhan W, Dong Y, Yang Z, Zhou C. Stiffness of the surrounding tissue of breast lesions evaluated by ultrasound elastography. *Eur Radiol*. (2014) 24:1659–67. doi: 10.1007/s00330-014-3152-7
42. Schoppmann SF, Bayer G, Aumayr K, Taucher S, Geleff S, Rudas M, et al. Prognostic value of lymphangiogenesis and lymphovascular invasion in invasive breast cancer. *Ann Surg*. (2004) 240:306–12. doi: 10.1097/01.sla.0000133355.48672.22
43. Ejlertsen B, Jensen M-B, Rank F, Rasmussen BB, Christiansen P, Kroman N, et al. Population-based study of peritumoral lymphovascular invasion and outcome among patients with operable breast cancer. *J Natl Cancer Inst*. (2009) 101:729–35. doi: 10.1093/jnci/djp090

44. Ocaña A, Diez-Gónzález L, Adrover E, Fernández-Aramburo A, Pandiella A, Amir E. Tumor-infiltrating lymphocytes in breast cancer: ready for prime time? *J Clin Oncol*. (2015) 33:1298–9. doi: 10.1200/JCO.2014.59.7286
45. Uematsu T. Focal breast edema associated with malignancy on T2-weighted images of breast MRI: peritumoral edema, prepectoral edema, and subcutaneous edema. *Breast Cancer*. (2015) 22:66–70. doi: 10.1007/s12282-014-0572-9
46. Zhao Y-C, Ni X-J, Li Y, Dai M, Yuan Z-X, Zhu Y-Y, et al. Peritumoral lymphangiogenesis induced by vascular endothelial growth factor C and D promotes lymph node metastasis in breast cancer patients. *World J Surg Oncol*. (2012) 10:165. doi: 10.1186/1477-7819-10-165
47. Balkwill F. TNF-α in promotion and progression of cancer. *Cancer Metastasis Rev*. (2006) 25:409. doi: 10.1007/s10555-006-9005-3

Conflict of Interest: LL was employed by the company Ultimage Lab.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 31 January 2020
doi: 10.3389/fonc.2020.00043

Correction for Magnetic Field Inhomogeneities and Normalization of Voxel Values Are Needed to Better Reveal the Potential of MR Radiomic Features in Lung Cancer

Maxime Lacroix[1,2], Frédérique Frouin[2], Anne-Sophie Dirand[2], Christophe Nioche[2], Fanny Orlhac[2], Jean-François Bernaudin[3], Pierre-Yves Brillet[1] and Irène Buvat[2]*

[1] Service d'Imagerie Médicale, AP-HP, Hôpital Avicenne, Bobigny, France, [2] Laboratoire IMIV, UMR 1023 Inserm-CEA-Université Paris Sud, ERL 9218 CNRS, Université Paris Saclay, Orsay, France, [3] INSERM UMR 1272 Hypoxie et Poumon, Université Paris 13, Bobigny, France

OPEN ACCESS

Edited by:
Xuelei Ma,
Sichuan University, China

Reviewed by:
Karen Drukker,
University of Chicago, United States
Ahmad Chaddad,
McGill University Health Centre, Canada
Satish E. Viswanath,
Case Western Reserve University, United States

***Correspondence:**
Frédérique Frouin
frederique.frouin@inserm.fr

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 18 September 2019
Accepted: 10 January 2020
Published: 31 January 2020

Citation:
Lacroix M, Frouin F, Dirand A-S, Nioche C, Orlhac F, Bernaudin J-F, Brillet P-Y and Buvat I (2020) Correction for Magnetic Field Inhomogeneities and Normalization of Voxel Values Are Needed to Better Reveal the Potential of MR Radiomic Features in Lung Cancer. Front. Oncol. 10:43. doi: 10.3389/fonc.2020.00043

Purpose: To design and validate a preprocessing procedure dedicated to T2-weighted MR images of lung cancers so as to improve the ability of radiomic features to distinguish between adenocarcinoma and other histological types.

Materials and Methods: A discovery set of 52 patients with advanced lung cancer who underwent T2-weighted MR imaging at 3 Tesla in a single center study from August 2017 to May 2019 was used. Findings were then validated using a validation set of 19 additional patients included from May to October 2019. Tumor type was obtained from the pathology report after trans-thoracic needle biopsy, metastatic lymph node or metastasis samples, or surgical excisions. MR images were preprocessed using N4ITK bias field correction and by normalizing voxel intensities with fat as a reference region. Segmentation and extraction of radiomic features were performed with LIFEx software on the raw images, on the N4ITK-corrected images and on the fully preprocessed images. Two analyses were conducted where radiomic features were extracted: (1) from the whole tumor volume (3D analysis); (2) from all slices encompassing the tumor (2D analysis). Receiver operating characteristic (ROC) analysis was used to identify features that could distinguish between adenocarcinoma and other histological types. Sham experiments were also designed to control the number of false positive findings.

Results: There were 31 (12) adenocarcinomas and 21 (7) other histological types in the discovery (validation) set. In 2D, preprocessing increased the number of discriminant radiomic features from 8 without preprocessing to 22 with preprocessing. 2D analysis yielded more features able to identify adenocarcinoma than 3D analysis (12 discriminant radiomic features after preprocessing in 3D). Preprocessing did not increase false positive findings as no discriminant features were identified in any of the sham experiments. The greatest sensitivity of the 2D analysis applied to preprocessed data was confirmed in the validation set.

Conclusion: Correction for magnetic field inhomogeneities and normalization of voxel values are essential to reveal the full potential of radiomic features to identify the tumor histological type from MR T2-weighted images, with classification performance similar to those reported in PET/CT and in multiphase CT in lung cancers.

Keywords: lung cancer, radiomics, histological types of lung cancer, T2-weighted MR images, bias field correction, MRI normalization

INTRODUCTION

Radiomics consists in the extraction of a large number of quantitative features from radiology images to describe the shape, intensity distribution, and texture characteristics of a region of interest (1–3). The assumption is that such image-derived features can outperform visual analysis to characterize abnormalities. In particular, in oncology, radiomic features might reflect tumor heterogeneity observed at the histological and genetic levels (4). Macroscopic structural heterogeneity can unveil differences in tumor biology, which cannot be identified by clinical data alone (5, 6). In medical images, the macroscopic heterogeneity corresponds to variations of image intensities between neighboring voxels, which are described by radiomic features. Radiomic features are thus expected to be related to the phenotype, genotype and microenvironment of the tumor, and thus be of interest to support therapeutic decisions (7). Radiomics is therefore largely investigated to assist cancer diagnosis, prognosis, and prediction of response to therapy (8, 9).

Many radiomic studies have been devoted to lung cancer, which is a major public health problem (10–16). These studies mostly focus on nodules detected on CT and/or PET scans and that can be removed surgically (8, 17, 18). CT and PET are indeed used in daily practice for managing lung cancer patients. MR images are also of interest to characterize tumors because of their excellent contrast (19–21). Yet, to the best of our knowledge, very few studies have investigated the usefulness of MR radiomic features in lung cancer patients. A study defined the optimal timing to extract radiomic features on T1-weighted images after contrast medium injection in order to predict 2-years progression-free survival (15). Another preliminary study suggested that MR-derived radiomic features (based on True Fast MR images with a Steady State Precession sequence) may improve the accuracy of models that predict the response to therapy and survivals at different time points compared to that of models based on CT features only (22). No radiomic study involving anatomical MR sequences has been reported. Predicting histology from functional MR data was reported in a meta-analysis (23) showing that diffusion MR sequences could distinguish between malignant and benign lung lesions. It was also suggested that small cell cancers had significantly lower Apparent Diffusion Coefficients than other subtypes (23). However, these functional MR sequences could not differentiate adenocarcinomas from squamous cell carcinomas.

CT intensities are expressed in Hounsfield Units (HUs) linearly related to the tissue attenuation coefficients at the energy of the CT scanner. PET images are expressed in Standardized Uptake Values (SUV) that are directly related to the tracer concentration. As a result, CT or PET image values have the same meaning, from a physics point of view, in all patient scans acquired with the same scanner and using the same protocol for the image acquisition and reconstruction. Still, the use of different scanners and/or different image acquisition and reconstruction protocols introduces some variability in voxel values hence in radiomic features (24) and some harmonization techniques have been proposed to realign radiomic features measured in different conditions (25). In anatomical MR sequences, images are initially expressed in arbitrary units, meaning that a given tissue type (for instance fat) will not always yield a similar voxel value, even when the images are acquired in the same patient and same conditions using the same scanner (26). Therefore, a measured voxel value cannot be readily interpreted in terms of well-understood physics quantity unlike in CT (a HU does correspond to a unique attenuation coefficient) and in PET (an SUV corresponds to a unique tracer concentration in a given patient). This makes radiomic studies more challenging in MR compared to PET and CT.

In that context, the present study had three objectives: (1) to design and validate a new normalization procedure dedicated to T2-weighted MR images of lung cancer patients, using subcutaneous fat as a reference tissue; (2) to perform a systematic comparison of a 3D analysis of radiomic features with a 2D analysis taking into account all slices; (3) to demonstrate the usefulness of the normalization procedure and the 2D analysis to identify relevant T2-weighted MR radiomic features for differentiating adenocarcinoma from other types of lung cancer.

MRI preprocessing, posterior to acquisitions, including magnetic bias field correction and normalization has been successfully applied to brain studies using white matter as the reference tissue (26, 27). This approach has been rarely applied to other organs. For prostate T2w images, a recent study considered muscle as the reference tissue, reporting mitigated results in terms of feature reproducibility (28). To the best of our knowledge, this normalization based on a reference tissue has never been used for lung MR images. Unlike many studies comparing 3D and 2D radiomic features where the 2D approach only exploits the slice presenting the largest tumor area or diameter, our 2D analysis calculates radiomic features in all slices and selects the median value of each feature as being representative of the tumor.

MATERIALS AND METHODS

Population

Patient imaging, pathology, and clinical data were selected from the single-center MRI-omics database built as part of a retrospective study approved by the Institutional Review Board

(protocol 32-2016, study number: 2016-A00813-48). All patients gave their written informed consents. From August 2017 to May 2019, patients with advanced lung cancer referred for brain MR imaging to detect cerebral metastases were proposed to undergo additional lung MRI sequences. Among 83 eligible patients, 8 were not included because the pathology of the tumor was not available, and 23 were excluded because of motion artifacts or incomplete MR protocol (**Figure 1**). The final population thus included 52 patients (34 men and 18 women) from 44 to 89 years old (mean age 66 years; SD 11.3 years). Among these patients, 42 subjects (80%) were active smokers and 2 (4%) were exposed to asbestos. These 52 patients defined the discovery set of the current study. In addition, the following 21 patients enrolled between May and October 2019 were used to create a validation set. Two were excluded because of inconsistent slice thickness. This additional population included 17 men and 2 women from 41 to 85 years old (mean age 71 years; SD 12.1 years). Among them, 13 (68%) were active smokers and 1 (5%) exposed to asbestos.

MRI Acquisition

All acquisitions were performed with a 3T MRI unit (Discovery MR750, GE Healthcare, Waukesha—WI, USA), using an 18-channel phased-array body coil on the thorax. All study participants were scanned in the supine position with the arms along the body. All patients had a T2-weighted (T2w) sequence with a vendor-specific implementation of the periodically rotated overlapping parallel lines with enhanced reconstruction technique (PROPELLER) acquired in free breathing. The T2w PROPELLER sequence was selected since it provided few motion artifacts (29) and a good image quality (30, 31). The main parameters of the sequence are given in **Table 1**.

Pathological Assessments of Tumor Samples

A dedicated pathologist (JB, with more than 30 years of experience in lung cancer pathology), blinded to the MR findings,

TABLE 1 | Parameters of MR images acquisition protocols.

Parameter	T2w PROPELLER
Plane	Axial
TR (ms)	9,677
TE (ms)	96
FA (degree)	160
FOV (mm)	500 × 500
Matrix	240 × 240
Slice thickness (mm)	4
Inter slice spacing (mm)	0
Frequence	384
NEx	1.5
Gating	Respiratory
Breath hold	No
Acquisition time (s)	65

TR, repetition time; TE, echo time; FA, flip angle; FOV, field of view; NEx, number of excitations.

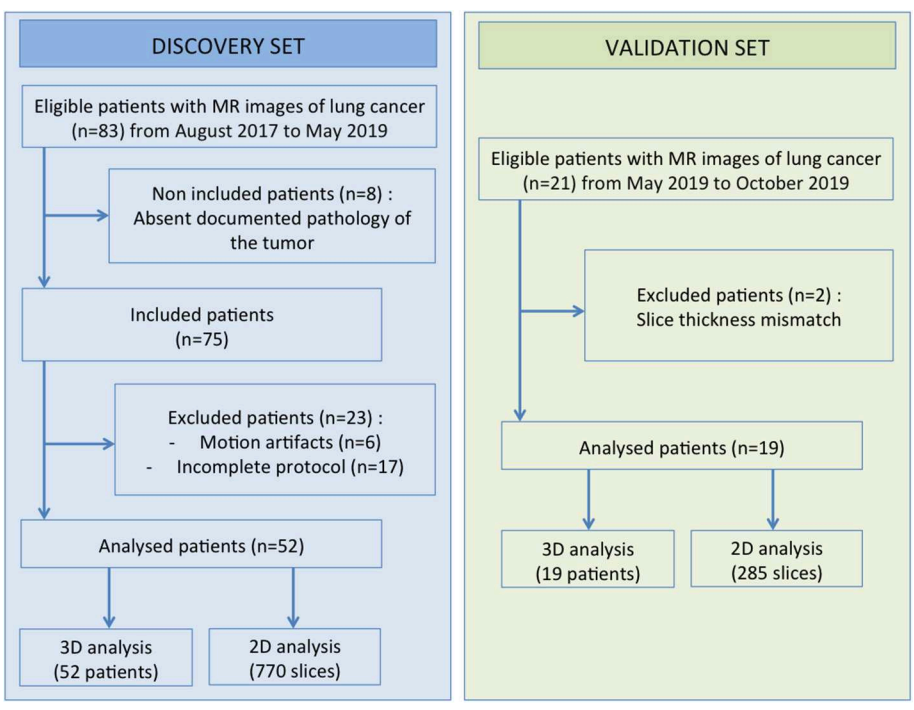


FIGURE 1 | Data selection pipelines.

reviewed all pathology reports and filled a structured pathology worksheet. The pathology information came from trans-thoracic core needle biopsy ($n = 31$ in the discovery set, $n = 11$ in the validation set), metastatic lymph node samples ($n = 11$ in the discovery set, $n = 3$ in the validation set), other metastatic location samples ($n = 6$ in the discovery set, $n = 4$ in the validation set), and surgical excisions ($n = 4$ in the discovery set, $n = 1$ in the validation set). Detailed characteristics are reported in **Table 2**.

Based on the pathology worksheet, two groups were defined: the first group included all patients with adenocarcinoma and the second group included all other patients.

MR Image Analysis

Lung MR images were retrieved from the Pictorial Archive and Communication System (Carestream 3.2. Carestream Health, Rochester, New York), anonymized and loaded in a workstation for radiomic analysis. Preprocessing of images included two steps: a correction for magnetic field (B1) inhomogeneity in order to reduce the signal intensity variation across the field of view, followed by a normalization of intensities based on the delineation of a reference tissue. Tumors were then segmented and 3D and 2D radiomic features were extracted for raw data, N4ITK corrected data, and normalized N4ITK-corrected data. Statistical analyses were performed to identify the features that could distinguish between the group of adenocarcinoma and the group including other tumor types.

TABLE 2 | Tumor characteristics.

Population	Discovery set	Validation set
Number of cases	52	19
Type of tumor: *n* (%)		
Adenocarcinoma	31 (60%)	12 (63%)
Other types	21 (40%)	7 (37%)
Squamous cell carcinoma	16 (76%)	4 (57%)
Small cell carcinoma	2 (9.5%)	1 (14%)
Sarcomatoid tumor	2 (9.5%)	2 (29%)
Large cell carcinoma	1 (5%)	0
Mean size in long axis (mm)	63.4 ± 23.2 (Range: 23–110)	67.7 ± 21.1 (Range: 27–109)
Location: *n* (%)		
Right upper lobe	24 (46%)	10 (53%)
Middle lobe	3 (6%)	0
Right lower lobe	8 (15%)	4 (21%)
Left upper lobe	10 (19%)	3 (16%)
Left lower lobe	7 (14%)	2 (10%)
T status (Lung-cancer TNM 8th edition): *n* (%)		
T1	3 (6%)	1 (5%)
T2	4 (8%)	2 (11%)
T3	10 (19%)	5 (26%)
T4	35 (67%)	11 (58%)
Invasion		
No parietal or mediastinal invasion	17 (33%)	6 (32%)
Parietal invasion	18 (35%)	4 (21%)
Mediastinal invasion	13 (25%)	8 (42%)
Parietal and mediastinal invasion	4 (7%)	1 (5%)

Correction of Magnetic Field Inhomogeneity

Magnetic field inhomogeneity artifacts were corrected based on the estimation of a bias field constrained to be spatially smooth (32). The bias field was estimated with the publicly available N4ITK algorithm using ANTs software (http://stnava.github.io/ANTs) with the standard setting of hyper-parameters. Each voxel value in the raw image was then modified by dividing its value by the corresponding voxel value in the bias field. This approach is widely used for brain studies (26), but not for other organs. It reduces variations of the mean intensity between similar tissues located at different positions within the field of view.

Image Intensity Normalization

Significant variations in mean intensity values measured in similar tissues (for instance subcutaneous fat for lung studies, white matter for brain studies) can be observed between different patients even when using a similar acquisition protocol on the same scanner (26). These variations are a major pitfall for radiomic studies (26). The intensity normalization aims at reducing the intensity variations between different patients. The proposed approach relies on the definition of a reference region that is always in the field of view of thoracic acquisitions, namely the fat. When compared to vertebra and muscle, fat was chosen as the most appropriate reference region because it showed the smallest intra-patient variability (see Results section). Three 2D regions-of-interest (ROIs) were therefore drawn in the normal subcutaneous fat while avoiding vessels, with each region drawn in a different slice. A linear transform was then applied to every image voxel v so that the mean value of the reference tissue was equal to 0 and its standard deviation was equal to 1:

$$\mathrm{Is(v)} = [\mathrm{I(v)} - \mathrm{F}]/\sigma,$$

where I(v) is the original intensity of each voxel in the bias field corrected image, F and σ are the mean intensity and associated standard deviation over all voxels belonging to the three fat regions, and Is(v) is the intensity in voxel v of the normalized image.

Segmentation

An expert radiologist with 4 years experience in thoracic imaging segmented the tumors using the LIFEx software (www.lifexsoft.org) (33). A coarse region surrounding the tumor was manually defined and then refined using an intensity threshold manually set for each patient to delineate the tumor from the lung tissue. The borders between the tumor, the mediastinum and the chest wall were manually delineated. The tumor volume was defined as a single 3D connected component. To further investigate the impact of the segmentation on the results, the original ROIs were modified by automatically shrinking the contours by two pixels.

Radiomic Feature Extraction

Radiomic features were computed in the tumor region for both the raw, the N4ITK-corrected and the normalized N4ITK-corrected MR images using the LIFEx software compliant with

the Image Biomarker Standardization Initiative guidelines (https://arxiv.org/abs/1612.07003). Features included shape features, first-order features that do not account for the spatial arrangement of voxel values, and second-order (textural) features that reflect how voxel values are spatially arranged. The definition of the matrices needed for textural feature calculations requires gray level quantization. For raw images, N4ITK corrected images, and normalized N4ITK images, fixed bin sizes were used for that gray level quantization step. The bin size was chosen so that 256 bins always encompassed all voxel values observed in the tumors, yielding a bin size of 15 units for raw images, 10 units for N4ITK corrected images, and 0.2 for normalized N4ITK corrected images.

For each patient and each image (without and with preprocessing), two sets of features were extracted. A first set of 48 3D radiomic features was obtained from the 3D tumor volume. All feature names are given in **Supplemental Tables 2**, **3** and precisely defined in the LIFEx online documentation (www.lifexsoft.org). A second set of 46 2D radiomic features was extracted for each slice of the tumor, by performing the calculations in the largest 2D connected component present in the slice (the two 3D shape features were not calculated in 2D analysis). Slices with too small regions (<64 pixels) were removed from the analysis, as calculating second-order features in regions with <64 pixels could be meaningless (34). Using the 2D approach, the median value of each feature among the whole set of slices encompassing the patient tumor was defined as the representative value of the corresponding feature for that tumor. Therefore, for each approach, called 2D and 3D in the following, each patient was associated with one 2D value and one 3D value for each feature. Our 2D approach was compared to the conventional 2D approach that consists in selecting the feature value measured from the slice including the largest tumor area.

Classification Tasks

To test the predictive power of each radiomic feature, we determined the ability of each feature to distinguish between adenocarcinoma (ADK) and other tumors (OTH). This task is further referred to as the ADK task. To check the relevance of our findings, a "sham" task was also used by randomly defining a sham ADK group and a sham OTH group. To do so, each patient was randomly assigned to the sham ADK or sham OTH group, whatever the actual tumor type of the patient, but still using the same prevalence of ADK as in the real data (31 patients in the sham ADK group and 21 patients in the sham OTH group). This task is further referred as the RAND task.

Statistical Analysis

To assess the impact of the correction for magnetic field inhomogeneity (N4ITK correction) on voxel values throughout the image volume, three ROIs were manually drawn in the vertebra body of Th3, Th4, and Th5, three ROIs were drawn in the pectoral muscles, in addition to the three ROIs defined in the fat (**Figure 2**). For each type of tissue, the coefficient of variation defined as the standard deviation divided by the mean over all the voxels belonging to the different ROIs of the same tissue was calculated in the original images and in the images after the N4ITK correction. Paired Wilcoxon signed-rank tests were used to determine whether the N4ITK correction significantly impacted the coefficients of variation.

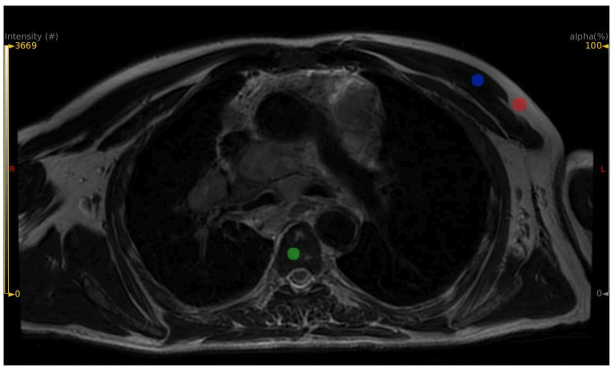


FIGURE 2 | Example of ROI positioning for three candidate reference tissues: subcutaneous fat in red color, vertebral body in green color, pectoral muscle in blue color.

To investigate the ability of radiomic feature to predict whether the tumor was an ADK or an OTH tumor, ROC analysis was performed for each feature, and resulting areas under the curve (AUC) were computed. 2D and 3D feature values were used, as calculated from the raw images, from the images after N4ITK correction and from the fully preprocessed images, i.e., normalized N4ITK corrected images. Following (35), the p-value of the Wilcoxon sum rank test was used to test whether the AUC differed significantly from 0.5. Features for which p-value was <0.05 were thus selected as candidate discriminant features. The same ROC analyses were performed for the RAND task. To reduce the possible false discovery rate, features that remained significant after Benjamini-Hochberg correction for multiple tests were also identified. All these analyses were performed separately using the discovery set and the validation set.

RESULTS

Pathological Data

The pathologic characteristics of tumors are listed in **Table 2**. In the discovery set, there were 31 adenocarcinomas (ADK group) and 21 other histological types (OTH group). The OTH group contained a majority of squamous cell carcinoma (76%). In the validation set, there were 12 adenocarcinomas and 7 other histological types, with a majority of squamous cell carcinoma (57%).

Impact of the Correction of Magnetic Field Inhomogeneity on Voxel Values

Figure 3 shows an example of a bias field as estimated by the N4ITK algorithm for one patient. As expected, the largest variations are observed between the center of the field of view and the periphery near the coil. For the three types of tissue (fat, vertebra, and pectoral muscle), the coefficients of variation demonstrated a statistically significant reduction after bias field correction (see **Supplemental Table 1**). As fat

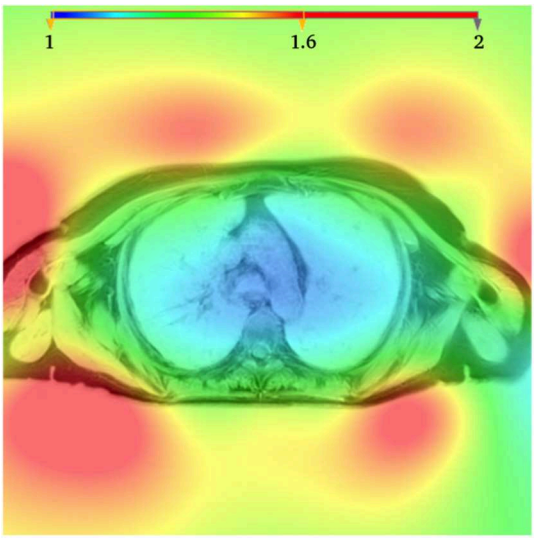


FIGURE 3 | Bias field as estimated using the N4ITK algorithm. The bias field is displayed in color and superimposed to the image in gray scale.

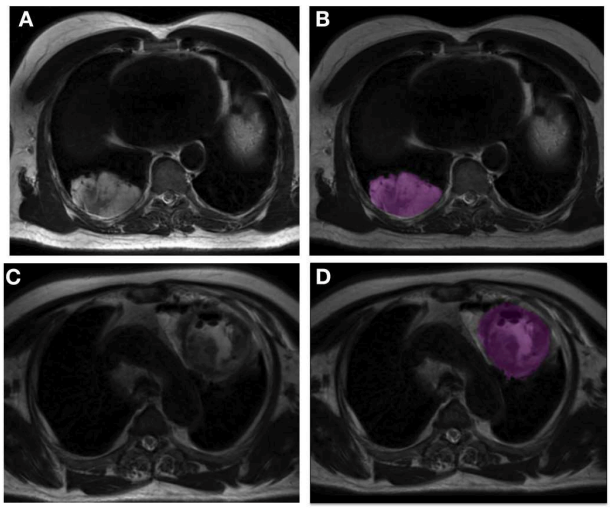


FIGURE 4 | Example of tumor segmentation for two patients. First row: patient with a lung adenocarcinoma of the right lower lobe (long axis: 77 mm). Raw image **(A)** and image after N4ITK correction with the segmented tumor volume in pink **(B)**. Second row: patient with a squamous cell carcinoma of the left upper lobe (long axis: 93 mm). Raw image **(C)** and image after N4ITK correction with the segmented tumor volume in pink **(D)**.

yielded the smallest coefficient of variation, it was chosen as the reference tissue.

Segmentation

Figure 4 shows two examples of tumor segmentation with LIFEx, highlighting the signal heterogeneity within these tumors.

Impact of Pre-processing on the Predictive Values of Radiomic Features

The discovery set was first analyzed. **Table 3** summarizes the significant features (AUC significantly greater that 0.5, $p < 0.05$) when using the 2D and 3D approaches, without preprocessing, with N4ITK correction only and with preprocessing involving the two steps (N4ITK correction and normalization), for both real data and sham data. **Supplemental Tables 2**, **3** provide the AUC and associated 95% confidence intervals for each feature. For the ADK classification task, 8 discriminant features are extracted systematically whatever the configuration tested (except GLRLM_GLNU for 3D N4ITK corrected data, and HISTO_Skewness for 2D N4ITK corrected data, both having a p-value of 0.054). These 8 features (HISTO_Skewness, SHAPE_Volume, GLCM_Correlation, GLRLM_GLNU, GLRLM_RLNU, NGLDM_Coarseness, GLZLM_GLNU, GLZLM_ZLNU) are subsequently called common discriminant features. For the RAND task, no feature yielded an AUC significantly different from 0.5. Correcting for the magnetic field inhomogeneity did not substantially change the number of predictive features. However, when combining bias field correction and normalization, some additional predictive features were observed especially for the 2D configuration with 14 new discriminant features in addition to the 8 common discriminant features. In the 3D analysis, the feature yielding the largest AUC was the "GLCM_Correlation" textural feature with an AUC of 0.77. The same feature yielded the largest AUC in the 2D analysis, with an AUC of 0.82. **Figure 5** shows the associated boxplot corresponding to the 2D analysis for the ADK and OTH groups.

The validation set was analyzed with the same approach. **Table 4** shows the significantly predictive features for all configurations. Compared with the discovery set, less significantly predictive features were identified, partly due to the lower number of patients hence larger confidence intervals (see examples in **Supplemental Table 4**). In all cases, the significant features were part of the 8 common discriminant features identified in the discovery set.

In the best configuration (2D analysis), a Wilcoxon signed-rank test showed that there was no statistically significant difference between the AUC of the eight common discriminant features for the different tested configurations: raw data vs. N4ITK corrected data, raw data vs. normalized N4ITK corrected data, and N4ITK corrected data vs. normalized N4ITK corrected data, and this was true for both the discovery and the validation sets (**Table 5**). There were statistically significant differences between the AUC of the 14 additional features (revealed on the discovery set) for the normalized N4ITK corrected data when compared to raw data or N4ITK corrected data (**Table 5**) and again, this was observed both for the discovery set and for the validation set.

Impact of Segmentation and of 2D Analysis on the Predictive Values of Radiomic Features

Supplemental Table 5 demonstrates the low impact of the tumor border erosion on the 2D discriminant features. Indeed eight

TABLE 3 | Number and list of features with an AUC significantly >0.5 for the different analyses (3D and 2D for raw data, N4ITK corrected data, and N4ITK corrected and normalized data—ADK task based on real data and RAND task based on sham data).

	Raw data	N4ITK corrected data	N4ITK corrected and normalized data	
3D FEATURES (DISCOVERY SET)				
ADK task	8 (**1**)	7 (**1**)	12 (**1**)	
RAND task	0	0	0	
Feature name	HISTO_Skewness SHAPE_Volume **GLCM_Correlation** GLRLM_GLNU GLRLM_RLNU NGLDM_Coarseness GLZLM_GLNU GLZLM_ZLNU	HISTO_Skewness SHAPE_Volume **GLCM_Correlation** GLRLM_RLNU NGLDM_Coarseness GLZLM_GLNU GLZLM_ZLNU	HISTO_Skewness SHAPE_Volume **GLCM_Correlation** GLRLM_LRE GLRLM_GLNU GLRLM_RLNU NGLDM_Coarseness NGLDM_Busyness GLZLM_SZE GLZLM_LZE GLZLM_GLNU GLZLM_ZLNU	
2D FEATURES (DISCOVERY SET)				
ADK task	8 (**5**)	9 (**4**)	22 (**20**)	
RAND task	0	0	0	
Feature name	HISTO_Skewness **SHAPE_Volume** **GLCM_Correlation** GLRLM_GLNU **GLRLM_RLNU** NGLDM_Coarseness **GLZLM_GLNU** **GLZLM_ZLNU**	**SHAPE_Volume** **GLCM_Correlation** GLCM_Entropy_log2 GLCM_Entropy_log10 GLRLM_GLNU **GLRLM_RLNU** NGLDM_Coarseness GLZLM_GLNU **GLZLM_ZLNU**	HISTO_Skewness **SHAPE_Volume** **GLCM_Homogeneity** **GLCM_Contrast** **GLCM_Correlation** **GLCM_Entropy_log2** **GLCM_Entropy_log10** **GLCM_Dissimilarity** **GLRLM_SRE** **GLRLM_LRE** **GLRLM_GLNU**	**GLRLM_RLNU** **GLRLM_RP** **NGLDM_Coarseness** **NGLDM_Contrast** **NGLDM_Busyness** **GLZLM_SZE** **GLZLM_LZE** GLZLM_SZHGE **GLZLM_GLNU** **GLZLM_ZLNU** **GLZLM_ZP**

Bold numbers in brackets give the numbers of significant features after Benjamini-Hochberg correction for multiple comparisons, with corresponding feature names in bold.

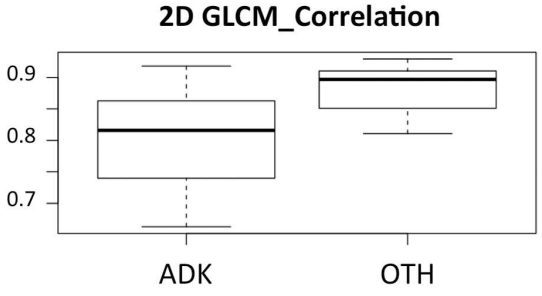


FIGURE 5 | Boxplot showing the values of the 2D "GLCM-correlation" feature for the group of patients with adenocarcinomas (ADK) and the group of patients having a different histological status (OTH).

features (the eight common discriminant features) were revealed using both the initial tumor regions and the eroded regions in the raw data, nine in the N4ITK corrected data (including seven of the eight common discriminant features), and 22 in the N4ITK corrected and normalized data (the eight common discriminant features and the 14 additional features shown in **Table 3**).

Supplemental Table 6 shows the interest of using our 2D approach, selecting the median value of the 2D features computed for all slices encompassing the tumor as opposed to the conventional 2D approach that calculates the feature value from the slice including the largest tumor area. Indeed the number of discriminant features was always superior with our 2D approach except for one supplemental feature for the raw data. For instance we found 6 additional discriminant features for N4ITK corrected and normalized data using our 2D approach instead of the conventional 2D approach.

DISCUSSION

In this study, we investigated the potential of MRI radiomics for lung cancer assessment and demonstrated the need for careful preprocessing of MR images to identify radiomic features correlated with the tumor pathology. While CT and PET scans are the standard imaging procedures to manage lung cancer patients, the clinical workflow can easily include additional lung acquisitions when MRI is prescribed for brain metastasis screening. Here, we focused on an anatomical T2w PROPELLER sequence that produced good quality images in the lung area as assessed by the radiologists in our department. We determined whether radiomic features calculated from these T2-weighted images could predict whether the tumor was an ADK, a question

TABLE 4 | Number and list of features with an AUC significantly >0.5 in the validation set for the 3D and 2D analyses (raw data, N4ITK corrected data, and N4ITK corrected and normalized data—ADK task based on real data).

	Raw data	N4ITK corrected data	N4ITK corrected and normalized data
3D FEATURES (VALIDATION SET)			
ADK task	0	1	1
Feature name		GLCM_Correlation	GLCM_Correlation
2D FEATURES (VALIDATION SET)			
ADK task	4	3	6
Feature name	SHAPE_Volume GLRLM_GLNU NGLDM_Coarseness GLZLM_GLNU	SHAPE_Volume GLRLM_RLNU NGLDM_Coarseness	SHAPE_Volume GLRLM_GLNU GLRLM_RLNU NGLDM_Coarseness GLZLM_GLNU GLZLM_ZLNU

TABLE 5 | Paired Wilcoxon signed rank tests to compare AUC between (1) raw data and N4ITK corrected data, (2) raw data and N4ITK corrected and normalized data, (3) N4ITK corrected data and N4ITK corrected and normalized data for the discriminant features (common and additional) using the discovery and the validation sets.

AUC comparison	2D analysis			
	Common discriminant features ($n = 8$)		Additional discriminant features ($n = 14$)	
	Discovery set	Validation set	Discovery set	Validation set
N4ITK corrected data vs. raw data	ns	ns	ns	ns
N4ITK corrected and normalized data vs. raw data	ns	ns	$p = 0.001^{**}$	$p = 0.001^{**}$
N4ITK corrected and normalized data vs. N4ITK corrected data	ns	ns	$p = 0.001^{**}$	$p = 0.003^{**}$

ns, not significant.
**Stands for p-values smaller than 0.005.

that has already been addressed using CT or PET radiomics (10, 36–40). The best reported performance from non-injected CT scans was an AUC of 0.72 (multivariate analysis) to differentiate ADK and squamous cell carcinomas (36). Using enhanced CT, an AUC of 0.86 was reported at the venous phase for the same classification task (10). Radiomic features extracted from PET/CT could also differentiate adenocarcinoma from other histological types (37–40) with an AUC of 0.81 reported in Kirienko et al. (38), and a radiomic signature to distinguish ADK from squamous cell carcinoma with an AUC of 0.90 reported in Zhu et al. (40).

Interestingly, we found that several MR radiomic features analyzed independently yielded an AUC >0.65 and up to 0.82. Yet, identifying these features required thorough preprocessing, without which up to 66% of the informative features (14 out of the 22 in the 2D approach, see **Supplemental Table 3**) were not identified as such.

The need for some preprocessing steps before extracting MRI radiomic features has been very recently acknowledged for other tumor types (41, 42). Although there is no consensus on the preprocessing methods that should be used, two main pitfalls that are specific to MRI have been identified. The first one results from the B1 magnetic field inhomogeneities (43, 44) caused by MR gradients that introduces variability in signal intensity of a given tissue type as a function of its location within the field of view. This bias is more severe in high field MR and was present in our 3T data. The second challenge is the significant variation in pixel values between different patients (42, 45, 46) in anatomical MR images, even when using the same scanner and the same acquisition sequences, due to the arbitrary units used to represent the anatomical MR images.

In our work, we proposed two complementary approaches to deal with these two issues. A bias field correction was performed using the N4ITK method, which is the state-of-art method for brain studies. N4ITK is a histogram based technique that estimates a slowly varying bias field by maximizing the high frequency histogram content in the image (32). Our goal was to validate its use in thoracic imaging and assess its impact on subsequent radiomic analysis. Using normal tissues, such as the vertebral bodies, subcutaneous fat, and pectoral muscle, we demonstrated that this correction was successful at reducing the variations of voxel values in all these tissues (**Supplemental Table 1**), the largest effect being observed in the subcutaneous fat. This is very likely due to the fat peripheral location. The identification of radiomic features able to predict ADK tumors was only slightly impacted by this correction (**Table 3**). Indeed this correction did not aim at increasing the identification of informative radiomic features, but at improving the subsequent normalization procedure, by reducing the coefficients of variation in the reference tissue. Other techniques of bias field correction, such as B1 mapping could also be of interest. Yet, a definite advantage of N4ITK is that it can be retrospectively used, which is especially useful as many radiomic studies are still performed retrospectively.

The challenge of image intensity normalization was addressed by defining a reference tissue. We chose the subcutaneous fat as it was always present in the thoracic field of view and showed the lowest coefficient of variation within patients (**Supplemental Table 1**). The principle of the normalization was to arbitrarily set the MR intensity to 0 in fat regions and its associated standard deviation to 1, similar to setting Hounsfield Units to 0 in water in CT imaging. Doing so, for any patient, the value will be 0 in the fat for these anatomical T2-weighted PROPELLER images, and all image values will be scaled linearly. The linear transformation is a very simple model with respect to the complexity of the MR signal intensity and more sophisticated models could certainly be used, but our aim was to determine whether this simple transformation could already reduce the variability of MR signal intensity across patients hence increase the statistical power of MR radiomic analysis. Our results suggest that the number of informative features for identifying ADK is substantially increased when using the image intensity normalization combined with the

bias field correction (**Table 3**). In the 3D approach, four additional features were identified while in the 2D analysis, 14 additional features were identified as discriminant. The eight common discriminant features that were already identified as informative before preprocessing remained informative after preprocessing, demonstrating that these features were robust with respect to the magnetic field heterogeneity and intensity scaling. Indeed five (SHAPE_Volume, GLRLM_GLNU, GLRLM_RLNU, GLZLM_GLNU, GLZLM_ZLNU) of these eight features are highly correlated with the volume of the region of interest (34), and that volume remains identical whatever the preprocessing steps. The question of whether the 14 new radiomic features identified as informative after preprocessing were truly informative for the classification task or were "false positive" features was answered by designing the sham experiment and by analyzing the validation set. In the sham experiment, we knew that we should not find any feature that would be related to the "fake" ADK or OTH status of the tumors, as each tumor was randomly assigned as ADK or OTH, whatever its actual pathological report. **Table 3** confirmed that without preprocessing, with N4ITK correction, and with full preprocessing, no feature was identified as informative of the fake tumor type. **Table 5** shows that the trends observed on the discovery set for the 14 additional features identified using the normalized N4ITK corrected data were confirmed on the validation set. All additional features were textural features, demonstrating the need for preprocessing to compute robust discriminant textural features. Altogether, these results demonstrate that the preprocessing does not produce an inflation of false positive and suggests that the additional features identified in the real classification task are truly informative.

In our cohort, all images had the same voxel size. It was thus not necessary to resample the images as previously proposed (47, 48) to reduce the variability induced by different voxel size. To characterize the tumor type based on the MR radiomic features, we compared two approaches: a 2D approach where radiomic features were computed in each slice and the median value over all slices was chosen as the representative value for the tumor, and a 3D approach in which the features directly pertain to the whole tumor volume. The 2D approach identified more informative radiomic features than then 3D approach (**Table 3**). Several hypotheses might explain this result. First, voxels are not isotropic, because the slice thickness (4 mm) is greater than the intra-plane voxel size (0.8 mm). As a result, 3D calculation of second-order feature is biased. Another reason might be the large size of most tumors in our study. All patients had advanced tumors with a mean diameter of 63 ± 23 mm, so each slice already contained a representative view of the tumor that might be sufficient to estimate the tumor type (see **Figure 4**). Two previous studies compared 2D and 3D radiomic feature performance for lung cancer in CT (49, 50). The first one did not find any significant difference between 2D and 3D results (49), while the second study reported better performance using the 3D analysis (50). Yet, for these two studies, the 2D analysis was limited to the slice that included the largest cross-section of the lesion, while in our so-called 2D approach, we still accounted for all slices encompassing the tumor. The selection of one single slice might lead to information loss while our 2D approach used all 2D slices to end up with a single feature value per tumor volume. Our 2D approach identified more discriminant features than the one-slice based 2D approach (**Supplemental Table 6**), especially for N4ITK corrected and normalized data.

The feature that yielded the largest AUC was GLCM_Correlation. This feature has actually already been reported as predictive in other MRI radiomic study: lower values of GLCM_Correlation on Diffusion Weighted Images and higher values of GLCM_Contrast on T2w sequences were shown to be correlated to an early disease progression in rectal cancers (51).

There are several limitations in our study. First, our results related to the prediction of ADK should be confirmed on a larger cohort. As all data were acquired in the same institution and using the same scanner, our findings would also need a multi-center validation. Another limitation is due to the fact that only one operator segmented the tumors and the robustness of the findings with respect to the tumor delineation should be further investigated (52). To investigate the impact of the tumor delineation on our results, all the segmented tumors were automatically eroded by an element of size 1.5 mm and results were similar, confirming the greater sensitivity of 2D analysis on normalized N4ITK corrected data to identify discriminant features (see **Supplemental Table 5**). This suggests that for tumors with large volumes as in our study, significant variations in results due to small changes in tumor contour delineation are unlikely. Histology was mostly determined by trans-thoracic core needle biopsy, which might not be representative of the whole tumor volume. This is a definite limitation as lung tumors may have heterogeneous histological types depending on the location in the lesion (53). Our task was to distinguish between ADK and all other tumor types, so this second tumor group was quite heterogeneous in itself. The reason why we did not separate the OTH group into different tumor types was to keep enough tumors in each group for the classification task. Last, the best prediction accuracy we obtained (AUC of 0.82) is not sufficient for clinical applicability (36). This accuracy might be limited by the fact that we used univariate models only, because of the relatively small size of our cohort. Our results warrant multivariate analyses based on larger patient cohorts. Also, we focused on one MR sequence only, while combining radiomic features from different MR sequences might be useful to enhance the accuracy of the classification.

CONCLUSION

We demonstrated that MRI T2-weighted sequences of lung cancer patients yielded radiomic features related to the pathological tumor type and that the number of informative radiomic features was significantly increased by appropriate processing of the MR images. Key preprocessing steps are correction for the magnetic field inhomogeneity and normalization of the voxel values to set a intensity scale common to all patient images. In addition, in our cohort, the 2D analysis selecting the median value of each feature among the different slices encompassing the tumor volume revealed more discriminant radiomic features than the 3D analysis. Based on these results, further exploration of the potential of MR radiomics in lung cancer patients is warranted.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available because they were extracted from an on-going clinical study (IRB protocol 32-2016, study number: 2016-A00813-48). Requests to access the datasets should be directed to Pierre-Yves Brillet, pierre-yves.brillet@aphp.fr.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comité de Protection des Personnes d'Ile de France Paris X. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ML, FF, P-YB, and IB contributed to the conception and design of the study. ML and P-YB collected the image and clinical data. J-FB did the pathological records reviewing. ML and FF performed the image analysis. CN and FO developed some specific software components. A-SD and IB performed the statistical analysis. ML wrote the first draft of the manuscript. FF and IB wrote some sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

The authors thank Dr. M. Kambouchner and Prof. A. Martin pathologists, Drs. B. Duchemann and K. Chouahnia clinical oncologists. ML thanks the Agence Régionale de Santé Ile de France for funding his research grant. All authors thank GE Healthcare for funding a technician to collect clinical data and Institut National de la Santé et de la Recherche Médicale for funding open access publication fees.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00043/full#supplementary-material

REFERENCES

1. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Mag Reson Imag.* (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
2. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJWL. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *Front Oncol.* (2015) 5:272. doi: 10.3389/fonc.2015.00272
3. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin.* (2019) 69:127–57 doi: 10.3322/caac.21552
4. Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, et al. Defining the biological basis of radiomic phenotypes in lung cancer. *eLife.* (2017) 6:e23421. doi: 10.7554/eLife.23421
5. Chaddad A, Kucharczyk MJ, Daniel P, Sabri S, Jean-Claude BJ, Niazi T, et al. Radiomics in glioblastoma: current status and challenges facing clinical implementation. *Front Oncol.* (2019) 9:374. doi: 10.3389/fonc.2019.00374
6. Jethanandani A, Lin TA, Volpe S, Elhalawani H, Mohamed ASR, Yang P, et al. Exploring applications of radiomics in magnetic resonance imaging of head and neck cancer: a systematic review. *Front Oncol.* (2018) 8:131. doi: 10.3389/fonc.2018.00131
7. Yi X, Pei Q, Zhang Y, Zhu H, Wang Z, Chen C, et al. MRI-based radiomics predicts tumor response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Front Oncol.* (2019) 9:552. doi: 10.3389/fonc.2019.00552
8. Thawani R, McLane M, Beig N, Ghose S, Prasanna P, Velcheti V, et al. Radiomics and radiogenomics in lung cancer: a review for the clinician. *Lung Cancer.* (2018) 115:34–41. doi: 10.1016/j.lungcan.2017.10.015
9. Chaddad A, Daniel P, Niazi T. Radiomics evaluation of histological heterogeneity using multiscale textures derived from 3D wavelet transformation of multispectral images. *Front Oncol.* (2018) 8:96. doi: 10.3389/fonc.2018.00096
10. E L, Lu L, Li L, Yang H, Schwartz LH, Zhao B. Radiomics for classifying histological subtypes of lung cancer based on multiphasic contrast-enhanced computed tomography. *J Comput Assist Tomogr.* (2019) 43:300–6. doi: 10.1097/RCT.0000000000000836
11. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. *Clin Lung Cancer.* (2016) 17:441–8.e6. doi: 10.1016/j.cllc.2016.02.001
12. Shen T-X, Liu L, Li W-H, Fu P, Xu K, Jiang Y-Q, et al. CT imaging-based histogram features for prediction of EGFR mutation status of bone metastases in patients with primary lung adenocarcinoma. *Cancer Imag.* (2019) 19:34. doi: 10.1186/s40644-019-0221-9
13. Yip SSF, Parmar C, Kim J, Huynh E, Mak RH, Aerts HJWL. Impact of experimental design on PET radiomics in predicting somatic mutation status. *Eur J Radiol.* (2017) 97:8–15. doi: 10.1016/j.ejrad.2017.10.009
14. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic mutations drive distinct imaging phenotypes in lung cancer. *Cancer Res.* (2017) 77:3922–30. doi: 10.1158/0008-5472.CAN-17-0122
15. Yoon SH, Park CM, Park SJ, Yoon J-H, Hahn S, Goo JM. Tumor heterogeneity in lung cancer: assessment with dynamic contrast-enhanced MR imaging. *Radiology.* (2016) 280:940–8. doi: 10.1148/radiol.2016151367
16. Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, et al. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLoS Med.* (2018) 15:e1002711. doi: 10.1371/journal.pmed.1002711
17. Park H, Sholl LM, Hatabu H, Awad MM, Nishino M. Imaging of precision therapy for lung cancer: current state of the art. *Radiology.* (2019) 293:15–29. doi: 10.1148/radiol.2019190173
18. Hochhegger B, Zanon M, Altmayer S, Pacini GS, Balbinot F, Francisco MZ, et al. Advances in imaging and automated quantification of malignant pulmonary diseases: a state-of-the-art review. *Lung.* (2018) 196:633–42. doi: 10.1007/s00408-018-0156-0
19. Yi CA, Shin KM, Lee KS, Kim B-T, Kim H, Kwon OJ, et al. Non-small cell lung cancer staging: efficacy comparison of integrated PET/CT versus 3.0-T whole-body MR imaging. *Radiology.* (2008) 248:632–42. doi: 10.1148/radiol.2482071822
20. Kim HS, Goh MJ, Kim N, Choi CG, Kim SJ, Kim JH. Which combination of MR imaging modalities is best for predicting recurrent glioblastoma? Study of diagnostic accuracy and reproducibility. *Radiology.* (2014) 273:831–43. doi: 10.1148/radiol.14132868
21. Hoeks CMA, Hambrock T, Yakar D, Hulsbergen-van de Kaa CA, Feuth T, Witjes JA, et al. Transition zone prostate cancer: detection and localization with 3-T multiparametric MR imaging. *Radiology.* (2013) 266:207–17. doi: 10.1148/radiol.12120281
22. Mahon RN, Hugo GD, Weiss E. Repeatability of texture features derived from magnetic resonance and computed tomography imaging and use in

predictive models for non-small cell lung cancer outcome. *Phys Med Biol.* (2019) 64:145007. doi: 10.1088/1361-6560/ab18d3
23. Shen G, Jia Z, Deng H. Apparent diffusion coefficient values of diffusion-weighted imaging for distinguishing focal pulmonary lesions and characterizing the subtype of lung cancer: a meta-analysis. *Eur Radiol.* (2016) 26:556–66. doi: 10.1007/s00330-015-3840-y
24. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring computed tomography scanner variability of radiomics features. *Invest Radiol.* (2015) 50:757–65. doi: 10.1097/RLI.0000000000000180
25. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting CT radiomics. *Radiology.* (2019) 291:53–9. doi: 10.1148/radiol.2019182023
26. Goya-Outi J, Orlhac F, Calmon R, Alentorn A, Nioche C, Philippe C, et al. Computation of reliable textural indices from multimodal brain MRI: suggestions based on a study of patients with diffuse intrinsic pontine glioma. *Phy Med Biol.* (2018) 63:105003. doi: 10.1088/1361-6560/aabd21
27. Shinohara RT, Sweeney EM, Goldsmith J, Shiee N, Mateen FJ, Calabresi PA, et al. Statistical normalization techniques for magnetic resonance imaging. *Neuroimage.* (2014) 6:9–19. doi: 10.1016/j.nicl.2014.08.008
28. Schwier M, van Griethuysen J, Vangel MG, Pieper S, Peled S, Tempany C, et al. Repeatability of multiparametric prostate MRI radiomics features. *Sci Rep.* (2019) 9:9441. doi: 10.1038/s41598-019-45766-z
29. Pandit P, Qi Y, King KF, Johnson GA. Reduction of artifacts in T2-weighted PROPELLER in high-field preclinical imaging. *Magn Reson Med.* (2011) 65:538–43. doi: 10.1002/mrm.22624
30. Meier-Schroers M, Kukuk G, Homsi R, Skowasch D, Schild HH, Thomas D. MRI of the lung using the PROPELLER technique: Artifact reduction, better image quality and improved nodule detection. *Eur J Radiol.* (2016) 85:707–13. doi: 10.1016/j.ejrad.2015.12.016
31. Scholz O, Denecke T, Böttcher J, Schwarz C, Mentzel H-J, Streitparth F, et al. MRI of cystic fibrosis lung manifestations: sequence evaluation and clinical outcome analysis. *Clin Radiol.* (2017) 72:754–63. doi: 10.1016/j.crad.2017.03.017
32. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imag.* (2010) 29:1310–20. doi: 10.1109/TMI.2010.2046908
33. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* (2018) 78:4786–9. doi: 10.1158/0008-5472.CAN-18-0125
34. Orlhac F, Soussan M, Maisonobe J-A, Garcia CA, Vanderlinden B, Buvat I. Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med.* (2014) 55:414–22. doi: 10.2967/jnumed.113.129858
35. Mason SJ, Graham NE. Areas beneath the relative operating characteristics (roc) and relative operating levels (rol) curves: statistical significance and interpretation. *Quart J R Meteorol Soc.* (2002) 128:2145–66. doi: 10.1256/003590002320603584
36. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, et al. Exploratory study to identify radiomics classifiers for lung cancer histology. *Front Oncol.* (2016) 6:71. doi: 10.3389/fonc.2016.00071
37. Kim D-H, Jung J, Son SH, Kim C-Y, Hong CM, Oh J-R, et al. Prognostic significance of intratumoral metabolic heterogeneity on 18F-FDG PET/CT in pathological N0 non–small cell lung cancer: *Clin Nucl Med.* (2015) 40:708–14. doi: 10.1097/RLU.0000000000000867
38. Kirienko M, Cozzi L, Rossi A, Voulaz E, Antunovic L, Fogliata A, et al. Ability of FDG PET and CT radiomics features to differentiate between primary and metastatic lung lesions. *Eur J Nucl Med Mol Imag.* (2018) 45:1649–60. doi: 10.1007/s00259-018-3987-2
39. Ha S, Choi H, Cheon GJ, Kang KW, Chung J-K, Kim EE, et al. Autoclustering of non-small cell lung carcinoma subtypes on 18F-FDG PET using texture analysis: a preliminary result. *Nucl Med Mol Imag.* (2014) 48:278–86. doi: 10.1007/s13139-014-0283-3
40. Zhu X, Dong D, Chen Z, Fang M, Zhang L, Song J, et al. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *Eur Radiol.* (2018) 28:2772–8. doi: 10.1007/s00330-017-5221-1
41. Um H, Tixier F, Bermudez D, Deasy JO, Young RJ, Veeraraghavan H. Impact of image preprocessing on the scanner dependence of multi-parametric MRI radiomic features and covariate shift in multi-institutional glioblastoma datasets. *Phy Med Biol.* (2019) 64:165011. doi: 10.1088/1361-6560/ab2f44
42. Chirra P, Leo P, Yim M, Bloch BN, Rastinehad AR, Purysko A, et al. Multisite evaluation of radiomic feature reproducibility and discriminability for identifying peripheral zone prostate tumors on MRI. *J Med Imag.* (2019) 6:1. doi: 10.1117/1.JMI.6.2.024502
43. Antunes J, Viswanath S, Rusu M, Valls L, Hoimes C, Avril N, et al. Radiomics analysis on FLT-PET/MRI for characterization of early treatment response in renal cell carcinoma: a proof-of-concept study. *Transl Oncol.* (2016) 9:155–62. doi: 10.1016/j.tranon.2016.01.008
44. Farjam R, Tyagi N, Veeraraghavan H, Apte A, Zakian K, Hunt MA, et al. Multiatlas approach with local registration goodness weighting for MRI-based electron density mapping of head and neck anatomy. *Med Phys.* (2017) 44:3706–17. doi: 10.1002/mp.12303
45. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS ONE.* (2014) 9:e102107. doi: 10.1371/journal.pone.0102107
46. Lakhman Y, Veeraraghavan H, Chaim J, Feier D, Goldman DA, Moskowitz CS, et al. Differentiation of uterine leiomyosarcoma from atypical leiomyoma: diagnostic accuracy of qualitative MR imaging features and feasibility of texture analysis. *Eur Radiol.* (2017) 27:2903–15. doi: 10.1007/s00330-016-4623-9
47. Li Q, Bai H, Chen Y, Sun Q, Liu L, Zhou S, et al. A fully-automatic multiparametric radiomics model: towards reproducible and prognostic imaging signature for prediction of overall survival in glioblastoma multiforme. *Sci Rep.* (2017) 7:14331. doi: 10.1038/s41598-017-14753-7
48. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys.* (2017) 44:1050–62. doi: 10.1002/mp.12123
49. Shen C, Liu Z, Guan M, Song J, Lian Y, Wang S, et al. 2D and 3D CT radiomics features prognostic performance comparison in non-small cell lung cancer. *Transl Oncol.* (2017) 10:886–94. doi: 10.1016/j.tranon.2017.08.007
50. Yang L, Yang J, Zhou X, Huang L, Zhao W, Wang T, et al. Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients. *Eur Radiol.* (2019) 29:2196–206. doi: 10.1007/s00330-018-5770-y
51. Nardone V, Reginelli A, Scala F, Carbone SF, Mazzei MA, Sebaste L, et al. Magnetic-resonance-imaging texture analysis predicts early progression in rectal cancer patients undergoing neoadjuvant chemoradiation. *Gastroenterol Res Pract.* (2019) 2019:8505798. doi: 10.1155/2019/8505798
52. Tixier F, Um H, Young RJ, Veeraraghavan H. Reliability of tumor segmentation in glioblastoma: impact on the robustness of MRI-radiomic features. *Med Phys.* (2019) 46:3582–91. doi: 10.1002/mp.13624
53. Travis WD. Pathology of lung cancer. *Clin Chest Med.* (2011) 32:669–92. doi: 10.1016/j.ccm.2011.08.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 31 January 2020
doi: 10.3389/fonc.2020.00028

Radiomics Signature as a Predictive Factor for EGFR Mutations in Advanced Lung Adenocarcinoma

Duo Hong[1], *Ke Xu*[1]*, *Lina Zhang*[1]*, *Xiaoting Wan*[1] and *Yan Guo*[2]

[1] Department of Radiology, The First Hospital of China Medical University, Shenyang, China, [2] GE Healthcare, Shanghai, China

Purpose: To develop and validate a radiomic signature to identify EGFR mutations in patients with advanced lung adenocarcinoma.

Methods: This study involved 201 patients with advanced lung adenocarcinoma (140 in the training cohort and 61 in the validation cohort). A total of 396 features were extracted from manual segmentation based on enhanced and non-enhance CT imaging after image preprocessing. The Lasso algorithm was used for feature selection, 6 machine learning methods were used to construct radiomics models. Receiver operating characteristic (ROC) curve analysis was applied to evaluate the performance of the radiomic signature between different data and methods. A nomogram was developed using clinical factors and the radiomics signature, then it was analyzed based on its discriminatory ability and calibration. Decision curve analysis (DCA) was implemented to evaluate the clinical utility.

Results: Ten features for contrast data and eleven features for non-contrast data were selected through LASSO algorithm. The performance of the radiomics signature for contrast images was better than that for non-contrast images in all of the 6 different machine learning methods. Finally, the best radiomics signature was built with logistic regression method based on enhanced CT imaging with an area under the curve (AUC) of 0.851 (95% CI, 0.750 to 0.951) in the validation cohort. A nomogram was developed using the radiomics signature and sex with a C-index of 0.908 (95% CI, 0.862 to 0.954) in the training cohort and 0.835 (95% CI, 0.825 to 0.845) in the validation cohort. It showed good discrimination and calibration (Hosmer-Lemeshow test, $P = 0.621$ for the training cohort and $P = 0.605$ for the validation cohort).

Conclusion: Radiomics signature can help to distinguish between EGFR positive and wild type advanced lung adenocarcinomas.

Keywords: advanced lung adenocarcinoma, tomography, epidermal growth factor receptor, mutation, radiomics

OPEN ACCESS

Edited by:
Rong Tian,
Sichuan University, China

Reviewed by:
Jung Hun Oh,
Memorial Sloan Kettering Cancer Center, United States
Seyedmehdi Payabvash,
School of Medicine, Yale University, United States

***Correspondence:**
Ke Xu
kxu@cmu.edu.cn
Lina Zhang
zhanglnda@163.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *21 August 2019*
Accepted: *09 January 2020*
Published: *31 January 2020*

Citation:
Hong D, Xu K, Zhang L, Wan X and Guo Y (2020) Radiomics Signature as a Predictive Factor for EGFR Mutations in Advanced Lung Adenocarcinoma. Front. Oncol. 10:28. doi: 10.3389/fonc.2020.00028

INTRODUCTION

Lung cancer is one of the most common malignant tumors in the world and the leading cause of cancer-related death worldwide (1). The World Health Organization (WHO) divides lung cancer into two major categories: non-small cell lung cancer (NSCLC), representing more than 85% of all cases, and small cell lung cancer (SCLC). Adenocarcinoma in NSCLC is the major histological

subtype, accounting for almost half of all lung cancer cases (2). The 5-year survival rate is >50% when the disease is still localized; however, 75% of cases are diagnosed at an advanced stage with unresectable lesions (3).

Over the last decade, advances in molecularly targeted drugs for thoracic oncology have led to a new emphasis on accurate analyses of biomolecular markers in a subset of lung adenocarcinoma (4). Patients with advanced lung adenocarcinoma harboring epidermal growth factor receptor (EGFR)-activating mutations showed a significant progression-free survival (PFS) benefit with reduced side effects by treatment with tyrosine kinase inhibitor (TKIs) (5). TKI therapy had already been used as first-line systemic therapy before chemotherapy (6, 7). Biopsy is the only widely used means to identify mutations of EGFR in unresectable lesions, but some patients refuse the procedure due to the risk of hemorrhage and pneumothorax. Furthermore, it is difficult to obtain tissue samples from inaccessible locations in some cases. Therefore, automatic, non-invasive, and cost-effective alternatives are desired (8). Radiomics refers to the systematic extraction and analysis of features from digital medical images with the intent of creating mineable databases to aid in diagnosis and treatment. Radiogenomics even involves specific features connecting genomic phenotypes and radiological images. The aim of this study was to develop a radiogenomic approach to identify EGFR mutations in advanced lung adenocarcinoma non-invasively.

MATERIALS AND METHODS

Patients

Institutional review board approval was obtained for this retrospective study, and with a waiver for the informed consent requirement. Consecutive patients (n = 449) with advanced lung adenocarcinoma who were admitted to the hospital from January 2014 to January 2016 were enrolled in this retrospective study. All cases were histologically confirmed by transthoracic biopsy and classified as stage IIIB-IV according to the Eighth Edition of the Lung Cancer Stage Classification (9). EGFR mutations in exons 18, 19, 20, and 21 were detected using human EGFR gene mutations detection kit (AmoyDx, China) via Amplification Refractory Mutation System (ARMS) real-time Polymerase Chain Reaction (PCR) technology. A total of 248 patients were excluded based on the following exclusion criteria: [1] examined by an unassigned CT scanner (n = 105); [2] received previous anticancer therapy or with other types of cancer (n = 19); [3] no EGFR mutation analysis available (n = 81); [4] difficulty in drawing regions of interest (ROIs) (n = 43).

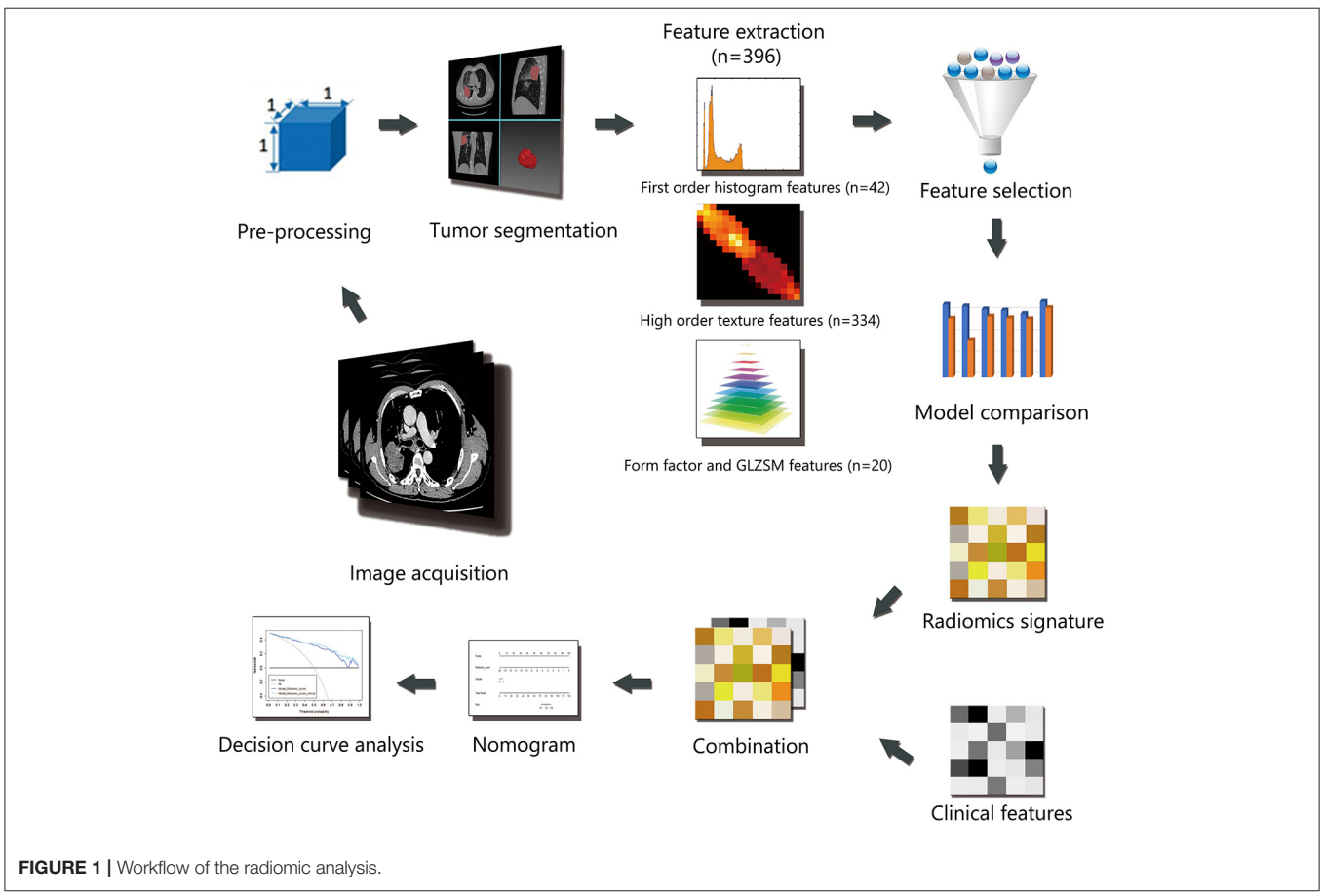


FIGURE 1 | Workflow of the radiomic analysis.

Finally, 201 patients were included in the study. The clinical data collected for analysis included sex, age, smoking status, and stage. The patients were randomly divided into two individual cohorts for training and validation at a ratio of 7:3. The workflow of the radiomic analysis is illustrated in **Figure 1**.

Image Acquisition

Contrast-enhanced computed tomography (CT) images were acquired at our hospital using either a Toshiba Aquilion One (Toshiba Medical Systems) or Phillips Brilliance iCT (Philips Medical Systems) system. The scanning parameters were as follows: 120 kVp; 100-200 mAs; detector collimation of 64× or 128× 0.625 mm; field of view of 350 × 350 mm; and matrix of 512 × 512. After routine CT, a dose of 85 mL non-ionic iodinated contrast material (350 mg iodine/mL, Omnipaque, GE Healthcare) was injected into the antecubital vein at a rate of 3.0 mL/s using an automated injector (Ulrich CT Plus 150, Ulrich Medical). CT scanning was performed again with a 25-second delay after the injection. All images were reconstructed at a slice thickness of 2 mm. Contrast and non-contrast images were retrieved separately from the Picture Archiving and Communication System (PACS) workstation (IMPAX, AGFA).

Image Preprocessing

Due to the use of different CT scans, image preprocessing (**Figure 2**) before segmentation and feature extraction was performed to improve the robustness of the radiomic features. The process included two steps: Step 1. To eliminate the intrinsic dependency on voxel size for the radiomic features, a resampling method with a linear interpolation algorithm was used to normalize the voxel size. Meanwhile, higher-order texture analysis features, such as GLCM and GLRLM features, were derived from different directions (also called "angles") and different scales (denoted here as "offsets"); thus, the anisotropic voxels scanned at 0.743 mm*0.743 mm*2.000 mm or other size were resampled to form isotropic voxels, i.e., 1.000 mm*1.000 mm*1.000 mm. Step 2. A Gaussian filter was used to remove "unwanted signals", i.e., noise beyond the scope of the ($\mu \pm 3\sigma$) CT values. The gray level was consistent across the different scanners, so gray level normalization was not used here.

Tumor Segmentation

ROIs were manually contoured along the boundaries of the tumor layer by layer in reference to images in both the mediastinum and lung windows. Segmentation was strictly performed by a chest radiologist (W.XT.) with 7-year experience in lung CT using ITK-Snap (version 3.4.0, www.itk-snap.org) software and confirmed by another chest radiologist (H.D.) with 13-year experience. Both radiologists were blinded to the diagnosis and EGFR mutation status.

Feature Extraction

Four types of radiomic features were extracted from both contrast and non-contrast CT images, and the details are shown in **Figure S1**. Features based on the three-dimensional volume of interest (3D VOI) were generated automatically using in-house software (Artificial Intelligence Kit, A.K., GE Healthcare).

Feature Selection

Some features might contribute to the positive performance of classification while others might add noise to it (10). The least absolute shrinkage and selection operator (LASSO) algorithm, which is suitable for high-dimensional low-sample size data with the problem of collinearity (11, 12), was used to select effective and predictable features in the training cohort after data split. Features with nonzero coefficients were chosen based on 10-fold cross-validation.

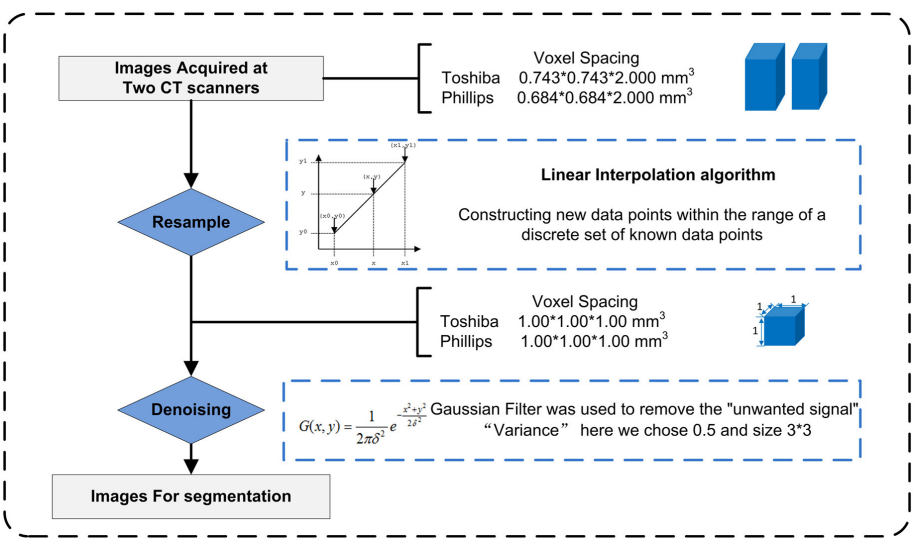


FIGURE 2 | Image preprocessing.

Model Construction

After feature selection, 6 machine learning methods were used to construct models which include NBC (Naive Bayesian Classifier), KNN (K-Nearest Neighbor), RF (Radom Forest), SVM (Support Vector Machine), DT (Decision Tree), LR (Logistic Regression). Their predictive performance was measured by using area under the curve (AUC) of receiver operating characteristic (ROC) curve analysis in the validation cohort. First, AUC of each model in contrast and non-contrast data were compared, and inferior data was abandoned, then in superior data, the optimal model was chosen for further analysis.

Nomogram Construction

The nomogram was constructed based on multivariable logistic regression analysis. Clinical factors and radiomics signature were included in a nomogram model for predicting EGFR mutations in the training cohort. The discriminative power of the model was evaluated by Harrell's concordance index (C-index) with 95% confidence intervals in both cohorts. The calibration curve was plotted to explore the predictive accuracy of the model. Decision curve analysis (DCA) was implemented to evaluate the clinical usefulness by quantifying the net benefits of the nomogram model in both the training and validation cohorts.

Statistical Analysis

All statistical tests were performed using R statistical software version 3.5.2. The "glmnet" package was used for executing the LASSO algorithm. For the baseline characteristic analyses, quantitative data were compared using Student's *t*-test, and categorical data were compared using the $\chi 2$ test. All statistical tests were two-tailed, and $p < 0.05$ indicated a significant difference.

RESULTS

The baseline clinical characteristics of the training and validation cohorts are listed in **Table 1**. There was no significant difference between training and validation cohorts in overall distribution of age, sex, smoking status or stage.

A total of 396 features were extracted. In the training cohort, 10 features for contrast images and 11 features for non-contrast images were evaluated to construct models through LASSO algorithm (**Figure S2**, **Table S1**).

The predictive performance of all six models based on contrast and non-contrast data were described in **Figure 3**. The predictive performance of all six models based on contrast

TABLE 1 | Demographic data of patients in the training and validation cohorts.

Variable	Training cohort			Validation cohort			*p*
	Mutant	Wild type	*p*	Mutant	Wild type	*p*	
Age (y, mean ± SD)	58.24 ± 11.05	57.93 ± 8.43	0.85	59.23 ± 7.62	57.07 ± 8.38	0.276	0.929
Sex, *n* (%)			0.007*			0.149	0.437
Male	28(40.0)	44(62.9)		15(46.6)	20(66.7)		
Female	42(60.0)	26(37.1)		16(53.4)	10(33.3)		
Smoking Status, *n* (%)			0.003*			0.09	0.396
Smoker	13(18.6)	29(41.4)		8(25.8)	14(46.7)		
Never smoker	57(81.4)	41(58.6)		23(74.2)	16(53.3)		
Stage, *n* (%)			0.002*			0.119	0.103
III	4 (5.7)	17 (24.3)		5 (16.1)	10 (33.3)		
IV	66 (94.3)	53 (75.7)		26 (83.9)	20 (66.7)		
Radiomic score, median (interquartile range)	1.42 (0.57 to 2.46)	−1.63 (−2.88 to 0.42)	<0.001*	1.01 (−0.53 to 2.21)	−1.93 (−4.47 to −1.22)	<0.001*	0.145

**P-value < 0.05.*

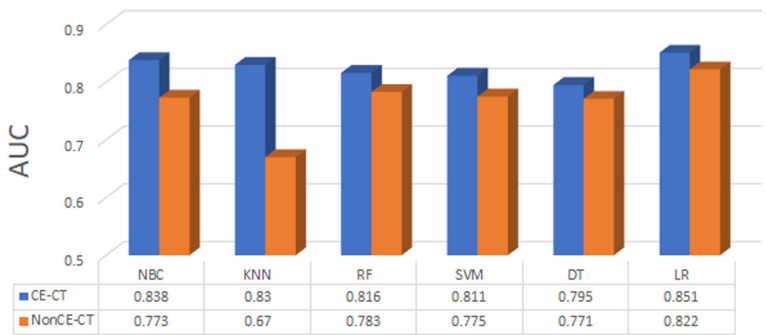


FIGURE 3 | The predictive performance of all machine learning methods based on contrast (CE-CT) and non-contrast (nonCE-CT) data.

and non-contrast data were described in **Figure 3**. Although there was no significant difference by Delong test in all results, the value of AUC in contrast images was better than non-contrast images in all models, hence, the non-contrast data was excluded from further analysis. The machine learning method of LR which could assign each patient a radiomic score (rad-score) obtained a better value than other models, therefore, the nomogram was built based on the LR model in contrast data.

Table 2 shows the results of multivariable logistic regression analysis including sex, age, smoking status, and rad-score. Sex and rad-score appeared to be independent prognostic predictors of mutations in this model. The model that incorporated the above independent predictors is presented as the nomogram (**Figure 4**). The model showed a favorable C-index of 0.908 (95% CI, 0.862 to 0.954) in the training cohort and 0.835 (95% CI, 0.825 to 0.845) in the validation cohort.

The calibration curve of the radiomic nomogram for the probability of EGFR mutations demonstrated good agreement between the predicted and observed results in both cohorts (**Figure 5**). The Hosmer-Lemeshow test showed no significant statistical difference between calibration curves and ideal curves ($P = 0.621$ for the training cohort and $P = 0.605$ for the validation cohort).

DCA was performed for the radiomic model (light blue line) and nomogram model (dark blue line) as shown in **Figure 6**. Using the radiomic model and the nomogram model to predict the EGFR status added more benefit than using the treat-all scheme or the treat-none scheme at any given threshold probability in the training cohort. For threshold probabilities > 20%, using the radiomic model and the nomogram model to predict the EGFR status added more benefit than using the treat-all scheme or the treat-none scheme in the validation cohort.

TABLE 2 | Multivariable logistic regression for nomogram construction.

	Coefficient	Odds ratio	95% CI		*p*
			Lower	Upper	
Intercept	−0.734				0.049*
Radiomic score	−1.023	0.359	0.256	0.504	< 0.001*
Sex#	1.139	3.124	1.116	9.742	0.030*
Smoking status†	0.450	1.569	0.521	4.726	0.424

#Male was denoted as 0, and Female as 1. The Odds Ratio was 3.124 means that female showed higher likelihood of EGFR (+).
† Smoker was denoted as 0, and Never smoker as 1. The Odds Ratio was 1.569 means that Never smoker showed higher likelihood of EGFR (+).
** P-value < 0.05, which showed significance.*

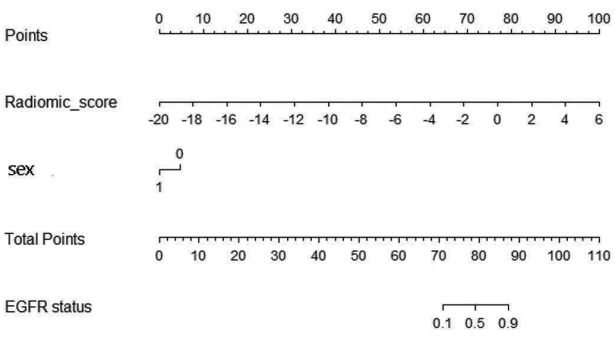


FIGURE 4 | Radiomic nomogram. In the training cohort, the nomogram incorporated the radiomic signature and sex.

DISCUSSION

The NCCN (2019, v3) recommended that testing for EGFR mutations should be applied in patients with non-squamous NSCLC or NSCLC NOS (not otherwise specified) so that patients with this genetic abnormality can receive effective treatment with targeted agents. Although patients with advanced adenocarcinoma benefit most from TKIs, accessibility to obtain transbronchial or transthoracic biopsy samples is not always satisfactory or safe in these patients. The adverse event rate in thoracic biopsy was reported to be 17.1% (13), and sufficient tissue for molecular analysis can only be obtained in 20–50% of NSCLC patients, even in large well-designed clinical trials (14). In addition, the heterogeneity of the tumor may mislead the clinical decision (15, 16).

We developed and validated a radiomics signature-based nomogram for the non-invasive detection of EGFR mutations in patients with advanced adenocarcinoma through preprocessing, parameters screening and model building from CT images. In the validation cohort, the AUC of radiomics signature was 0.851 (95% CI, 0.750 to 0.951). Previous studies have demonstrated such correlations in all stages of peripheral lung adenocarcinoma (17, 18), with AUC of 0.709 (95% CI, 0.645 to 0.766) and 0.751 (95% CI, 0.631 to 0.848), respectively. For early-stage resectable adenocarcinoma, the detection is less important, whereas for advanced-stage patients with EGFR mutations, TKIs are the first-line standard modality for the treatment today (19), so the detection is urgently needed. Thus, it is of greater significance to establish relatively inexpensive and safe imaging biomarker for the advanced-stage patients to help making treatment decision. However, stage selection brought limitation at the same time. The signature could not act as an alone biomarker in patients with unknown pulmonary nodules and it is also a time-consuming thing to stage before using the biomarker.

Previous articles on pulmonary tumor radiomics were generally based on non-contrast CT images (20–22). Some studies have used contrast images alone (23, 24), and some have used both, but no comparisons or descriptions regarding which type of image is better for further analysis have been reported (25). In this study, we managed contrast and non-contrast data separately and compared their diagnostic value using ROC curve analysis; we finally chose the contrast data for subsequent analysis. This result is consistent with clinical applications. Contrast-enhanced CT can be used to better delineate and

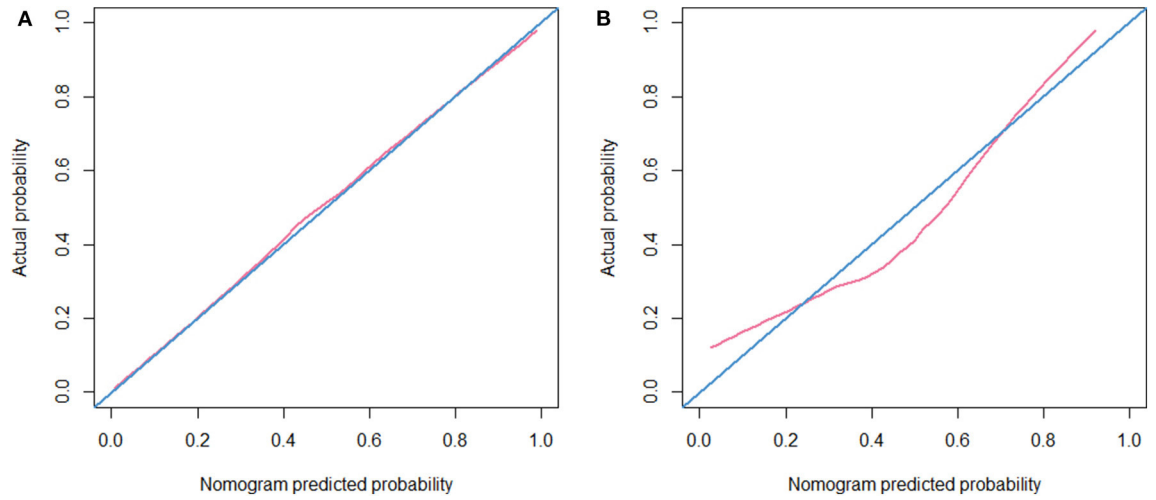


FIGURE 5 | Calibration curve. **(A)** Calibration curve of the nomogram in the training cohort. **(B)** Calibration curve of the nomogram in the validation cohort.

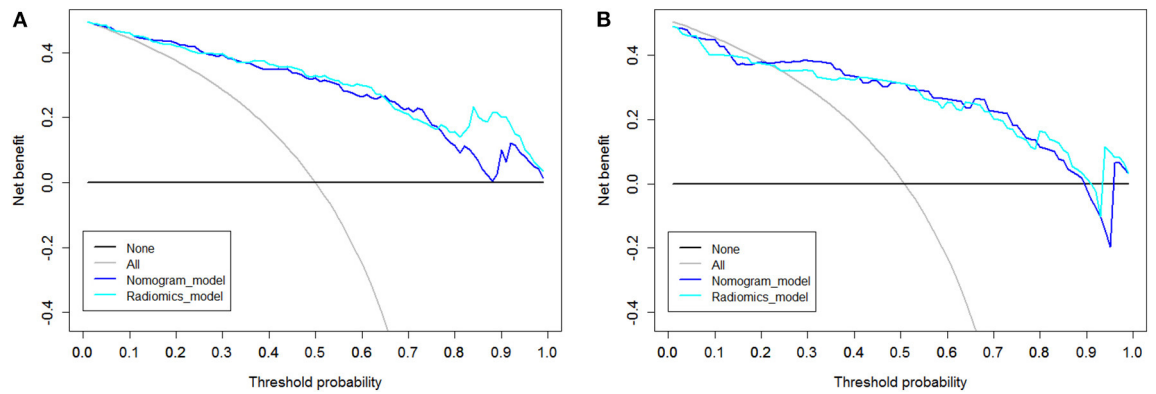


FIGURE 6 | Decision curve analysis (DCA). The y axis represents the net benefit, which was determined by calculating the difference between the expected benefit and the expected harm associated with each proposed model [net benefit = true-positive rate (TPR) – (false-positive rate (FPR)× weighting factor), where the weighting factor = threshold probability/ (1-threshold probability)]. The gray line represents the assumption that all tumors were EGFR (+) (the treat-all scheme). The black line represents the assumption that all tumors were EGFR (-) (the treat-none scheme). **(A)** DCA in the training cohort. Using the radiomic model and the nomogram model to predict the EGFR status added more benefit than using the treat-all scheme or the treat-none scheme at any given threshold probability. **(B)** DCA in the validation cohort. For threshold probabilities >20%, using the radiomic model and the nomogram model to predict the EGFR status added more benefit than using the treat-all scheme or the treat-none scheme.

define tumor regions in relationship to surrounding structures than non-contrast CT, and also demonstrates the increased vascularity that occurs within malignancies and provides additional information on the tumor's physiology and active blood supply. All of this information is reflected by radiomic features, leading to better models.

However, some limitations to this work still exist. First, although image acquisition was confined to two CT systems and all the images were preprocessed before segmentation, differences between devices may influence the results. Second, in the baseline clinical characteristics, there was no significant difference in the overall distribution of age, sex, smoking status or stage between the training and validation cohorts, thus we believed that there was no bias for the training and validation cohorts. But when taking into consideration the distribution in mutant and wild-type EGFR patients, sex and smoking status showed significant differences between the two groups in the training cohort but no significance in the validation cohort, which we considered may due to the small sample size in the validation cohort. Third, less sample size and lack of external validation of the model, more multicenter studies and prospective studies should be carried out to increase the generalizability and robustness of the radiomic findings. Fourth, all samples were obtained through biopsy. They were smaller than those obtained by surgery, which could better represent the tumor heterogeneity. Further studies may also include testing for cell-free tumor DNA (ctDNA) and circulating tumor cells (CTCs) to ensure the homogeneity of mutations (14).

CONCLUSION

In conclusion, radiomics signature can help to distinguish between EGFR positive and wild type advanced lung adenocarcinomas. Compared with non-contrast CT, contrast-enhanced CT provided more value for radiomic predication.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of the First Hospital of China Medical University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

DH conceived of the project, performed the experiments and wrote the paper. YG and XW analyzed the data. KX and LZ provided expert guidance and reviewed the manuscript. All the authors gave the final approval of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00028/full#supplementary-material

Figure S1 | Details of four types of radiomic features extracted from CT images.

Figure S2 | Feature selection using the LASSO algorithm in the training cohort for contrast images. **(A)** The 10-fold cross-validation process was repeated to select the optimal penalization coefficient lambda. The value of lambda yielded the minimum average binomial deviance that was used to select features. **(B)** LASSO coefficient profile plot of the 396 features against log(lambda); the optimal lambda resulted in 10 non-zero features.

Table S1 | The formulas for calculation of the radiomics signature.

REFERENCES

1. Didkowska J, Wojciechowska U, Manczuk M, Lobaszewski J. Lung cancer epidemiology: contemporary and future challenges worldwide. *Ann Transl Med*. (2016) 4:150. doi: 10.21037/atm.2016.03.11
2. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger K, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society: international multidisciplinary classification of lung adenocarcinoma: executive summary. *Proc Am Thorac Soc*. (2011) 8:381–5. doi: 10.1513/pats.201107-042ST
3. Ramon Andrade DM, Filipa Soares P, Dania Sofia M, Júlio O, Ana R, Marta S, et al. EGFR exon mutation distribution and outcome in non-small-cell lung cancer: a Portuguese retrospective study. *Tumor Biol*. (2012) 33:2061–8. doi: 10.1007/s13277-012-0465-5
4. Hong SJ, Kim TJ, Choi YW, Park JS, Chung JH, Lee KW. Radiogenomic correlation in lung adenocarcinoma with epidermal growth factor receptor mutations: imaging features and histological subtypes. *Eur Radiol*. (2016) 26:3660–8. doi: 10.1007/s00330-015-4196-z
5. Li S, Ding C, Zhang H, Song J, Wu L. Radiomics for the prediction of EGFR mutation subtypes in non-small cell lung cancer. *Med Phys*. (2019) 46:4545–52. doi: 10.1002/mp.13747
6. Sequist LV, James Chih-Hsin Y, Nobuyuki Y, Kenneth OB, Vera H, Tony M, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol*. (2013) 31:3327. doi: 10.1200/JCO.2012.44.2806
7. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol*. (2012) 13:239–46. doi: 10.1016/S1470-2045(11)70393-X
8. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, et al. Exploratory study to identify radiomics classifiers for lung cancer histology. *Front Oncol*. (2016) 6:71. doi: 10.3389/fonc.2016.00071
9. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The eighth edition lung cancer stage classification. *Chest*. (2017) 151:193–203. doi: 10.1016/j.chest.2016.10.010
10. Jia TY, Xiong JF, Li XY, Yu W, Xu ZY, Cai XW, et al. Identifying EGFR mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling. *Eur Radiol*. (2019) 29:4742–50. doi: 10.1007/s00330-019-06024-y
11. Gui J, Li H. Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics*. (2005) 21:3001–8. doi: 10.1093/bioinformatics/bti422
12. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology*. (2016) 281:947–57. doi: 10.1148/radiol.2016152234
13. Overman MJ, Modak J, Kopetz S, Murthy R, Yao JC, Hicks ME, et al. Use of research biopsies in clinical trials: are risks and benefits adequately discussed? *J Clin Oncol*. (2013) 31:17–22. doi: 10.1200/JCO.2012.43.1718
14. Sun W, Yuan X, Tian Y, Wu H, Xu H, Hu G, et al. Non-invasive approaches to monitor EGFR-TKI treatment in non-small-cell lung cancer. *J Hematol Oncol*. (2015) 8:95. doi: 10.1186/s13045-015-0193-6
15. Bai H, Wang Z, Wang Y, Zhuo M, Zhou Q, Duan J, et al. Detection and clinical significance of intratumoral EGFR mutational heterogeneity in Chinese patients with advanced non-small cell lung cancer. *PLoS ONE*. (2013) 8:e54170. doi: 10.1371/journal.pone.0054170
16. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J*. (2019) 53:1800986. doi: 10.1183/13993003.00986-2018
17. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. *Clin Lung Cancer*. (2016) 17:441–8.e6. doi: 10.1016/j.cllc.2016.02.001
18. Sacconi B, Anzidei M, Leonardi A, Boni F, Saba L, Scipione R, et al. Analysis of CT features and quantitative texture analysis in patients with lung adenocarcinoma: a correlation with EGFR mutations and survival rates. *Clin Radiol*. (2017) 72:443–50. doi: 10.1016/j.crad.2017.01.015
19. Fukuoka M, Wu Y-L, Thongprasert S, Sunpaweravong P, Leong S-S, Sriuranpong V, et al. Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer in Asia (IPASS). *J Clin Oncol*. (2011) 29:2866–74. doi: 10.1200/JCO.2010.33.4235
20. Emaminejad N, Qian W, Guan Y, Tan M, Qiu Y, Liu H, et al. Fusion of quantitative image and genomic biomarkers to improve prognosis assessment of early stage lung cancer patients. *IEEE Trans Bio-med Eng*. (2016) 63:1034–43. doi: 10.1109/TBME.2015.2477688
21. Choi W, Oh JH, Riyahi S, Liu CJ, Jiang F, Chen W, et al. Radiomics analysis of pulmonary nodules in low-dose CT for early detection

of lung cancer. *Med Phys*. (2018) 45:1537–49. doi: 10.1002/mp.12820
22. Li XY, Xiong JF, Jia TY, Shen TL, Hou RP, Zhao J, et al. Detection of epithelial growth factor receptor (EGFR) mutations on CT images of patients with lung adenocarcinoma using radiomics and/or multi-level residual convolutionary neural networks. *J Thorac Dis*. (2018) 10:6624–35. doi: 10.21037/jtd.2018.11.03
23. Lee G, Park H, Sohn I, Lee SH, Song SH, Kim H, et al. Comprehensive computed tomography radiomics analysis of lung adenocarcinoma for prognostication. *Oncologist*. (2018) 23:806–13. doi: 10.1634/theoncologist.2017-0538
24. Digumarthy SR, Padole AM, Gullo RL, Sequist LV, Kalra MK. Can CT radiomic analysis in NSCLC predict histology and EGFR mutation status? *Medicine*. (2019) 98:e13963. doi: 10.1097/MD.0000000000013963
25. Chen X, Fang M, Dong D, Wei X, Liu L, Xu X, et al. A Radiomics signature in preoperative predicting degree of tumor differentiationin patients with non-small cell lung cancer. *Acad Radiol*. (2018) 25:1548–55. doi: 10.1016/j.acra.2018.02.019

Conflict of Interest: YG was employed by the company GE Healthcare, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 07 February 2020
doi: 10.3389/fonc.2020.00071

Machine Learning Decision Tree Models for Differentiation of Posterior Fossa Tumors Using Diffusion Histogram Analysis and Structural MRI Findings

Seyedmehdi Payabvash[1,2]*, *Mariam Aboian*[1,2], *Tarik Tihan*[3] and *Soonmee Cha*[2]

[1] Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, United States, [2] Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA, United States, [3] Department of Pathology, University of California, San Francisco, San Francisco, CA, United States

We applied machine learning algorithms for differentiation of posterior fossa tumors using apparent diffusion coefficient (ADC) histogram analysis and structural MRI findings. A total of 256 patients with intra-axial posterior fossa tumors were identified, of whom 248 were included in machine learning analysis, with at least 6 representative subjects per each tumor pathology. The ADC histograms of solid components of tumors, structural MRI findings, and patients' age were applied to construct decision models using Classification and Regression Tree analysis. We also compared different machine learning classification algorithms (i.e., naïve Bayes, random forest, neural networks, support vector machine with linear and polynomial kernel) for dichotomized differentiation of the 5 most common tumors in our cohort: metastasis ($n = 65$), hemangioblastoma ($n = 44$), pilocytic astrocytoma ($n = 43$), ependymoma ($n = 27$), and medulloblastoma ($n = 26$). The decision tree model could differentiate seven tumor histopathologies with terminal nodes yielding up to 90% accurate classification rates. In receiver operating characteristics (ROC) analysis, the decision tree model achieved greater area under the curve (AUC) for differentiation of pilocytic astrocytoma ($p = 0.020$); and atypical teratoid/rhabdoid tumor ATRT ($p = 0.001$) from other types of neoplasms compared to the official clinical report. However, neuroradiologists' interpretations had greater accuracy in differentiating metastases ($p = 0.001$). Among different machine learning algorithms, random forest models yielded the highest accuracy in dichotomized classification of the 5 most common tumor types; and in multiclass differentiation of all tumor types random forest yielded an averaged AUC of 0.961 in training datasets, and 0.873 in validation samples. Our study demonstrates the potential application of machine learning algorithms and decision trees for accurate differentiation of brain tumors based on pretreatment MRI. Using easy to apply and understandable imaging metrics, the proposed decision tree model can help radiologists with differentiation of posterior fossa tumors, especially in tumors with similar qualitative imaging characteristics. In particular, our decision tree model provided more accurate differentiation of pilocytic astrocytomas from ATRT than by neuroradiologists in clinical reads.

Keywords: posterior fossa tumor, diffusion histogram, machine learning, decision tree model, MRI

OPEN ACCESS

Edited by:
Lei Deng,
Jacobi Medical Center, United States

Reviewed by:
Deepak Ranjan Nayak,
Indian Institute of Information Technology Design & Manufacturing Kancheepuram, India
John Crawford,
University of California, San Diego, United States

***Correspondence:**
Seyedmehdi Payabvash
Sam.Payabvash@yale.edu

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *10 August 2019*
Accepted: *15 January 2020*
Published: *07 February 2020*

Citation:
Payabvash S, Aboian M, Tihan T and Cha S (2020) Machine Learning Decision Tree Models for Differentiation of Posterior Fossa Tumors Using Diffusion Histogram Analysis and Structural MRI Findings. Front. Oncol. 10:71. doi: 10.3389/fonc.2020.00071

INTRODUCTION

The current standard of care for patients presenting with posterior fossa tumors is maximal safe resection of tumor, decompression to eliminate mass effect, and radiochemotherapy. While histopathological evaluation is currently the gold standard for brain tumors diagnosis, there is growing body of evidence that combination of quantitative imagining and machine learning algorithms can help with non-invasive differentiation of brain neoplasms based on pre-treatment MRI (1, 2). An accurate presurgical diagnosis can play an important role in surgical planning, determining the extent of resection (3, 4), evaluating the need for neoadjuvant therapy, defining radiation therapy field, and counseling patients and their families (5).

The apparent diffusion coefficient (ADC) values are reflective of tumor cellularity, and help with diagnostic and prognostic assessment of posterior fossa tumors (6, 7). Recent studies demonstrate the added value of quantitative diffusion analysis in differentiation of posterior fossa tumors, besides conventional structural MRI findings such as peritumoral edema, enhancement pattern, location, or extension through the foramina of Luschka/Magendie (8–10). However, prior studies were limited by restricting their analysis pool to select tumor types, analyzing few ADC quantitative metrics (e.g., ADC means, median, or minimum), or only evaluating ADC values on a single slice, thus not accounting for tumor heterogeneity (11, 12).

In current study, we assessed the volumetric voxel-based ADC histogram analysis of the tumor solid components in a large sample of posterior fossa neoplasms. Using machine learning algorithms, we utilized clinical variables, quantitative ADC histogram metrics, and qualitative MRI imaging features extracted by 2 neuroradiologists on presurgical MRI to devise decision trees for accurate diagnosis of posterior fossa tumors. We chose unequivocal imaging metrics, which can be reliably assessed on widely available image viewer software in all hospitals, and thus can be readily used in neuroradiology and neuro-oncology practices. We also compared different machine learning classification models for differentiation of the most common posterior fossa tumors, which presents as a challenge in clinical practice. Including a large number of patients with a variety of pathologies allowed us to devise comprehensive differentiation models that represent a broad range of tumor types and imitate the real-world practice in a tertiary referral center.

METHODS

Patients' Characteristics

Clinical and imaging records of all patients with posterior fossa tumor and surgical pathology results, between January 200 Re reviewer's comment #and December 2015 at our institution, were reviewed. Patients were included if they had (1) intra-axial or intra-ventricular posterior fossa tumor, (2) surgical pathology diagnosis of a neoplasm (**Table 1**), and (3) a presurgical MRI including ADC map, T2-weighted, Fluid Attenuated Inversion Recovery (FLAIR), and post contrast T1-weighted sequences. The exclusion criteria were pathological diagnosis other than a malignant process (e.g., cavernoma), extra-axial location except intraventricular tumors, and an ADC map quality precluding histogram analysis. In addition, tumor pathologies with <6 subjects in our cohort were excluded from univariate, and machine learning analyses (i.e., Choroid plexus papilloma, $n = 4$; Rosette-forming glioneuronal tumor, $n = 2$; Ganglioglioma, $n = 1$; Anaplastic pleomorphic xanthoastrocytoma, $n = 1$). The Institutional Review Board approved the study design, granting a waiver of informed consent given the retrospective nature of study.

MRI Acquisition

The presurgical MRI was performed on 1.5 and 3 Tesla MRI scanners using surgical navigation (BrainLab) imaging protocol—which included axial 2D T1 weighted images, axial diffusion-weighted images (DWI), 3D T2 weighted images, 3D FLAIR, axial susceptibility weighted imaging, dynamic contrast enhancement perfusion, and 3D post contrast T1 sequences. In majority of patients, spin-echo echo-planar DWI was performed in 2D axial plane on a GE Discovery MR750 3T scanner (Waukesha, WI), with image acquisition at b = 0 s/mm^2 and b = 1,000 s/mm^2; repeat time = 8,300 ms, echo time = 65 ms, section thickness of 2 mm, field of view of 250 mm, and matrix size of 128 × 128.

Qualitative Assessment of Posterior Fossa Tumors

All MRI scans were reviewed independently by two board-certified neuroradiologists (SP and MA), each with 8 years of experience in interpretation of brain tumor MRI. Except for the patients' age, the reviewers were blinded to clinical information, radiology report, and pathological diagnosis at the time of review. Both SP and MA predicted the single most likely differential diagnosis for each tumor based on presurgical brain MRI. In addition, the official "clinical report"

TABLE 1 | List of (intra-axial/intra-ventricular) posterior cranial fossa neoplasms ($n = 256$).

Surgical pathology diagnosis	Patients number (frequency)
Metastasis	65 (25.4%)
Hemangioblastoma	44 (17.2%)
Pilocytic astrocytoma	43 (16.8%)
Ependymoma	27 (10.5%)
Medulloblastoma	26 (10.2%)
Low grade glioma/astrocytoma	10 (3.9%)
Lymphoma	8 (3.1%)
Anaplastic astrocytoma	7 (2.7%)
Atypical teratoid/rhabdoid tumor	6 (2.3%)
Glioblastoma multiforme	6 (2.3%)
Subependymoma	6 (2.3%)
Choroid plexus papilloma	4 (1.6%)
Rosette-forming glioneuronal tumor	2 (0.8%)
Ganglioglioma	1 (0.4%)
Anaplastic pleomorphic xanthoastrocytoma	1 (0.4%)

in the electronic medical records were examined to identify the foremost differential diagnosis in the "impression," which considered as the most likely diagnosis for comparison purposes. In addition, the imaging characteristics—listed in **Table 2**—were extracted and corroborated with the official clinical report. In case of discrepancy between the neuroradiologist reviewers and official clinical report, the senior author (SC) reviewed the scan to reach consensus. The lesion morphology was categorized as predominantly solid (>80% solid component), mixed solid and cystic, cystic/necrotic (>80%) with irregular wall, and cystic (>80%) with smooth mural nodule. A "T2 hyperintense" solid component was determined by T2 signal greater than the gray matter (**Figure 1**) (13). The presence of prominent vascular flow void was assessed on T2-weighted images and confirmed on post contrast series. The tumor volumes, including both solid and cystic components, were calculated after manual segmentation on post-contrast T1 images with attention to T2/FLAIR series for non-enhancing component. We also measured the maximum radial width of FLAIR hyperintensity surrounding the tumor on axial slices as a surrogate for peritumoral edema.

ADC Histogram Analysis

On a GE Advantage Workstation (GE healthcare, Milwaukee, WI), we manually segmented the solid component of tumors on ADC maps with attention to post-contrast T1-weighted, T2-weighted, and FLAIR imaging. The volumetric voxel-based ADC histograms of the solid component were calculated and normalized to the average ADC value from cerebrospinal fluid in the body of lateral ventricles, as described previously (8). For each tumor, a total of 24 histogram metrics were calculated—including 21 ADC percentile values with 5 percentile increments (i.e., minimum, 5th, 10th, 15th percentile…) as well as the mean, kurtosis, and skewness. The schematic mean ADC histograms of different tumor types were developed for visual comparison.

Decision Tree Model

For development of decision trees, we applied the "rpart" R package for Classification and Regression Tree (CART) models (14). At each split/node, a variable is selected to maximize the variance explanation of dependent variable. The patients' characteristics (age and gender), structural MRI findings (**Table 2**), and ADC histogram metrics were included as input for the model. By default, a 10-fold cross-validation and fitting at each sub-tree were applied. The final classification of the decision tree model was separately compared with the top differential diagnosis from the clinical report and two independent neuroradiologists using the receiver operating characteristics (ROC) analysis implemented by "pROC" package in R (**Table 3**). We also determined the Cohen's Kappa inter-rater agreement coefficient (**Table 4**). The CART models were first applied for differentiation of all tumors, and then separately for dichotomized classification of the 5 most common posterior fossa tumors. The block diagram in **Figure 2** summarizes the analysis steps in decision tree and machine learning models.

Machine Learning Classification

Different machine learning models were applied and compared for dichotomized classification of the 5 most common posterior fossa tumors—including the naïve Bayes, random forest, support vector machine (SVM), and neural networks (**Figure 2**). For development of the naïve Bayes models, we applied the "naivebayes" R package with a Laplace smoothing value of 0, as suggested by the package developers. The "randomForest" package was used for random forest ensemble learning classification (15). For each random forest model, 500 trees were constructed applying a randomly-selected one-third of variables at each split (16–19). In our preliminary experiments, the error rate consistently plateaued after constructing 160–300 tree splits in random forests; thus, the default recommendation of 500 trees by the R package deemed adequate to achieve optimal accuracy in our models. For SVM algorithms, we used the "e1071" R package to construct dichotomized classification models. We applied both linear and non-linear kernels for data classification—in this series, polynomial kernel was used for non-linear kernel function. During finetuning of the hyperparameters for SVM models, a cost of 0.1 yielded the optimal error rate, and was applied for all linear kernels. For polynomial kernel, we used a sigma of 1 as the optimal cutoff. For neural networks, we applied the "neuralnet" package, and used the "rBayesianOptimization" package to optimize the number of nodes in the neural network hidden layer.

In order to present realistic estimates from our cohort and minimize the risk of overfitting, we opted to report the averaged results from "stratified" cross-validation. Given the uneven distribution of tumor types in our cohort, a "stratified" sampling strategy seemed appropriate to ensure that enough number of each tumor type is allocated in every training and validation sample. Using stratified random sampling, we applied 5-fold cross validation, preserving the tumor subtype percentage in both training and validation samples. The random sampling was repeated 100 times, and the averaged results from "stratified" cross validation across 500 permutations are presented. In each pair of randomly selected training/validation samples, the model was constructed on training sample, and tested on corresponding validation sample. A confusion matrix was constructed based on prediction results in each training and validation sample, and the corresponding accuracy (number of correctly classified subjects divided by sample size), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated. In addition, the ROC area under the curve (AUC) and 95% confidence interval (CI) were computed using 2000 stratified bootstrap replicates per the default implementation in pROC package. The average test characteristics across 500 training and validation samples are reported.

All models were first applied and compared for differentiation of the five most common posterior fossa tumors—with each combination of tumor type and machine learning analyzed separately. Given that random forest models yielded higher accuracy for classification of tumor types compared to other algorithms, we applied the random forest for multi-class

TABLE 2 | Structural MRI findings and clinical characteristics among various posterior fossa neoplasms.

	MET (*n* = 65)	HB (*n* = 44)	PA (*n* = 43)	EP (*n* = 27)	MB (*n* = 26)	LGG (*n* = 10)	LYM (*n* = 8)	AA (*n* = 7)	ATRT (*n* = 6)	GBM (*n* = 6)	SEP (*n* = 6)	*P*-value
Patients' characteristics												
Age (years)	57.6±12.2	49.3±17.9	18.7±11.2	26.1±20.1	21.8±16.8	35.6±27.9	63.2±12.9	36.1±23.7	1.3±1.0	31.7±22.3	50.2±13.3	**<0.001**
Gender (male)	18 (33%)	19 (49%)	11 (61%)	2 (22%)	3 (50%)	2 (40%)	1 (25%)	2 (22%)	3 (50%)	2 (40%)	1 (25%)	0.304
Tumor localization												
Cerebellar hemisphere	54 (83%)	39 (89%)	18 (42%)	3 (11%)	9 (35%)	5 (50%)	6 (75%)	1 (14%)	2 (33%)	4 (67%)	0 (0%)	**<0.001**
Fourth ventricle	3 (5%)	3 (7%)	13 (30%)	21 (78%)	17 (65%)	0 (0%)	1 (13%)	0 (0%)	4 (67%)	0 (0%)	6 (100%)	**<0.001**
Vermis/midline	4 (6%)	2 (5%)	10 (23%)	2 (7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**0.013**
Brainstem	4 (6%)	0 (0%)	2 (5%)	1 (4%)	0 (0%)	5 (50%)	1 (13%)	6 (86%)	0 (0%)	2 (33%)	0 (0%)	**<0.001**
Cerebellar peduncle involvement	4 (6%)	6 (14%)	4 (9%)	4 (15%)	5 (19%)	5 (50%)	1 (13%)	4 (57%)	2 (33%)	4 (67%)	0 (0%)	**<0.001**
Lesion morphology												
Predominantly solid (>80%)	46 (71%)	6 (14%)	7 (16%)	13 (48%)	17 (65%)	6 (60%)	8 (100%)	6 (86%)	3 (50%)	4 (67%)	6 (100%)	**<0.001**
Mixed solid and cystic	6 (9%)	12 (27%)	20 (47%)	14 (52%)	9 (35%)	4 (40%)	0 (0%)	1 (14%)	3 (50%)	1 (17%)	0 (0%)	**<0.001**
Cystic (>80%) with mural nodule	4 (6%)	26 (59%)	15 (35%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**<0.001**
Necrotic with irregular wall	9 (14%)	0 (0%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (17%)	0 (0%)	**0.013**
Enhancement pattern												
Homogenous enhancement	16 (25%)	39 (89%)	8 (19%)	0 (0%)	5 (19%)	0 (0%)	5 (62%)	1 (14%)	0 (0%)	0 (0%)	1 (17%)	**<0.001**
Heterogeneous enhancement	49 (75%)	5 (11%)	35 (81%)	26 (96%)	20 (77%)	6 (60%)	3 (38%)	4 (57%)	6 (100%)	5 (83%)	5 (83%)	**<0.001**
No enhancement	0 (0%)	0 (0%)	0 (0%)	1 (4%)	1 (4%)	4 (40%)	0 (0%)	2 (29%)	0 (0%)	1 (17%)	0 (0%)	**<0.001**
Extension along the neuroaxis												
Multiple lesions	16 (25%)	4 (9%)	2 (5%)	0 (0%)	2 (8%)	1 (10%)	2 (25%)	1 (14%)	0 (0%)	2 (33%)	0 (0%)	**0.016**
Leptomeningeal drop metastasis	5 (8%)	0 (0%)	1 (2%)	0 (0%)	1 (4%)	1 (10%)	0 (0%)	0 (0%)	0 (0%)	1 (17%)	0 (0%)	0.354
T2/FLAIR findings												
Prominent vascular flow voids	2 (3%)	25 (57%)	0 (0%)	3 (11%)	6 (23%)	0 (0%)	0 (0%)	0 (0%)	2 (33%)	0 (0%)	0 (0%)	**<0.001**
Surrounding FLAIR (cm)	1.8±0.9	1.7±0.9	0.5±0.5	0.4±0.6	0.7±0.6	0.5±0.4	1.8±0.6	1.0±1.0	0.4±0.3	1.0±0.9	0.1±0.1	**<0.001**
T2 hyperintense solid component*	9 (14%)	17 (39%)	33 (77%)	10 (37%)	6 (23%)	10 (100%)	1 (13%)	7 (100%)	0 (0%)	1 (17%)	0 (0%)	**<0.001**
Mass effect												
Volume (mL)	11.4±9.3	19.8±15.5	29.7±30.1	21.6±18.0	27.4±19.8	14.2±13.8	10.6±9.1	21.4±12.3	42.8±35.6	12±9.0	4.8±4.3	**<0.001**
Hydrocephalus	22 (34%)	23 (52%)	27 (63%)	20 (74%)	19 (73%)	5 (50%)	2 (25%)	4 (57%)	5 (83%)	2 (33%)	0 (0%)	**<0.001**

Results of univariate comparison between different neoplasm types.

**A "T2 hyperintense" solid component was determined by T2 signal greater than the gray matter (13).*

AA, anaplastic astrocytoma; ATRT, atypical teratoid/rhabdoid tumors; EP, Ependymoma; GBM, glioblastoma multiforme; HB, hemangioblastoma; LGG, low-grade glioma/astrocytoma; LYM, lymphoma; MB, medulloblastoma; MET, metastasis; PA, pilocytic astrocytoma; SEP, subependymoma.

A p-value < 0.05 was considered statistically significant, and depicted in bold.

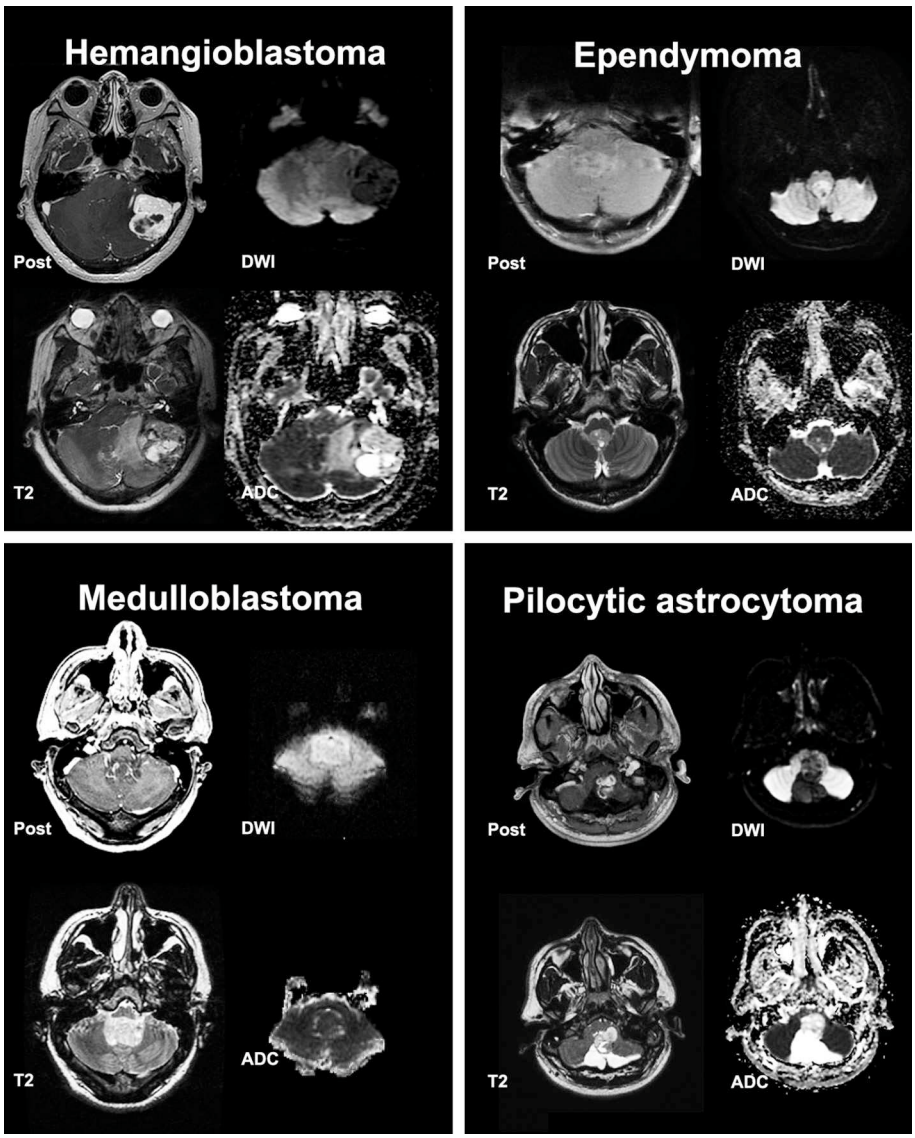


FIGURE 1 | Post contrast T1, T2, Diffusion Weighted Imaging (DWI) and Apparent Diffusion Coefficient (ADC) scans from representative posterior cranial fossa tumors.

differentiation of all tumor types. The average multiclass ROC AUC was determined for random forest models and neuroradiologist interpretations using the "multiROC" package. Notably, comparison and calculation of 95% CI for multiclass averaged AUC is not feasible. In addition, the averaged "mean decrease in Gini coefficient" are reported to depict the relative effect of each variable on random forest model accuracy if the variable is deleted.

Statistical Analysis

The data are expressed as mean ± standard deviation, and frequency (percentage). Kolmogorov–Smirnov test confirmed normal distribution of continuous variables in our analysis. For univariate comparison between different tumor types, the ANOVA with Tukey *post-hoc* analysis was used for continuous variables, and Chi square test was used for nominal variables. MANOVA was applied to evaluate the effects of 1.5 vs. 3 Tesla scanners on ADC measurement. In addition to R package (https://cran.r-project.org/), we used SPSS 22.0 (IBM, Somers, NY) for statistical analysis.

RESULTS

Posterior Fossa Tumors

Of 403 consecutive patients with pathologic diagnosis of posterior cranial fossa neoplasm over 12-year period, 256 patients had intra-axial/intra-ventricular tumors. We excluded 136 patients with extra-axial tumors (except intraventricular tumors), and 11 subjects with poor quality of MRI. Excluded extra-axial tumors were schwannoma, meningioma, metastases,

TABLE 3 | Comparing the accuracy of decision tree model (**Figure 5**) with the official clinical interpretation and independent neuroradiologist reviewers.

Decision tree model		Clinical report		Reader #1		Reader #2	
	ROC AUC (95% CI)	ROC AUC (95% CI)	*P*-value	ROC AUC (95% CI)	*P*-value	ROC AUC (95% CI)	*P*-value
Metastasis	0.857 (0.810–0.904)	0.958 (0.935–0.982)	**0.001**	0.917 (0.877–0.957)	0.056	0.956 (0.935–0.975)	**0.001**
Hemangioblastoma	0.885 (0.826–0.944)	0.891 (0.829–0.951)	0.9148	0.913 (0.858–0.968)	0.510	0.847 (0.779–0.916)	0.448
Pilocytic astrocytoma	0.885 (0.825–0.946)	0.792 (0.718–0.867)	**0.020**	0.855 (0.788–0.922)	0.403	0.830 (0.759–0.901)	0.151
Ependymoma	0.759 (0.663–0.855)	0.857 (0.780–0.934)	0.0841	0.773 (0.678–0.867)	0.838	0.876 (0.801–0.953)	**0.018**
Medulloblastoma	0.8545 (0.767–0.942)	0.788 (0.692–0.884)	0.345	0.856 (0.769_0.944)	0.969	0.912 (0.841–0.983)	0.178
Low grade glioma/astrocytoma	0.6358 (0.515–0.785)	0.721 (0.557–0.885)	0.392	0.737 (0.574–0.901)	0.308	0.815 (0.664–0.965)	0.1049
Atypical teratoid/rhabdoid tumor	0.913 (0.749–1.000)	0.579 (0.415–0.742)	**0.001**	0.742 (0.522–0.961)	0.306	0.750 (0.531–0.969)	0.329

*The receiver operating characteristics (ROC) area under the curve (AUC) with 95% confidence interval (CI) were calculated for the decision tree model (**Figure 5**) vs. official clinical interpretation, and independent neuroradiologists (separately) in differentiation of posterior fossa tumors. A p-value < 0.05 was considered statistically significant, and depicted in bold.*

TABLE 4 | Inter-rater agreement in differentiation of posterior cranial fossa neoplasms.

Diagnosis	Cohen's Kappa		
	Clinical report vs Reader #1	Clinical report vs Reader #2	Reader #1 vs Reader #2
Metastasis	0.802	0.839	0.787
Hemangioblastoma	0.788	0.820	0.728
Pilocytic astrocytoma	0.749	0.662	0.679
Ependymoma	0.664	0.781	0.677
Medulloblastoma	0.724	0.615	0.823
Low grade glioma/astrocytoma	0.718	0.569	0.423
Lymphoma	0.388	−0.006	−0.006
Anaplastic astrocytoma	0.435	−0.006	−0.006
Atypical teratoid/rhabdoid tumor	0.057	−0.012	0.593
Glioblastoma multiforme	0.559	0.535	0.291
Subependymoma	0.000	0.000	0.063
Choroid plexus papilloma	−0.008	−0.005	−0.005
Ganglioglioma	0.000	0.000	0.000
Rosette-forming glioneuronal tumor	0.000	0.000	0.000

The Cohen's Kappa co-efficient for inter-rater agreement in classification of each neoplasm type among the official clinical report, and two independent neuroradiologist reviewers.

and hemangiopericytoma. Among tumors included in our analysis (**Table 1**), metastasis, hemangioblastoma, pilocytic astrocytoma, ependymoma, and medulloblastoma were the most common types, comprising 205/256 (80%) subjects. Representative tumors from different pathologies are depicted in **Figure 1**.

Patients' Characteristics and Qualitative MRI Analysis

A summary of the univariate analysis comparing various posterior fossa tumors is shown in **Table 2**. The patients' age at presentation, tumor lesion localization, tumor morphology, enhancement pattern, degree of peritumoral FLAIR hyperintensity, whole tumor volume, and presence of hydrocephalus were significantly different among tumor types in univariate analyses. The results of *post-hoc* analysis for patients' age, peri-tumor FLAIR hyperintensity width, and tumor volume between different neoplasms are depicted in **Figure 3**.

ADC Histogram Analysis

Figure 4 depicts the schematic representation of the averaged ADC percentile values among different posterior fossa tumors. Medulloblastomas, followed by ATRT and lymphomas had the lowest ADC histogram percentile values; whereas, pilocytic astrocytomas, followed by hemangioblastomas had the highest ADC histogram percentile values (**Figure 4**). Using ANOVA, there was significant difference in all ADC percentile metrics, average, skewness, and kurtosis values among 11 different tumor types (with $\geq$6 subjects) in our cohort (p-values < 0.001). There has been no significant difference in ADC histogram metrics between DWI series from 1.5 Tesla ($n = 34$) vs. 3 Tesla ($n = 214$) scanners in MANOVA.

Decision Tree Models for Differentiation of Posterior Fossa Tumors

The CART decision tree model successfully differentiated 7 types of neoplasms in our cohort (**Figure 5**). The first

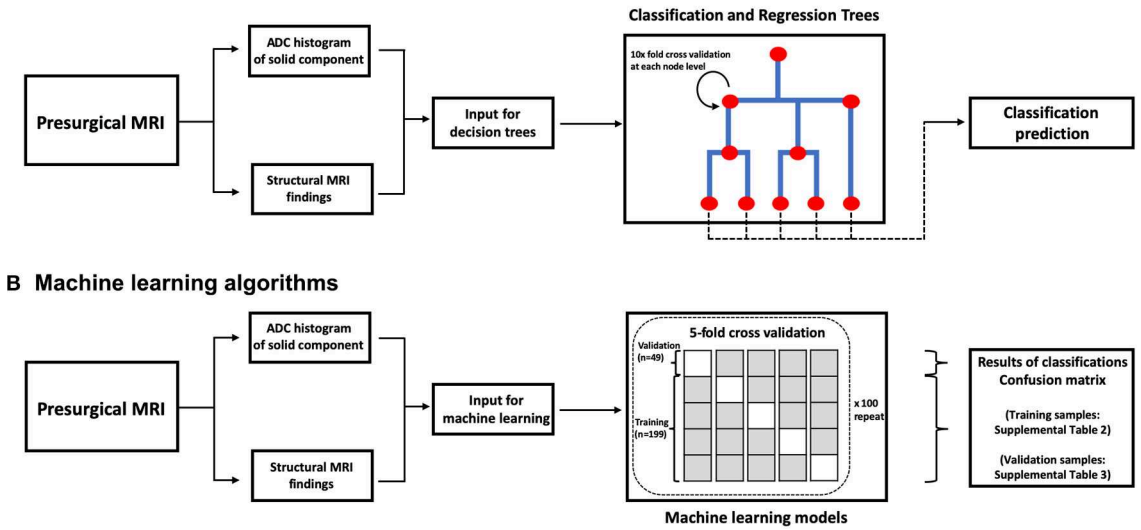


FIGURE 2 | A block diagram of decision tree model **(A)** and machine learning algorithms **(B)**. ADC, apparent diffusion coefficient.

decision node identified by the model was patients' age with a cut off of 35 years. Subsequent nodes used ADC histogram values, presence/absence of prominent flow voids on T2 weighted images, homogenous enhancement pattern, solid tumor morphology, and the fourth ventricle localization, respectively, for further tumor classification (**Figure 5**). The terminal nodes (leaves) of the decision tree yielded 30 to 90% correct classification ratios. Moreover, the likelihood of each tumor type, based on classification criteria is calculated in each terminal node (**Figure 5**).

Using ROC curve analysis, we compared the accuracy of the decision tree model with clinical interpretation, and each of independent neuroradiologists (**Table 3**). The decision tree model yielded a greater AUC compared to the clinical interpretation in differentiation of the pilocytic astrocytoma ($p = 0.020$) and ATRT ($p = 0.001$) from other neoplasm subtypes; whereas, the clinical interpretation and reviewer #2 had higher ROC AUC in differentiation of metastasis ($p = 0.001$) from other tumors. The Cohen's Kappa analysis, showed substantial inter-rater agreement (>0.6) between the clinical interpretation and neuroradiologists among the 5 most common tumor types; however, the agreement rates were lower for the less common tumors (**Table 4**).

In order to further delineate specific imaging characteristics of common posterior fossa tumors and achieve higher classification accuracy, we also developed separate CART decision trees models for dichotomized classification of the 5 most common tumors in our cohort (**Figure 6**). The patients' age, ADC histogram metrics, peritumoral FLAIR hyperintensity width, presence of prominent flow void, enhancement pattern, presence of cystic component, fourth ventricle location, cerebellar hemisphere localization, extension through foramina of Luschka/Magendie, and tumor volume, were included in these CART decision tree models (**Figure 6**).

Machine Learning Algorithm for Tumor Classification

The ratio of tumor types included in the stratified training ($n = 199$) and validation ($n = 49$) datasets are tabulated in **Supplemental Table 1**. In separate classification models devised for the dichotomized differentiation of the 5 most common posterior fossa tumors, random forest models achieved the highest ROC AUC, sensitivity, specificity, PPV, and NPV across training and validation samples from the ×100 repeat of 5-fold stratified cross validation (**Figure 7**, **Supplemental Tables 2, 3**).

Then, we applied random forest model for multiclass differentiation of posterior fossa tumor types. Using a multiclass ROC analysis, the average AUC of random forest models was 0.961 in training datasets, and 0.873 in validation dataset. Using multiclass ROC analysis in same 248 patients, the average AUC of clinical interpretation, reviewer #1, and reviewer #2 were 0.832, 0.799, and 0.834, respectively. There was significant correlation between pathological diagnosis and random forest model prediction in the training (averaged $r = 0.96$, $p < 0.001$), and validation (averaged $r = 0.51$, $p < 0.001$) datasets. The patients' age, width of peritumoral FLAIR hyperintensity, cerebellar hemisphere location, involvement of cerebellar peduncle, tumor volume, and ADC histogram metrics had the greatest impact on accuracy of random forest models (**Figure 8**).

DISCUSSION

Using the CART decision tree model analysis, we have devised differentiation algorithms for posterior fossa tumors based on patients' age, ADC histogram analysis, and qualitative imaging features on pretreatment MRI. The proposed decision tree model (in **Figure 5**) could differentiate 7 histopathologies with 30 to 90 % accurate classification rates in terminal nodes. This

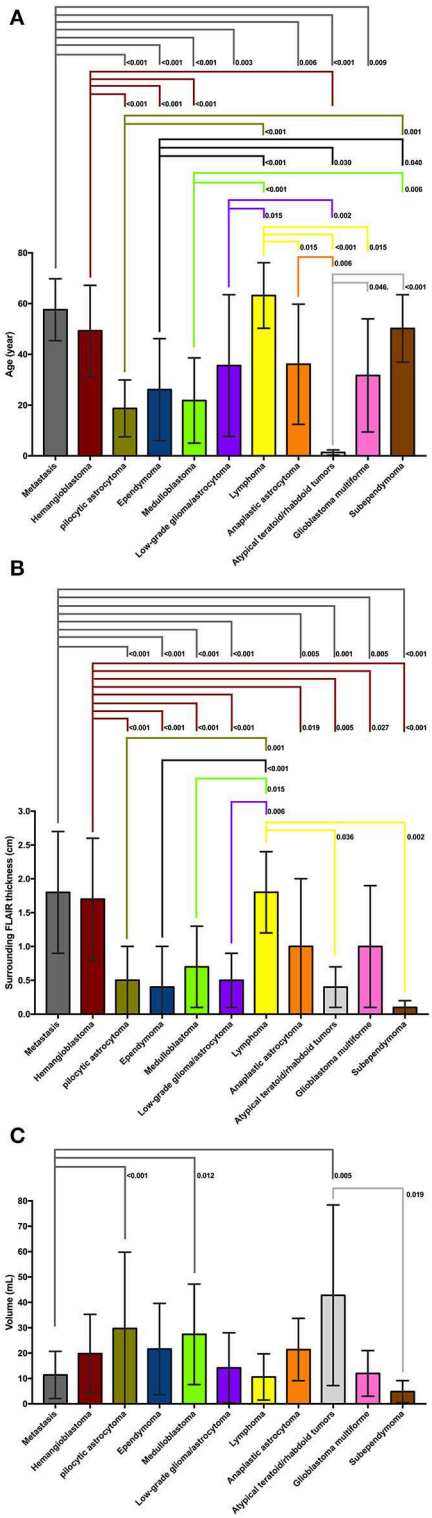


FIGURE 3 | (A) On *post-hoc* analysis, patients with atypical teratoid/rhabdoid tumors (ATRT) had significantly lower age at presentation compare to all other tumor types except medulloblastoma and pilocytic astrocytoma. Patients with medulloblastoma, pilocytic astrocytoma, and ependymoma were significantly younger compared to those with metastasis, lymphoma, hemangioblastoma, *(Continued)*

FIGURE 3 | and subependymoma. Patients with metastasis and hemangioblastomas were also older than those with anaplastic astrocytoma, low-grade glioma, and glioblastoma multiforme. **(B)** On *post hoc* analysis, metastases, hemangioblastomas, and lymphomas had larger diameter of peritumoral FLAIR hyperintensity compared to medulloblastoma, pilocytic astrocytomas, ependymomas, low-grade glioma, ATRT, and subependymomas—likely since latter tumors tend to be intraventricular with virtually no peritumoral edema. Also, the peritumoral FLAIR hyperintensity surrounding metastases, and hemangioblastomas was larger in diameter compared to anaplastic astrocytoma, and glioblastoma multiforme. **(C)** On *post hoc* analysis of tumor volumes, pilocytic astrocytomas, medulloblastomas, and ATRTs had larger size compared to metastases. The ATRTs were also significantly larger compared to subependymomas. ATRT, atypical teratoid/rhabdoid tumors; FLAIR, fluid attenuated inversion recovery.

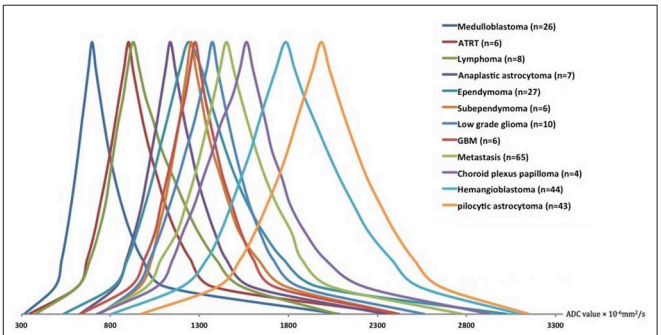


FIGURE 4 | The schematic representation of the averaged ADC histogram distribution among different posterior fossa neoplasms. Medulloblastomas, ATRT, and lymphomas had the lowest; whereas, pilocytic astrocytomas, and hemangioblastomas had the highest ADC histogram percentile values. The average percentile values for each tumor type were calculated, and representative averaged histograms were modified so that the median values would be depicted at the same height on the y axis. ADC, apparent diffusion coefficient; ATRT, atypical teratoid/rhabdoid tumors; GBM, glioblastoma multiforme.

decision tree model appears to be most helpful in differentiation of pilocytic astrocytoma and ATRT - as it achieved higher accuracy compared to clinical report in our cohort. We also demonstrated the feasibility of random forest machine learning algorithms in devising classification models for differentiation of posterior fossa tumors. Applying multiclass ROC analysis, we achieved an averaged AUC of 0.961 in training datasets, and 0.873 in validation dataset, as compared to 0.799 and 0.834 by neuroradiologists.

Recent studies demonstrated the feasibility of machine learning algorithms in prediction of glioma histopathological grade, and classification of the most common pediatric posterior fossa tumors (20–22). These studies, however, utilized small training datasets with few select types of tumors, therefore clinical application of these models may be limited (23). In our study, we used a large comprehensive cohort of patients including 15 different types of posterior fossa tumors and at least 6 representative patients for each tumor histology. Our decision tree models rely on qualitative imaging features and quantitative histogram analysis, which can be easily translated to commercially-available image viewer systems in clinical practice; thus, providing a ready-to-apply tool for neuroradiologists to

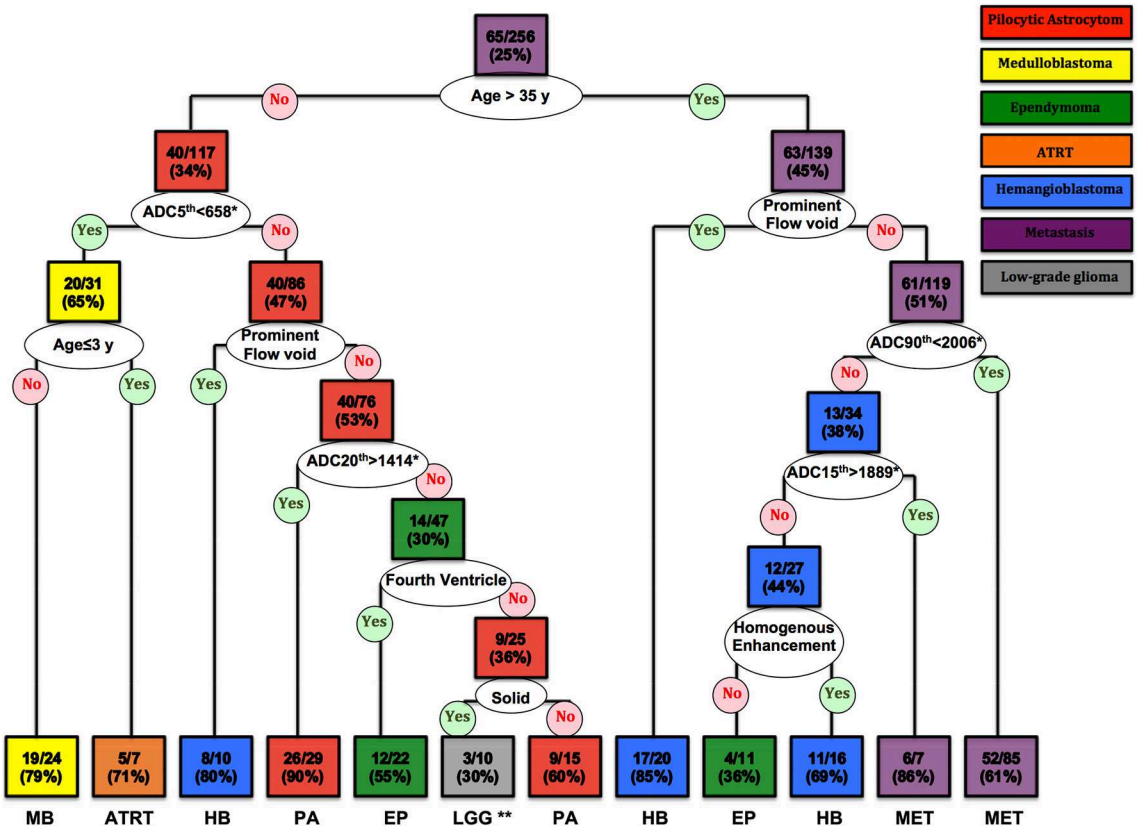


FIGURE 5 | The CART decision tree model could differentiate 7 histopathologies among posterior fossa tumors. In each box (node/leaf), the accurate classification rate (positive predictive value) is expressed as ratios of the most common tumor type per total number of cases fulfilling the criteria. Each box is colored to represent the tumor type, and the intensity of the color reflects the accurate classification ratio. ADC, apparent diffusion coefficient; ATRT, atypical teratoid/rhabdoid tumors; CART, Classification and Regression Tree; EP, Ependymoma; HB, Hemangioblastoma; LGG, low-grade glioma/astrocytoma; MET, metastases; MB, Medulloblastoma; PA, pilocytic astrocytoma. *The ADC n^{th} refers to the ADC histogram n^{th} percentile value $\times 10^{-6}$ mm^2/s; for example, ADC 5th percentile $<658 \times 10^{-6}$ mm^2/s. **Both low-grade glioma and anaplastic astrocytoma had 3/10 (30%) ratios in this terminal node.

formulate their differential diagnosis before pathology results becomes available. The decision tree model and the "mean decrease in Gini coefficient" in random forest models also provide an insight into the innerworkings of machine learning models in their prediction decision.

The patients' age is one of the most important factors in differentiation of posterior fossa tumors, and it is well established that adult and pediatric patients are prone to different types of posterior fossa tumors. In this study, instead of using preset age cutoffs as inclusion or exclusion criteria, we applied CART models to identify data driven and tumor-specific age thresholds for differentiation of various neoplasms; and indeed, an age cutoff of 35 years was the first step in decision tree model for differentiation of various tumor types (**Figure 5**). Moreover, in our cohort, an age cutoff of ≤3 years was identified as differentiation criteria for ATRT from other tumors—including medulloblastoma (**Figure 5**); and an age cutoff of <27 years was helpful for differentiation of pilocytic astrocytoma from rest of tumors (**Figure 6C**).

While similar qualitative and quantitative imaging characteristics were previously used to differentiate posterior fossa tumors, our decision tree model provides step-wise approach for differentiation of wide various posterior fossa tumors (**Figures 5**, **6**, **8**). For example, high ADC values in solid component of the tumor and young age at presentation could help differentiate pilocytic astrocytoma from other tumors in posterior fossa (**Figures 5**, **6C**). In dichotomized analysis, age younger than 27 years, high ADC percentile values (10th percentile $>1{,}055 \times 10^{-6}$ mm^2/s and 95th percentile $>2{,}805 \times 10^{-6}$ mm^2/s), and presence of cystic component had 96% accurate classification rate (positive predictive value) for differentiation of pilocytic astrocytoma from other tumors (**Figure 6C**). On the other hand, age at presentation of <35 year and low ADC values are suggestive of medulloblastoma (or ATRT), regardless of extension through foramina of Luschka/Magendie, tumor localization, or enhancement pattern (**Figures 5**, **6E**). Although the number of ATRT patients in our cohort was too small to draw a firm conclusion, we found that an age cutoff ≤3 years can help differentiate ATRT from other neoplasms of posterior fossa—including medulloblastoma (**Figure 5**). Indeed, an ADC 5th percentile $<658 \times 10^{-6}$ mm^2/s and

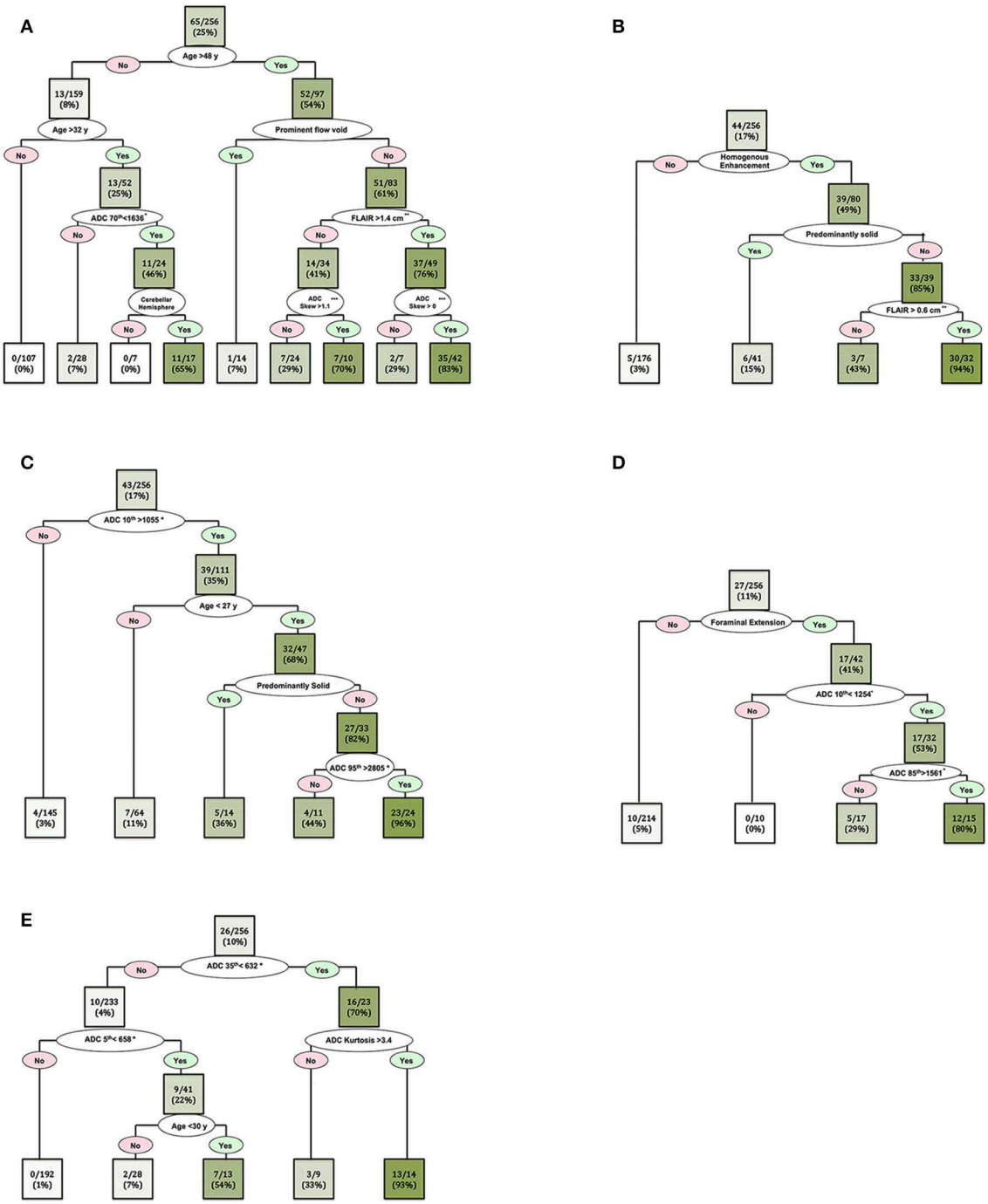


FIGURE 6 | Decision tree models for dichotomized differentiation of the 5 most common posterior cranial fossa neoplasms: **(A)** metastasis, **(B)** hemangioblastomas, **(C)** pilocytic astrocytoma, **(D)** ependymomas, and **(E)** medulloblastomas. In each box (node/leaf), the accurate classification rate (positive predictive value) is expressed as ratios of the most common tumor type per total number of cases fulfilling the criteria. The intensity of the green color reflects the accurate classification ratio. Froaminal extension refers to tumoral extension through the foramina of Luschka and/or Magendie. *The ADC n^{th} refers to the ADC histogram n^{th} percentile value $\times 10^{-6}$ mm^2/s; for example, ADC 70th percentile $<$1,636 $\times 10^{-6}$ mm^2/s. **Peritumoral FLAIR hyperintensity width. ***ADC histogram skewness.

age $\leq$3 years of age had 71% accuracy (positive predictive value) for differentiation of ATRT from other posterior fossa tumors (**Figure 5**).

In this study, we included consecutive patients with posterior cranial fossa tumors, which is a distinction from many prior studies restricting their cohorts based on age, select tumor

Metastasis

	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
Naïve Bayes	0.771	78.0%	48.4%	89.9%	68.2%	82.7%
Random Forest	0.823	83.0%	55.6%	92.6%	73.9%	85.7%
SVM - Linear Kernel	0.793	77.8%	52.1%	88.1%	71.9%	81.0%
SVM - Polynomial Kernel	0.786	80.3%	50.4%	91.4%	67.2%	83.7%
Neural Network	0.781	79.5%	51.0%	91.1%	69.2%	83.2%

Hemangioblastoma

	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
Naïve Bayes	0.864	87.0%	55.2%	97.6%	87.5%	88.6%
Random Forest	0.921	92.7%	63.7%	99.1%	94.5%	92.6%
SVM - Linear Kernel	0.879	88.7%	59.2%	96.6%	90.3%	87.6%
SVM - Polynomial Kernel	0.884	88.0%	59.2%	97.4%	87.3%	91.1%
Neural Network	0.876	90.2%	58.4%	95.9%	89.8%	88.4%

Pilocytic astrocytoma

	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
Naïve Bayes	0.939	95.7%	91.0%	97.4%	92.4%	95.3%
Random Forest	0.991	99.9%	99.2%	99.9%	99.9%	99.8%
SVM - Linear Kernel	0.949	95.9%	94.4%	96.2%	95.9%	94.6%
SVM - Polynomial Kernel	0.946	96.2%	92.7%	98.4%	92.4%	96.8%
Neural Network	0.959	96.4%	93.4%	97.2%	95.4%	96.6%

Ependymoma

	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
Naïve Bayes	0.823	83.9%	32.7%	96.9%	57.1%	86.9%
Random Forest	0.863	90.1%	41.4%	98.6%	63.1%	91.1%
SVM - Linear Kernel	0.808	85.6%	38.1%	94.6%	59.9%	85.4%
SVM - Polynomial Kernel	0.841	85.4%	36.9%	95.9%	57.9%	87.9%
Neural Network	0.828	86.1%	36.8%	95.4%	58.4%	88.9%

Medulloblastoma

	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
Naïve Bayes	0.832	88.0%	45.1%	96.1%	64.0%	91.4%
Random Forest	0.872	92.7%	53.6%	97.1%	71.5%	94.9%
SVM - Linear Kernel	0.830	88.2%	49.3%	93.1%	68.8%	88.4%
SVM - Polynomial Kernel	0.845	90.0%	48.6%	95.6%	66.3%	93.7%
Neural Network	0.835	88.5%	48.8%	95.9%	68.8%	93.2%

FIGURE 7 | Heat map summary for classification performance of different machine learning algorithms in dichotomized differentiation of the 5 most common posterior fossa tumors. The test characteristics were calculated in validation datasets from ×100 repeats of 5-fold cross validation– details in **Supplemental Tables 2, 3**. NPV, negative predictive value; PPV, positive predictive value; SVM, support vector machine.

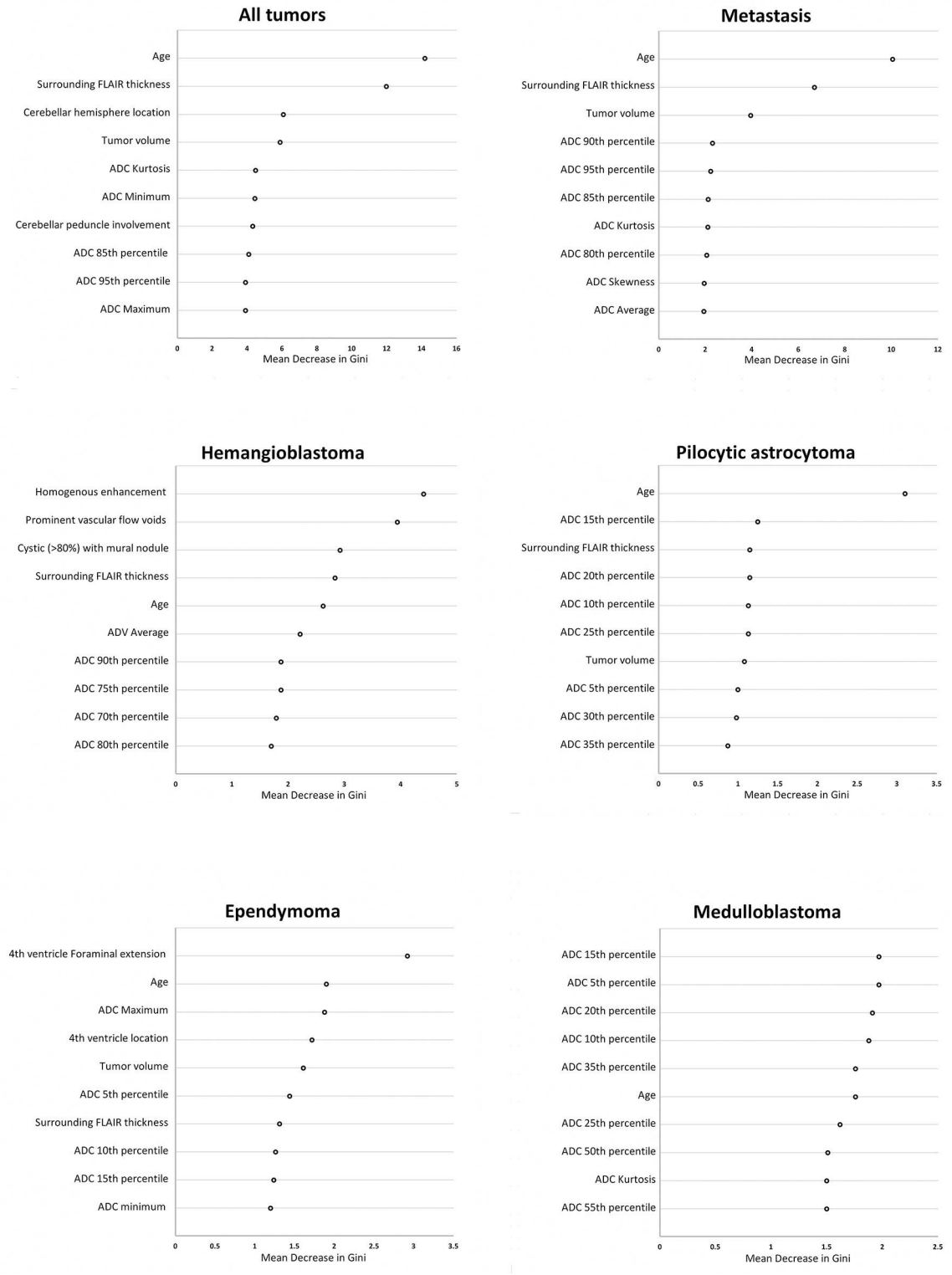


FIGURE 8 | The "mean decrease in Gini coefficient" in random forest models for differentiation of posterior cranial fossa tumors. Separate random forests models were developed for differentiation of all posterior fossa tumors from each other in multiclass analysis, as well as dichotomized classification for the 5 most common posterior fossa neoplasms. The top 10 variables with the highest averaged "mean decrease in Gini coefficient" among from ×100 repeats of 5-fold cross validation are reported. ADC, apparent diffusion coefficient; FLAIR, fluid attenuated inversion recovery.

types, or location (5, 8, 11, 12, 24). In terms of image analysis, prior research focused on 2D region of interest measurements, restricted use of ADC histogram metrics, and exclusion of qualitative features from analysis (6, 8, 11, 12). The strength of our analysis is defined by volumetric voxel-based ADC histogram analysis, use of comprehensive ADC histogram metrics, and incorporation of qualitative imaging analysis and patient's age (6, 8, 11, 12).

Almost all prior machine learning schemes for differentiation of posterior fossa tumors have limited their study to differentiation of ependymoma, medulloblastoma, and pilocytic astrocytoma (25). Rodriguez Gutierrez et al. have applied support vector machine classification of 17 medulloblastomas, 16 pilocytic astrocytomas, and 7 ependymomas (8). They reported that combination of the ADC histogram 25th percentile, 75th percentile, and skewness values could achieve the highest accuracy of 91% (8). Orphanidou-Vlachou et al. have applied principal component analysis for feature selection in combination with probabilistic neural network to classify 21 medulloblastomas, 14 pilocytic astrocytomas and 5 ependymomas patients based on T1- and T2-weighted image texture features (8, 25). In leave-one-out cross validation, they achieved 85.8% overall accuracy (25). Fetit et al. analyzed 21 medulloblastomas, 20 pilocytic astrocytomas and 7 ependymomas patients; and compared naïve Bayes, classification tree, k nearest neighbor, SVM, artificial neural network, and logistic regression classification models using three-dimensional texture data (26). In leave-one-out cross-validation, the SVM and artificial neural networks achieved the highest accuracy of 92% (26). In 2017, Zarinabad et al. reported the results of 1.5 Tesla 1H-MR spectroscopy for differentiation of 42 pilocytic astrocytoma, 38 medulloblastomas, and 10 ependymomas, comparing Naïve Bayes, SVM, artificial neural networks, and linear discriminative analysis (27). Using AdaBoost ensemble technique and synthetic minority oversampling technique (SMOTE), they could achieve an averaged balanced accuracy rate of 91% in oversampled-data based on metabolite concentration (27). In 2018, Zarinabad et al. reported the results of 3 Tesla 1H-MR spectroscopy for differentiation of 17 medulloblastomas, 20 pilocytic astrocytomas, and 4 ependymomas, and could achieve the highest Balanced Accuracy Rate of 86% using SVM classifiers (28). In our study, there was no exclusion based on the patients' age or the tumor histopathology, and we could achieve 0.873 averaged AUC among validation datasets for differentiation of 11 posterior fossa tumor types in the multiclass random forest analysis.

Of note, ADC histogram metrics were among variables with the greatest effects on accuracy of random forest models for differentiation of posterior fossa tumors (**Figure 8**). While prior studies have shown the value of ADC maps in differentiation of posterior fossa tumors (8–10), current results depict how combination of ADC histogram analysis and qualitative MR imaging features defined by neuroradiologists can help with diagnostic differentiation of these tumors. For example, among adult posterior fossa tumors, both metastases and hemangioblastomas present with prominent surrounding vasogenic edema (**Figure 3**). However, homogenous enhancement pattern, presence of prominent vascular flow voids, and higher ADC histogram percentile values favor the diagnosis of hemangioblastoma over metastasis (**Figures 5**, **6**).

While individual CART decision trees are prone to overfitting, random forest ensemble learning method theoretically reduces the potential overfitting. In addition, we opted to report the averaged results of machine learning models among 500 randomly selected training and validation cohorts to represent a realistic reflection of machine learning algorithm accuracy for prediction of tumor type, and compensate for potential overfitting. By doing so, however, we could not directly compare the performance of machine learning algorithm with clinical interpretation or neuroradiologist results. Nevertheless, the results of multiclass ROC analysis as well as the ROC AUC of reviewers in **Table 3** and averaged ROC AUC of machine learning models can provide an indirect comparison between neuroradiologist interpretation and machine learning models.

One of the strengths of our study is the use of a large cohort of patients presenting with posterior fossa tumor, which is representative of a patient population in a tertiary care center. A large and diverse cohort allowed us to have an appropriate training set for development of an accurate machine learning based model that differentiates a wide variety of posterior fossa tumors encountered in a tertiary care center practice. The natural next step of current study is training of machine learning models for prediction of molecular subtypes in specific posterior tumors. Future studies and prospective validation of decision tree models can also determine the impact of proposed machine learning algorithms on pretreatment diagnosis and therapy strategies in patients with posterior fossa tumors. Our results, however, provide the first step in devising a "no priori" and "data driven" decision models for differentiation of posterior fossa tumors, and are a new guide for methodological design of future machine learning classifiers. In addition, combination of clinical, and genetic biomarkers with imaging features can provide multivariate wholistic models for accurate prognostication and targeted therapy plan.

The major limitations of current study are the small number of rare tumor types; and the lack of molecular subtyping in medulloblastomas and ependymomas, which affect neoplasm prognosis and treatment planning (29). The study is also inherently limited in devising statistically powerful diagnostic models for less frequent posterior fossa tumors. Moreover, we only included subjects with known posterior fossa tumor; whereas, the machine learning model should preferably differentiate non-neoplastic tumor-mimics from tumors. However, designing the selection criteria for inclusion of potential tumor-mimic lesions for training machine learning models can be challenging due to lack of consensus on which lesions are qualified as tumor-mimic. Manual segmentation of brain tumors and measurement of peritumoral FLAIR hyperintensity can be challenging and a source of variability, particularly in non-enhancing T2 hyperintense glial tumors. Acquisition of MRIs in two different field strengths and on various scanners may also introduce heterogeneity in our data, although ADC obtained with repetition time >3,000 ms and b

value of 1,000 s/mm^2 are not substantially affected by scanner magnet strength (30). Additionally, there was no homogenous standardized imaging performed. Finally, the difference in imaging protocols, heterogeneity of patients' population, and age group can limit generalizability of our models.

CONCLUSION

We developed objective and quantitative decision tree models for differentiation of posterior fossa tumors based on ADC histogram metrics, patients' age, and qualitative MR imaging features that can easily be extracted on common image viewer platforms by radiologists. In addition, we have compared different machine learning classifiers for prediction of the most common posterior fossa tumors, and found that random forest models achieved greater accuracy in tumor differentiation. However, the results of our study need to be used with caution; and the proposed differentiation model should be validated in a larger prospective cohort before being used for clinical decision making. Pending prospective validation, such quantitative and objective diagnostic tools can potentially guide surgical planning or treatment decision for presurgical neoadjuvant therapy.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board at UCSF. Written consent requirement was waved given the retrospective review nature of the study.

AUTHOR CONTRIBUTIONS

SP and SC: guarantors of integrity of entire study, concepts and design, and data acquisition or data analysis/interpretation. All authors: manuscript drafting and final version approval.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00071/full#supplementary-material

REFERENCES

1. Chen C, Ou X, Wang J, Guo W, Ma X. Radiomics-based machine learning in differentiation between glioblastoma and metastatic brain tumors. *Front Oncol*. (2019) 9:806. doi: 10.3389/fonc.2019.00806
2. Tian Z, Chen C, Fan Y, Ou X, Wang J, Ma X, et al. Glioblastoma and anaplastic astrocytoma: differentiation using MRI texture analysis. *Front Oncol*. (2019) 9:876. doi: 10.3389/fonc.2019.00876
3. Wang Q, Zhang S, Cheng J, Liu W, Hui X. Radiologic features and surgical strategy of hemangioblastomas with enhanced cyst wall. *World Neurosurg*. (2017) 108:143–50. doi: 10.1016/j.wneu.2017.08.155
4. Lonser RR, Vortmeyer AO, Butman JA, Glasker S, Finn MA, Ammerman JM, et al. Edema is a precursor to central nervous system peritumoral cyst formation. *Ann Neurol*. (2005) 58:392–9. doi: 10.1002/ana.20584
5. Bull JG, Saunders DE, Clark CA. Discrimination of paediatric brain tumours using apparent diffusion coefficient histograms. *Eur Radiol*. (2012) 22:447–57. doi: 10.1007/s00330-011-2255-7
6. Poussaint TY, Vajapeyam S, Ricci KI, Panigrahy A, Kocak M, Kun LE, et al. Apparent diffusion coefficient histogram metrics correlate with survival in diffuse intrinsic pontine glioma: a report from the Pediatric Brain Tumor Consortium. *Neuro Oncol*. (2016) 18:725–34. doi: 10.1093/neuonc/nov256
7. Rumboldt Z, Camacho DL, Lake D, Welsh CT, Castillo M. Apparent diffusion coefficients for differentiation of cerebellar tumors in children. *AJNR Am J Neuroradiol*. (2006) 27:1362–9.
8. Rodriguez Gutierrez D, Awwad A, Meijer L, Manita M, Jaspan T, Dineen RA, et al. Metrics and textural features of MRI diffusion to improve classification of pediatric posterior fossa tumors. *AJNR Am J Neuroradiol*. (2014) 35:1009–15. doi: 10.3174/ajnr.A3784
9. Payabvash S, Tihan T, Cha S. Differentiation of cerebellar hemisphere tumors: combining apparent diffusion coefficient histogram analysis and structural MRI features. *J Neuroimag*. (2018) 28:656–65. doi: 10.1111/jon.12550
10. Payabvash S, Tihan T, Cha S. Volumetric voxelwise apparent diffusion coefficient histogram analysis for differentiation of the fourth ventricular tumors. *Neuroradiol J*. (2018) 31:554–64. doi: 10.1177/1971400918800803
11. Pierce T, Kranz PG, Roth C, Leong D, Wei P, Provenzale JM. Use of apparent diffusion coefficient values for diagnosis of pediatric posterior fossa tumors. *Neuroradiol J*. (2014) 27:233–44. doi: 10.15274/NRJ-2014-10027
12. Pierce TT, Provenzale JM. Evaluation of apparent diffusion coefficient thresholds for diagnosis of medulloblastoma using diffusion-weighted imaging. *Neuroradiol J*. (2014) 27:63–74. doi: 10.15274/NRJ-2014-10007
13. Arai K, Sato N, Aoki J, Yagi A, Taketomi-Takahashi A, Morita H, et al. MR signal of the solid portion of pilocytic astrocytoma on T2-weighted images: is it useful for differentiation from medulloblastoma? *Neuroradiology*. (2006) 48:233–7. doi: 10.1007/s00234-006-0048-5
14. Marshall RJ. The use of classification and regression trees in clinical epidemiology. *J Clin Epidemiol*. (2001) 54:603–9. doi: 10.1016/S0895-4356(00)00344-9
15. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324
16. Lahmiri S, Boukadoum M. Hybrid discrete wavelet transform and gabor filter banks processing for features extraction from biomedical images. *J Med Eng*. (2013) 2013:104684. doi: 10.1155/2013/104684
17. Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *Neuroimage*. (2013) 65:167–75. doi: 10.1016/j.neuroimage.2012.09.065
18. Nayak DR, Dash R, Majhi B. Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests. *Neurocomputing*. (2016) 177:188–97. doi: 10.1016/j.neucom.2015.11.034
19. Mitra J, Bourgeat P, Fripp J, Ghose S, Rose S, Salvado O, et al. Lesion segmentation from multimodal MRI using random forest following ischemic stroke. *Neuroimage*. (2014) 98:324–35. doi: 10.1016/j.neuroimage.2014.04.056
20. Emblem KE, Pinho MC, Zöllner FG, Due-Tonnessen P, Hald JK, Schad LR, et al. A generic support vector machine model for preoperative glioma survival associations. *Radiology*. (2015) 275:228–34. doi: 10.1148/radiol.14140770
21. Tian Q, Yan LF, Zhang X, Zhang X, Hu YC, Han Y, et al. Radiomics strategy for glioma grading using texture features from multiparametric MRI. *J Magn Reson Imaging*. (2018) 48:1518–28. doi: 10.1002/jmri.26010
22. Zhang X, Yan LF, Hu YC, Li G, Yang Y, Han Y, et al. Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features. *Oncotarget*. (2017) 8:47816–30. doi: 10.18632/oncotarget.18001
23. Park YW, Han K, Ahn SS, Bae S, Choi YS, Chang JH, et al. Prediction of IDH1-mutation and 1p/19q-codeletion status using preoperative MR imaging

phenotypes in lower grade gliomas. *AJNR Am J Neuroradiol*. (2018) 39:37–42. doi: 10.3174/ajnr.A5421
24. Cha J, Kim ST, Nam DH, Kong DS, Kim HJ, Kim YK, et al. Differentiation of hemangioblastoma from metastatic brain tumor using dynamic contrast-enhanced MR imaging. *Clin Neuroradiol*. (2017) 27:329–34. doi: 10.1007/s00062-016-0508-1
25. Orphanidou-Vlachou E, Vlachos N, Davies NP, Arvanitis TN, Grundy RG, Peet AC. Texture analysis of T1 - and T2 -weighted MR images and use of probabilistic neural network to discriminate posterior fossa tumours in children. *NMR Biomed.* (2014) 27:632–9. doi: 10.1002/nbm.3099
26. Fetit AE, Novak J, Peet AC, Arvanitits TN. Three-dimensional textural features of conventional MRI improve diagnostic classification of childhood brain tumours. *NMR Biomed*. (2015) 28:1174–84. doi: 10.1002/nbm.3353
27. Zarinabad N, Wilson M, Gill SK, Manias KA, Davies NP, Peet AC. Multiclass imbalance learning: improving classification of pediatric brain tumors from magnetic resonance spectroscopy. *Magn Reson Med*. (2017) 77:2114–24. doi: 10.1002/mrm.26318
28. Zarinabad N, Abernethy LJ, Avula S, Davies NP, Rodriguez Gutierrez D, Jaspan T, et al. Application of pattern recognition techniques for classification of pediatric brain tumors by *in vivo* 3T (1) H-MR spectroscopy-A multi-center study. *Magn Reson Med*. (2018) 79:2359–66. doi: 10.1002/mrm.26837
29. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 world health organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. (2016) 131:803–20. doi: 10.1007/s00401-016-1545-1
30. Ogura A, Tamura T, Ozaki M, Doi T, Fujimoto K, Miyati T, et al. Apparent diffusion coefficient value is not dependent on magnetic resonance systems and field strength under fixed imaging parameters in brain. *J Comput Assist Tomogr*. (2015) 39:760–5. doi: 10.1097/RCT.0000000000000266

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 25 February 2020
doi: 10.3389/fonc.2020.00198

Performance of Multiparametric Functional Imaging and Texture Analysis in Predicting Synchronous Metastatic Disease in Pancreatic Ductal Adenocarcinoma Patients by Hybrid PET/MR: Initial Experience

Jing Gao†, Xinyun Huang†, Hongping Meng, Miao Zhang, Xiaozhe Zhang, Xiaozhu Lin and Biao Li*

Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

OPEN ACCESS

Edited by:
Xuelei Ma,
Sichuan University, China

Reviewed by:
Laurence Gluch,
The Strathfield Breast Centre, Australia
Athanasios G. Zafeirakis,
Army Share Fund Hospital (NIMTS), Greece

****Correspondence:***
Xiaozhu Lin
lxz11357@rjh.com.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *16 September 2019*
Accepted: *05 February 2020*
Published: *25 February 2020*

Citation:
Gao J, Huang X, Meng H, Zhang M, Zhang X, Lin X and Li B (2020) Performance of Multiparametric Functional Imaging and Texture Analysis in Predicting Synchronous Metastatic Disease in Pancreatic Ductal Adenocarcinoma Patients by Hybrid PET/MR: Initial Experience. Front. Oncol. 10:198. doi: 10.3389/fonc.2020.00198

Objectives: To assess the imaging biomarkers of glucose metabolic activity and diffusion-weighted imaging (DWI) derived from pretreatment integrated ^{18}F-fluorodeoxyglucose positron emission tomography-magnetic resonance (^{18}F-FDG PET/MR) imaging as potential predictive factors of metastasis in patients with pancreatic ductal adenocarcinoma (PDAC).

Patients and Methods: We retrospectively included 17 consecutive patients with pathologically confirmed PDAC by pretreatment ^{18}F-FDG PET/MR. The study subjects were divided into a non-metastatic group (M0, six cases) and a metastatic group (M1, 11 cases). The ^{18}F-FDG PET/MR images were reviewed independently by two board certificated nuclear medicine physicians and one radiologist. Conventional characteristics and quantitative parameters from both PET and apparent diffusion coefficient (ADC) were assessed. The texture features were extracted from LIFEx packages (www.lifexsoft.org), and a 3D tumor volume of interest was manually drawn on fused PET/ADC images. Chi-square tests, independent-samples *t*-tests and Mann–Whitney *U*-tests were used to compare the differences in single parameters between the two groups. A logistic regression analysis was performed to determine independent predictors. A receiver operating characteristic (ROC) curve analysis was performed to assess the discriminatory power of the selected parameters. Correlations between metabolic parameters and ADC features were calculated with Spearman's rank correlation coefficient test.

Results: For conventional parameters, univariable analysis demonstrated that the M1 group had a significantly larger size and a higher peak of standardized uptake value (SUV_{peak}), metabolic tumor volume (MTV), and total lesion glycolysis (TLG) than those of the M0 group ($p < 0.05$ for all). TLG remained significant predictor in the multivariable analysis, but there were no significant differences for the area under the ROC curve (AUC) among the four conventional features in differential diagnoses ($p > 0.05$ for all). For the texture features, there were four features from the PET image and 13 from the ADC map that showed significant differences between the two groups. Multivariate analysis indicated that one feature from PET and three from the ADC

were significant predictors. TLG was associated with ADC-GLRLM_GLNU ($r = 0.659$), ADC-GLRLM_LRHGE ($r = 0.762$), and PET-GLRLM_LRHGE ($r = 0.806$).

Conclusions: Multiple parameters and texture features of primary tumors from ^{18}F-FDG PET/MR images maybe reliable biomarkers to predict synchronous metastatic disease for the pretreatment PDAC.

Keywords: pancreas—adenocarcinoma, metastasis, PET/MR hybrid imaging, multiparametric, texture analysis

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) has poor prognosis, and ranks the fourth among cancer-related death. It often presents at a late stage, and exhibits a 5-year overall survival rate of <8% (1). Distant metastasis is still frequently encountered in the operation of patients with potentially resectable PDAC (2, 3). Currently surgical resection is the only curative treatment for PADC. But it is very challenging to identify occult metastatic disease (OMD) by conventional images in the patients with resectable tumor before surgery, which makes further development of preoperative imaging essential. The accurate diagnosis of pancreatic cancer is important for determining the optimal management strategy. The predicting of patients with poor prognosis in advance would help in initial management, including the use of neoadjuvant chemotherapy or radiation, or adopting adjuvant therapy after surgery. Although OMD in PDAC is common, the mechanism and risk factors of its development are largely unknown.

Positron emission tomography/magnetic resonance (PET/MR) imaging is a newly developed technology that combines the anatomical and functional characteristics of MR imaging with the metabolic information of PET in one-stop examination. Hybrid PET/MR has been introduced into the clinical application setting since 2011. Studies on the feasibility and potential applications of PET/MR imaging have been reported soon after that, and oncology was one of the hot topic (4–8). Because multiparametric PET/MR imaging can provide many biomarkers of the studied diseases non-invasively, it was widely used in oncological research, especially for the tumor diagnosis, treatment planning, surveillance, and follow-up. Compared with PET/CT plus contrast-enhanced multidetector CT (MDCT), ^{18}F-FDG PET/MR imaging obtained a similar diagnostic performance in the preoperative staging and resectability assessment of pancreatic neoplasms (9).

MRI with diffusion weighted imaging (DWI) has incremental value in detecting small hepatic metastasis and peritoneal implants when combined with FDG PET imaging, which can avoid unnecessary surgery (10, 11). With the integration of the advantages of PET and MR imaging, PET/MR imaging bears great potential in detecting and diagnosing of metastatic disease in PDAC patients.

By extracting and analyzing a large number of putative imaging features, which may reflect the heterogeneity of tissues, texture analysis and radiomics played an increasingly important role in cancer research (12). The rationale is that image texture features and radiomics characteristics may contain information of tumor phenotypes, which can reflect patient prognosis indirectly. Texture analysis and radiomics using CT images, which are widely available, has been used to predict aggressiveness, disease-free survival (DFS), and overall survival (OS) in patients with PDAC (13–15). DWI can reflect the tissue cellularity, and has been used in texture analysis in many other studies (16–19). Quantitative parameters obtained from current-generation hybrid imaging can provide complementary information of morphology and function simultaneously, which might be related to tumor biological behavior (16, 20). In the present study, we first explored the value of three-dimensional texture analysis based on hybrid ^{18}F-FDG PET/ADC images in predicting of metastatic disease in PDAC patients.

Our hypothesis is that different kind of imaging parameters and features from pretreatment multiparametric PET/MR can be used to predict synchronous distant metastasis in patients with PDAC. In addition, the automated analysis of quantitative imaging features may complement conventional imaging metrics for prognostic evaluation. The purpose of this study was to assess conventional PET/MR findings and tumor texture features on pretreatment PET/MR imaging as potential predictive factors of metastasis for PDAC.

MATERIALS AND METHODS

Subjects

This retrospective study was approved by the Institutional Ethics Committee of Ruijin Hospital, and informed consent was obtained from the patients who participated in another clinical study (application of abdominal PET/MR sequentially after whole body ^{18}F-FDG PET/CT). No written informed consent was required for the other patients who underwent whole body PET/MR according to clinical indications. From March 2018 to January 2020, 29 consecutive patients (mean age, 60.8 ± 10.1 years; men/women, 12/17) with suspected pancreatic cancer underwent hybrid multiparametric ^{18}F-FDG PET/MR with DWI before treatment. The patients were considered eligible based on the following criteria: (1) histopathological examination via either biopsy or surgical procedure; (2) hybrid ^{18}F-FDG PET/MR scans (with DWI) performed before biopsy and surgical intervention; and (3) no local or systemic treatments to pancreatic cancer. Of the 29 patients, 12 patients without a pathological-confirmed diagnosis were excluded. Finally, 17 patients (mean age, 57.4 ± 10.1 years; range, 40–75 years; eight men, nine women) with PDAC were included in our study population. All patients tolerated this examination. Tumor size was measured according to MRI images, and the maximum

diameter was recorded. Synchronous distant metastases were confirmed with imaging techniques and, if possible, by either surgical operation or biopsy. The study subjects were divided into two groups [without synchronous distant metastasis (M0 group) and with synchronous distant metastasis (M1 group)]. The patient characteristics are summarized in **Table 1**.

PET/MR Protocol

Whole-body PET/MR was performed using an integrated PET/MR system (Biograph mMR; Siemens Healthineers, Erlangen, Germany). All participants were fasted for at least 6 h before the study and given intravenous ^{18}F-FDG 2.5 to 6 MBq/kg at 40–100 min before each PET/MR study. For whole body examination, PET was performed from the mid-thighs to the skull base in four bed positions (acquisition time, 4 min/position) with the patient in a supine arm-down position, and head was scanned with 1 bed position for 8 min. Simultaneous MRI with axial T2-weighted 2D half-Fourier acquisition single-shot turbo spin-echo sequences(HASTE), axial DWI with echo planar sequence(b-values, 50 and 800 s/mm^2), and axial T1-weighted imaging (T1WI) with a DIXON sequence were performed and PET data were acquired at each bed position. For abdominal examination, the simultaneous acquisition of PET and MRI data was performed. Unenhanced studies, including coronal T2WI half-Fourier acquisition single-shot fast spin-echo, axial and coronal T2WI with fat saturation, axial T1-weighted fat-suppressed three-dimensional gradient-recalled echo imaging were performed. DWI was performed by using a single-shot echo-planar imaging sequence with *b* values of 50 and 800 sec/mm^2. The ADC map was calculated using a monoexponential function (b-values, 50 and 800 s/mm^2; **Supplementary Table 1**).

The PET images were reconstructed with an ordered-subset, expectation-maximization, iterative algorithm (4 iterations, 21 subsets), with a 4-mm post reconstruction Gaussian filter and a matrix of 172 * 172. Attenuation correction of PET data was obtained by a 4-tissue-class (air, lung, fat, soft tissue) segmented attenuation map from a 2-point Dixon MR pulse sequence. Eight patients were subjected to abdominal PET/MR (after whole body PET/CT), one patient was subjected to whole body PET/MR, and eight patients were subjected to whole body plus abdominal PET/MR.

Image Analysis

The focal ^{18}F-FDG uptake at the primary tumor, the lymph nodes and distant metastases were reviewed independently by two board certificated nuclear medicine physicians (12 and 4 years of experience) on PET/MR images. A radiologist who specialized in abdominal MRI with 13 years of experience and 2 years of experience in nuclear medicine read the PET/MR studies. The nuclear medicine physicians and radiologists independently performed their analyses on the workstation. Any disagreement was resolved by discussion. The volume of interest (VOI) was manually drawn on the PET image, and a region of interest (ROI) was drawn manually on ADC maps with consensus by three readers, and the ADC values and PET parameters of the pancreatic tumor were measured.

The PET-related parameters included maximum standardized uptake value (SUV_{max}), mean SUV (SUV_{mean}), maximum average SUV within a 1 cm^3 spherical volume (SUV_{peak}), standard deviation of SUV(SUV_{sd}), MTV, and TLG. The SUV_{max} and SUV_{mean} were defined as the maximum and mean radioactivity concentration of images enclosed by the VOI divided by the whole body concentration of the injected

TABLE 1 | Basic characteristics of the study participants (17 cases).

Patient number	Gender	Age (years)	Height (cm)	Body weight (Kg)	Tumor location	Tumor size (cm)	Location of metastasis	Group
1	Female	62	164	55	Body/tail	4.7	Peritoneum	Metastatic
2	Female	40	154	40	Body/tail	5.1	Liver	Metastatic
3	Male	66	170	65	Head/neck	4.3		Non-metastatic
4	Male	61	172	57	head/neck	4.6		Non-metastatic
5	Male	66	170	57	Head/neck	3.9	Liver	Metastatic
6	Female	55	160	45	Body/tail	5.7	peritoneum	metastatic
7	Female	47	159	60	Head/neck	2.2		Non-metastatic
8	Male	49	173	64	Body/tail	4.6	Multiple*	Metastatic
9	Male	72	170	78	Head/neck	2.5		Non-metastatic
10	Female	75	160	55	Body/tail	4.5	Supraclavicular lymph node	Metastatic
11	Female	43	155	56	Head/neck	3.2		Non-metastatic
12	Male	57	170	66	Body/tail	6.0	Peritoneum	Metastatic
13	Female	51	164	47	Body/tail	4.7	Liver	Metastatic
14	Male	63	170	70	Body/tail	4.0	Liver	Metastatic
15	Female	65	163	60	Body/tail	2.7	Liver	Metastatic
16	Male	48	180	69	Head/neck	3.3		Non-metastatic
17	Female	56	160	47	Body/tail	3.2	Liver, peritoneum	Metastatic

Multiple, liver, left adrenal gland, remote lymph nodes, bones.*

radioactivity. SUV_{max}, SUV_{mean}, $SUV_{peak,}$ and MTV values were then measured automatically using commercial software (Syngo Via Workstation; Siemens Healthineers, Erlangen, Germany). The peak of the SUV (SUV_{peak}) was determined using a 1 cm^3 spherical volume of interest automatically centered on the tumor area with the maximum uptake. The MTV was determined by segmentation of the tumor based on a 40% threshold of SUV_{max}. TLG was calculated as SUV_{mean}* MTV.

To measure the ADC, ROIs were manually drawn on the ADC map along the contour of the tumor on a single slice containing the largest area of the tumor. The DWI parameters included the mean ADC (ADC_{mean}), standard deviation of ADC value (ADC_{sd}), and minimum ADC (ADC_{min}). The lowest ADC value in an ROI, ADC_{min}, represented the greatest tumor cellularity.

Among all 17 patients enrolled, the following imaging biomarkers were recorded for the primary tumor: SUV_{mean}, SUV_{max}, SUV_{peak}, SUV_{sd}, MTV, TLG, ADC_{mean}, ADC_{min}, ADC_{sd}, and tumor size (maximum diameter of the tumor from MRI). A total of 10 PET/MR parameters were applied for differentiation.

TNM staging system of American Joint Committee on Cancer (8th edition) was applied for the study patients by a multidisciplinary team for pancreatic cancer at our hospital. Among those who did not receive curative surgery, the stage was determined by biopsy and all available image results.

Computerized Textual Analysis

Features of the primary tumor were extracted using the Local Image Features Extraction (LIFEx) package (http://www.lifexsoft.org). The texture analysis was performed inside the VOI retrieved from the fused PET/ADC images. The VOI was manually drawn with consensus by three nuclear medicine-certified physicians and radiologist together. Histogram-based features, the gray-level cooccurrence matrix (GLCM), the neighborhood gray-level different matrix (NGLDM), the gray level run length matrix (GLRLM) and the gray level zone length matrix (GLZLM) were obtained. There were 37 texture indices analyzed in this study (**Supplementary Table 2**). The ^{18}F-FDG uptake intensity data were rescaled using 64 discrete values to reduce the image noise.

Statistical Analysis

Summary statistics are presented as the mean ± SD for quantitative variables or frequency for qualitative variables. Appropriate statistical tests were used to assess differences in ^{18}F-FDG PET/MR imaging biomarkers between patients with and without synchronous metastatic disease. We first performed univariate analyses on a series of variables, followed by multivariate analyses on selected variables with significant differences in the univariate analysis. The patient gender and tumor location between two groups were compared using the Chi-square test with Fisher's exact test. The patient age, height, body weight, and tumor size between the two groups were compared using an independent-samples *t*-test. The ADC values, PET parameters, and textural parameters between the two groups were compared using the independent-samples Mann–Whitney *U*-test. Multivariable analysis was investigated using the stepwise forward logistic regression model with significant parameters. Receiver operating characteristic (ROC) analyses were performed to evaluate the diagnostic accuracy of predicting synchronous metastatic disease (M1 or M0), and the area under the ROC curve (AUC) was calculated to identify the optimal cut-off values for each parameter. The parameter was most likely to accurately identify a positive instance (with synchronous metastatic disease) when the AUC value was high ($p < 0.05$). The 95% confidence intervals (CI) for AUC and *p*-values for comparison of related ROC curves were obtained with the method described by DeLong and coworkers (21). The relationship between metabolic parameters and texture features from the ADC map was also evaluated using Spearman's rank correlation coefficient test. A $p < 0.05$ was considered statistically significant, and all *p*-values presented were two-sided. Data were analyzed using SPSS software (SPSS for Windows 23; IBM Corp., Armonk, USA) and MedCalc for Windows, version 11.4 (MedCalc Software, Ostend, Belgium).

RESULTS

Patient Characteristics

Six patients without synchronous metastatic disease (M0) and 11 patients with synchronous metastatic disease (M1) were included in this study. The average age was 56.2 ± 11.8 years (range from 43 to 72 years) in M0 patients and 58.1 ± 9.5 years (range from 40 to 75 years) in M1 patients. There were four males and two females in the M0 patient group and four males and seven females in the M1 patient group. The age, gender, height, and body weight did not differ significantly between the two groups ($p > 0.05$ for all). The characteristics of the patients are summarized in **Table 1**.

Conventional Parameters

Tumor location, tumor size, SUV_{peak}, MTV, and TLG differed significantly ($p < 0.05$ for all) between M0 and M1 patients. More tumors were located in the body/tail in the M1 group than in the M0 group ($p = 0.001$). The M1 group showed a larger tumor size than that in the M0 group ($p = 0.039$). Patients with synchronous metastatic disease demonstrated increased SUV_{peak}, MTV, and TLG in the primary tumor. SUV_{max}, $SUV_{mean,}$ and SUV_{sd} did not differ significantly between the two groups ($p > 0.05$ for all). ADC_{mean}, ADC_{min}, and ADC_{sd} did not differ significantly between the two groups ($p > 0.05$ for all). **Table 2** shows the conventional quantitative parameters of the two groups. Three of the 11 patients in M1 group had FDG-negative metastatic lesions. One patient had metastatic foci in the liver (**Figure 1**), and two patient had metastatic peritoneal lesions. One of the six patients in M0 group had FDG-negative primary tumors (**Tables 1**,**2**).

The conditional logistic regression model using significant parameters identified TLG as an independent predictor for synchronous metastatic disease diagnosis. The other parameters did not reach significance. Based on multivariate regression analysis, and we performed an ROC analysis for the selected parameters. The AUC was 0.848 for TLG.

TABLE 2 | The diagnostic performance of conventional quantitative ^{18}F-FDG PET/MR parameters for predicting synchronous distant metastasis in pancreatic ductal adenocarcinoma patients.

Parameter	Comparison of mean value			Receiver operating characteristic (ROC) analysis					
	M0 group	M1 group	*p*	AUC	95% CI	*p*	Optimal cutoff value	Se (%)	Sp (%)
Age (years)	56.2 ± 11.8	58.1 ± 9.5	0.719†						
Height (cm)	167.7 ± 9.1	164.4 ± 5.8	0.373†						
Body weight (Kg)	64.2 ± 8.4	55.1 ± 9.5	0.07†						
Tumor size (cm)	3.4 ± 1.0	4.5 ± 1.0	**0.039**†	0.803	0.543–0.952	0.006	>3.3	81.8	66.7
SUV_{mean}	2.6 ± 1.1	3.6 ± 1.3	0.149‡						
SUV_{max}	4.5 ± 2.0	6.3 ± 2.3	0.216‡						
SUV_{peak}	3.0 ± 1.1	4.8 ± 1.5	**0.037**‡	0.818	0.560–0.960	0.004	>4.06	72.7	100.0
SUV_{sd}	0.6 ± 0.3	0.8 ± 0.3	0.149‡						
MTV	8.4 ± 6.1	20.7 ± 13.1	**0.037**‡	0.818	0.560–0.960	0.003	>15.04	63.6	100.0
TLG	21.3 ± 16.7	67.7 ± 42.1	**0.020**‡	0.848	0.595–0.973	<0.001	>41.3	72.7	100.0
ADC_{mean} (s/mm^2)	1192 ± 625	1311 ± 219	0.884‡						
ADC_{min} (s/mm^2)	1093 ± 274	974 ± 334	0.733‡						
ADC_{sd} (s/mm^2)	128 ± 25	134 ± 20	0.525‡						

†Independent-samples t-test, bold value indicates p-value is significant <0.05; ‡Independent-samples Mann-Whitney U-test, bold value indicates p-value is significant <0.05. ADC, apparent diffusion coefficient; ADC_{mean}, mean apparent diffusion coefficient; ADC_{min}, minimum apparent diffusion coefficient; ADC_{sd}: standard deviation of apparent diffusion coefficient; AUC, area under receiver operating characteristic (ROC) curve; M0, no synchronous distant metastasis; M1, with synchronous distant metastasis; MTV, metabolic tumor volume; Se, sensitivity; Sp, specificity; SUV, standardized uptake values; SUV_{max}, maximum standardized uptake value; SUV_{mean}, mean standardized uptake value; SUV_{peak}, the peak of SUV in 1 ml; SUV_{sd}, standard deviation of standardized uptake value; TLG, total lesion glycolysis; 95% CI, 95% confidence interval.

When the optimal cut-off point was 41.3, the TLG showed a sensitivity of 72.7% and a specificity of 100.0% (**Table 2**). There were no significant differences in the AUC among tumor size, SUV_{peak}, MTV, and TLG ($p > 0.05$ for all; **Figure 2**).

Texture Features

Regarding the texture features, four features from the PET image (two GLRLM, one NGLDM and one GLZLM) and 13 features from the ADC map (two histogram based, seven GLRLM, and four GLZLM) showed significant differences between the two groups (**Table 3** and **Supplementary Table 3**). Conditional logistic regression analysis demonstrated that Long-Run High Gray-level Emphasis of Gray-Level Run Length Matrix (GLRLM_LRHGE) from PET image, Long-Run High Gray-level Emphasis (LRHGE), Gray-level Non-Uniformity for run (GLNU), and Run Length Non-Uniformity (RLNU) of Gray-Level Run Length Matrix (GLRLM) from the ADC map were significant independent predictors for predicting synchronous metastatic disease in PDAC. The metastatic group showed significantly higher PET-GLRLM_LRHGE, ADC-GLRLM_LRHGE, ADC-GLRLM_GLNU, and ADC-GLRLM_RLNU ($p < 0.05$ for all). The AUC was 0.939, 0.894, 0.924, and 0.909 for PET-GLRLM_LRHGE, ADC-GLRLM_LRHGE, ADC-GLRLM_GLNU, and ADC-GLRLM_RLNU, respectively. The logistic regression model with proposed features obtained an AUC of 1.000 (95% CI 0.805–1.000, $p < 0.001$), but there were no significant differences in the AUC for a single parameter vs. that for the logistic regression model ($p > 0.05$ for all, **Figure 3**).

Correlations Between PET/MR Parameters and Texture Features

The PET parameter of TLG showed positive correlations with the texture feature of ADC-GLRLM_GLNU ($r = 0.659$, $P = 0.004$), ADC-GLRLM_LRHGE ($r = 0.762$, $P < 0.001$), and PET-GLRLM_LRHGE ($r = 0.806$, $P < 0.001$).

DISCUSSION

In this study, we demonstrated differences in multiparametric ^{18}F-FDG PET/MR imaging biomarkers obtained from the primary tumor of PDAC between patients with and without synchronous metastasis. Then, we identified prognostic PET/MR imaging signatures in patients with PDAC by using conventional parameters and a texture analysis approach. We found that metastatic PDAC patients showed significantly larger tumor sizes, more frequent body/tail locations and higher SUV_{peak}, MTV, and TLG values in the primary tumor than those in non-metastatic patients ($p < 0.05$ for all). In addition, TLG remained significant predictor in the multivariable analysis. Regarding the texture features, we found that GLRLM_RLNU, GLRLM_GLNU, and GLRLM_ LRHGE from the ADC map, and GLRLM_LRHGE from PET image were also significant predictors of synchronous metastatic disease. In addition, TLG was associated with ADC-GLRLM_GLNU, ADC-GLRLM_LRHGE, and PET-GLRLM_LRHGE.

Regarding the tumor size and location, our results were consistent with previous studies (2, 3, 22, 23). The larger the tumor, the more likely it is to have distant metastasis. The cut-off value of tumor size was similar between our study

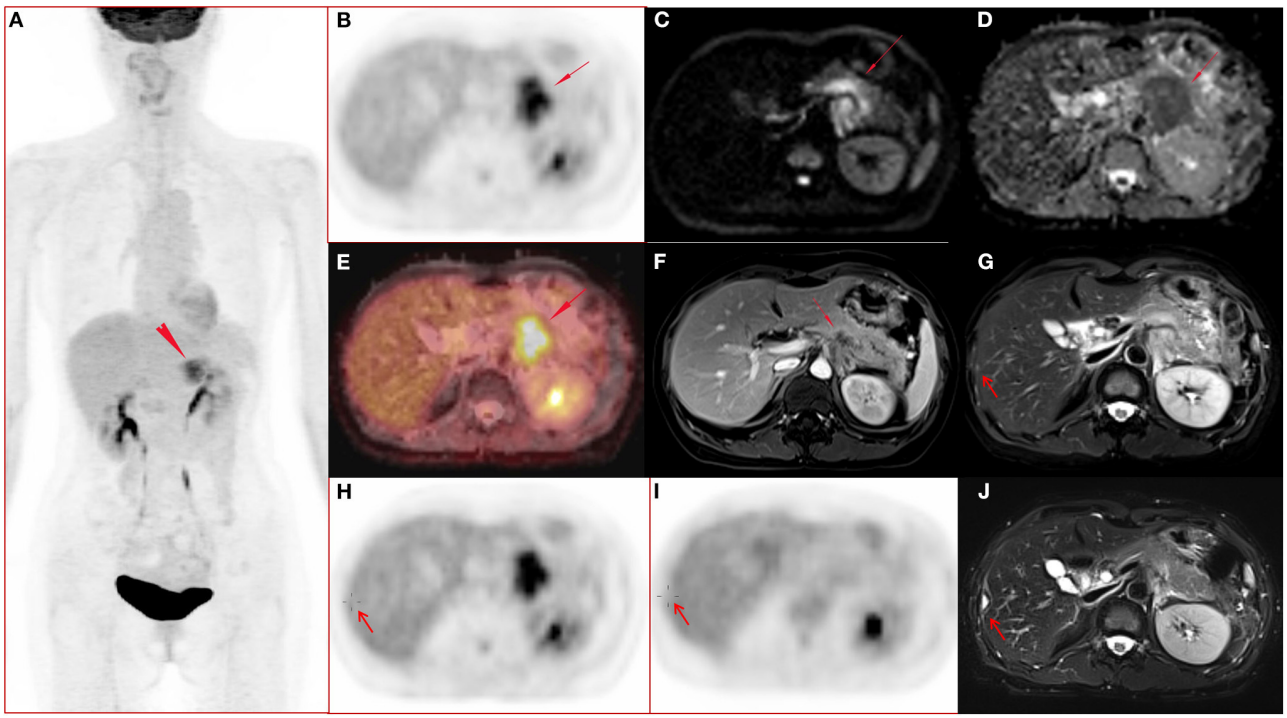


FIGURE 1 | A 51-year-old female with pancreatic ductal adenocarcinoma in body and tail with hepatic metastases. **(A–J)** Whole body PET image with maximum intensity projection (MIP) **(A)** and axial abdominal PET image **(B)** showed FDG metabolism increased lesion in pancreatic body and tail, with SUV_{max} 7.52, SUV_{peak} 6.39, and MTV 17.43 cm^3. Diffusion weighted imaging (DWI, b = 800) **(C)** and apparent diffusion coefficient (ADC) map **(D)** showed a diffusion restricted lesion in pancreatic body and tail. **(E)** Fused image of PET and ADC showed a diffusion restricted lesion with hyper FDG metabolism. **(F)** Contrast enhanced (CE) T1 weighted image (T1WI) with fat suppression (fs) on late arterial phase showed hypo-vascular lesion and dilated main pancreatic duct, and the maximum diameter of the lesion was 4.7 cm. **(G,H)** Metastasis in the right lobe of the liver (arrow) confirmed by surgery operation (2 days after the initial PET/MR examination) and histo-pathological examination, and the lesion showed slightly hyper-intensity on T2 weighted image with fat saturation **(G)**, no FDG avid lesion on PET image **(H)**. (**I–J)** Follow up PET/MR 112 days after operation showed the operated region with hyper-intensity in T2 weighted image with fat saturation **(J)** and without abnormal FDG uptake on PET image **(I)**.

(3.3 cm) and the studies of Liu et al. (4.0 cm) and Karabicak et al. (4.2 cm) (2, 22). In a cohort of 1,423 patients with PDAC who underwent pancreatectomies, the occurrence of occult metastatic disease in PDAC accounted for 8% of cases, and multivariable analysis defined four independent predictors for occult metastatic disease (3). Patients with abdominal pain, preoperative CA 19-9 > 192 U/ml, tumor bigger than 3 cm, and indeterminate lesions on preoperative CT had high risk of occult metastatic disease (3). The cut-off value of tumor size was slightly smaller in the study of Gemenetzis et al. (3) than that in our study, which might be because that the patients were potentially resectable with occult but not obvious metastasis and the sample size was large in that study. Another study of 110 patients with PDAC (22), patients with high CA 19-9 levels and large size tumor located in body-tail are at greater risk for latent distant organ metastasis or peritoneal metastasis. Tumors located in the body/tail of the pancreas are more likely to metastasize (22, 23), which was also confirmed in our study. The metabolic parameters of ^{18}F-FDG PET could reflect biological aggressiveness and predict prognosis in various studies (24–29), and we demonstrated similar results in this study. A study of 93 patients with pathologic T3 (pT3) resectable pancreatic cancer showed that tumor with high MTV2.5 is associated with both lymph node metastasis and early systemic metastasis (24). Patients who developed metastatic disease during follow-up after chemoradiotherapy had higher SUV_{max} (3.8 vs. 8.6), SUV_{peak} (2.5 vs. 7.5), SUV_{mean} (1.8 vs. 3.3), SUV_{median} (1.7 vs. 3.0), and TLG (26.9 vs. 115.9) than did those without metastatic disease (25). The average SUV_{peak} was 3.0 and 4.8 for M0 and M1 group in our study. The SUV_{peak} of PDAC without metastasis was similar between the two studies, and the SUV_{peak} of metastatic PDAC was slightly higher in the study of Wilson et al. (25) than that of the present study. Other recent studies (26–29) which made use of the PET/CT technique, unlike PET/MR, as was the case in our study, have addressed PET-derived parameters (TLG, MTV, or SUV_{peak}) as independent predictors for OS and PFS outcome in patients with pancreatic adenocarcinoma. A PET/CT scoring system with combination of quantitative parameters helps to improve the prognostication significantly (28).

According to our knowledge only two studies about overall survival(OS), prognosis, and imaging biomarkers of PDAC and periampullary cancer have been published using integrated PET/MR imaging (30, 31). In a study with 60

PET/MRI of pancreatic and periampullary cancer patients, the imaging biomarkers (ADC_{min}, Choline levels, TLG, MTV, MTV/ADC_{min} ratio) may predict clinical stage and progression-free survival (PFS) of the patients (30). Recently, Chen et al. have showed that multiparametric PET/MR imaging biomarkers of pancreatic cancer patients were associated with their OS (31). The application of PET/MR has just started, and more research is needed to find out the potential value of PET/MR. And more sophisticated methods are needed to improve the existing diagnostic capabilities. Radiomics in nuclear medicine is fastly developing. The advantage of radiomics should be fully explored from now on to improve the clinical value of multiparametric imaging, such as PET/CT and PET/MR, in predicting disease phenotypes and personalized diagnosis and treatment. In this study, texture analysis showed significant differences between M0 and M1 PDAC for two first-level (histogram skewness and kurtosis from ADC map) and for 15 third-level features(four from PET and 11 from ADC map). ADC-HISTO_Skewness and ADC-HISTO_Kurtosis were the first-level features with significant differences between the two groups based on the ROC analysis. According to the literature, ADC histogram analysis has the potential to provide valuable information on tumor biology and to predict tumor behavior in several malignancies (17, 18, 32, 33). The skewness and kurtosis were higher in cervical cancer patients with metastatic lymph nodes than those with negative nodal status (33). Another study showed that skewness and kurtosis of histogram analysis from ADC map were able to differentiate thyroid carcinoma with lymph node metastasis from that without metastasis (32). In the study of non-small cell lung cancer, higher ADC skewness and kurtosis were associated with lymphovascular invasion and pleural invasion (34). In a study of pediatric diffuse intrinsic potine glioma using ^{18}F-FDG PET and MRI ADC histogram, higher ADC skewness and kurtosis of the enhancing

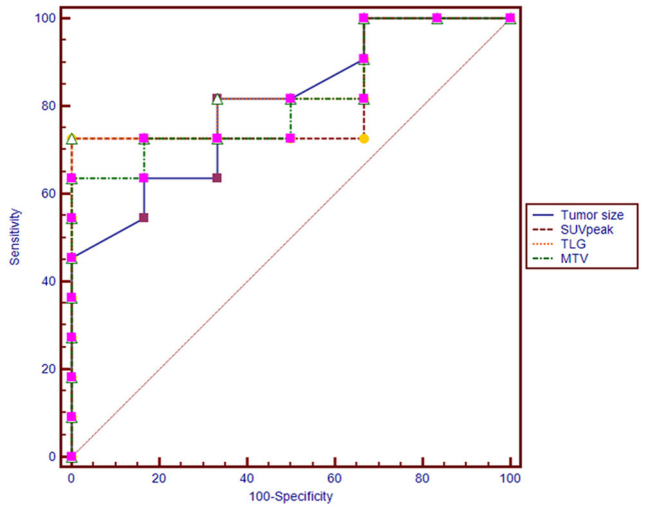


FIGURE 2 | Receiver operating characteristic (ROC) curves of tumor size, SUV_{peak}, MTV, and TLG for diagnosing synchronous metastatic disease in pancreatic ductal adenocarcinoma.

TABLE 3 | The diagnostic performance of texture features derived from simutanous ^{18}F-FDG PET image and the ADC map for predicting synchronous distant metastasis in pancreatic ductal adenocarcinoma patients.

Texture feature	M0 group	M1 group	P1	AUC	95% confidence intervals	P2	Optimal cut-off value
PET-GLRLM_RLNU	467 ± 173	1391 ± 738	0.002	0.939	0.711–0.998	<0.0001	>751
PET-GLRLM_LRHGE	172 ± 77	288 ± 96	0.037	0.818	0.560–0.960	0.005	>120.6
PET-NGLDM_Coarseness	0.015 ± 0.008	0.007 ± 0.003	0.037	0.803	0.543–0.952	0.034	<=0.01
PET-GLZLM_GLNU	5.4 ± 2.3	11.3 ± 6.8	0.020	0.848	0.595–0.973	0.0004	>8.3
ADC-HISTO_Skewness	−0.02 ± 0.59	0.72 ± 0.80	0.048	0.795	0.534–0.948	0.001	>0.14
ADC-HISTO_Kurtosis	3.32 ± 1.16	5.25 ± 2.01	0.048	0.803	0.543–0.952	0.008	>3.84
ADC-GLRLM_LRE	36 ± 10	66 ± 24	0.007	0.894	0.650–0.989	<0.0001	>42.98
ADC-GLRLM_SRHGE	786 ± 51	654 ± 131	0.048	0.803	0.543–0.952	0.006	<=691
ADC-GLRLM_LRLGE	0.008 ± 0.002	0.019 ± 0.009	0.005	0.879	0.631–0.984	<0.0001	>0.012
ADC-GLRLM_LRHGE	0.15E+6 ± 0.04E+6	0.28E+6 ± 0.10E+6	0.007	0.894	0.650–0.989	<0.0001	>0.18E+6
ADC-GLRLM_GLNU	227 ± 96	442 ± 142	0.003	0.924	0.690–0.996	<0.0001	>269.4
ADC-GLRLM_RLNU	30 ± 8	45 ± 9	0.005	0.909	0.670–0.993	<0.0001	>32.6
ADC-GLRLM_RP	0.212 ± 0.023	0.165 ± 0.033	0.007	0.879	0.631–0.984	<0.0001	<0.18
ADC-GLZLM_LZE	1.53E+6 ± 1.67E+6	9.50E+6 ± 8.71E+6	0.010	0.879	0.631–0.984	<0.0001	>4.62E+6
ADC-GLZLM_LZLGE	362 ± 395	2249 ± 2061	0.010	0.879	0.631–0.984	<0.0001	>1094.1
ADC-GLZLM_LZHGE	0.65E+10 ± 0.70E+10	4.01E+10 ± 3.68E+10	0.010	0.879	0.631–0.984	<0.0001	>1.95E+10
ADC-GLZLM_ZP	0.0016 ± 0.0017	0.0006 ± 0.0004	0.050	0.795	0.534–0.948	0.010	<=0

ADC, apparent diffusion coefficient; AUC, area under receiver operating characteristic (ROC) curve; FDG, fluorodeoxyglucose; M0, no synchronous distant metastasis; M1, with synchronous distant metastasis; NA, not applicable; P1, P-value for independent- samples Mann-Whitney U-test, indicates p-value is significant <0.05; P2, p-value for AUC, indicates P-value is significant <0.05; PET, positron emission tomography; 95% CI, 95% confidence interval; GLRLM, Gray level run length matrix; GLZLM, Gray level zone length matrix; SRHGE, Short-nun high gray-level emphasis; GLNU, Gray-level non-uniformity; RLNU, Run length non-uniformity; LZE, Long-zone emphasis; NGLDM, Neighborhood gray-level different matrix; SZHGE, Short-zone high gray-level emphasis; LZHGE, Long-zone high gray-level emphasis.

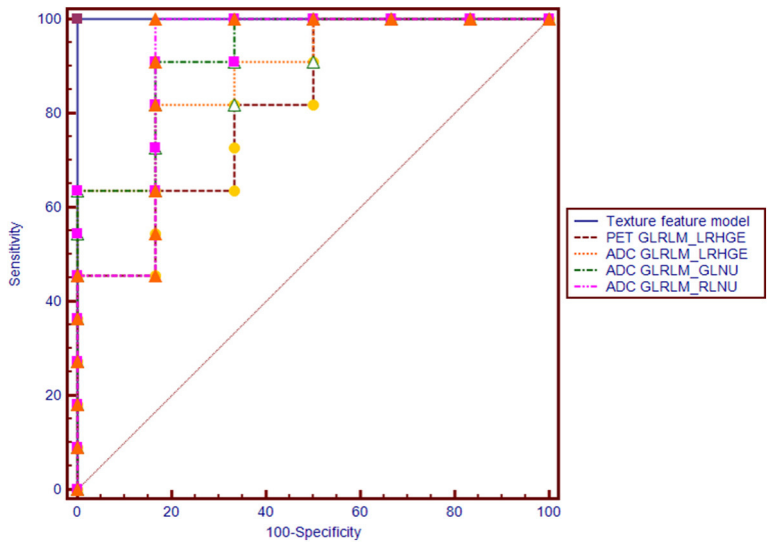


FIGURE 3 | Receiver operating characteristic (ROC) curves of PET-GLRLM_LRHGE, ADC-GLRLM_LRHGE, ADC-GLRLM_GLNU, ADC-GLRLM_RLNU, and logistic regression model with combination of four texture features for diagnosing synchronous metastatic disease in pancreatic ductal adenocarcinoma.

portion of the tumor were associated with shorter PFS (16). HISTO_Skewness is the asymmetry of the gray-level distribution in the histogram. If the peak of the frequency distribution shifts to the left, the long tail extends to the right, which is called a positive skewed distribution. Kurtosis reflects the sharpness of the histogram peak. So in this study, in M1 patients most voxels containing an ADC less than the mean. The lower ADC value indicates the higher cellularity and aggressiveness. Unexpectedly, the conventional ADC values (ADC_{mean}, ADC_{min}, ADC_{sd}) had no significant differences between metastatic and non-metastatic PDAC. Considering that the sample size of this pilot study is too small, it would be hasty to draw any conclusions from this negative finding.

The first-level texture feature describes the characteristics related to the voxel intensity distribution, while the meaning of second- and third-level features is non-figurative. In this study, the texture features of GLRLM_RLNU, GLRLM_GLNU, and GLRLM_LRHGE from the ADC map, and GLRLM_LRHGE from PET image were independent predictors of synchronous metastatic disease. GLRLM reflects the comprehensive information of the image grayscale with respect to direction, adjacent interval, and variation amplitude. GLRLM is a set of statistical feature extracted from medical images and applied in radiomics frequently (35–37). HGRE measures the distribution of sections of high intensity, and its value is expected to be large if the number of sections of high intensity is high. In a study of breast cancer using ^{18}F-FDG PET texture analysis, triple negative breast cancer (TNBC) had higher RHGE value than non-TNBC and exhibited more tumor heterogeneity (38). Gray-level Non-Uniformity for run (GLNU) measures similarity of values of gray-level and Run Length Non-Uniformity (RLNU) measures similarity of run length. In this study, the PET-GLRLM_LRHGE, ADC-GLRLM_LRHGE, ADC-GLRLM_GLNU, and ADC-GLRLM_RLNU values were higher in M1 patients than in M0 patients. This may indicate that the higher heterogeneity in PET image and ADC map of PDAC, the higher risk of metastasize. ADC-GLRLM_LRHGE and ADC-GLRLM_GLNU were positively correlated with TLG, which indicate that PDAC with higher TLG might have higher heterogeneity of ADC map. Whether there is an intrinsic link between these texture features and tumor biological behaviors requires further research. Ultimately, like the readings of a radiologist, texture analyses should contain all image sequences. Since such research has just begun, separate and gradually deepening study may be a suitable method. Although radiomics is a promising tool for high-tech hybrid imaging technology such as PET/CT and PET/MR (39–41), but many factors such as attenuation correction techniques, different uptake times and voxel size may influence the radiomic features (42, 43), which makes the application value obscure. We need to be cautious about the results of the present study, and continue to increase the sample size and research centers to further investigate the exact value of texture features and radiomics in PET/MR.

This study has several limitations. First, this study is a retrospective study, and the number of patients is limited. Second, evaluating of serum tumor markers vs. imaging parameters was not included in this study. Third, there is a certain degree of subjectivity in manually delineating the tumor boundaries. Therefore, prospective studies with a larger sample and multicenter studies are needed to confirm the present findings. Another limitation of this study is that some of the patients underwent only abdominal PET/MR sequentially after a whole body ^{18}F-FDG PET/CT.

In conclusion, our preliminary study showed that multi-parameter and textural features of primary tumors in ^{18}F-FDG PET/MR images are reliable biomarkers for predicting synchronous metastatic disease in pretreatment PDAC, which might be helpful for the selection of optimal therapeutic methods. This technique may provide a convenient and non-invasive approach to evaluate the prognosis of PDAC in clinical practice. However, multicenter studies with a large population are needed to confirm these results.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ruijin Hospital Ethics Committee Shanghai Jiao Tong University School of Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XL conceived the idea of the study. XH and JG collected the data. XH, MZ, and XL performed image interpretation and analysis. JG performed the statistical analysis. JG and XL drafted the manuscript. HM and XZ scanned the patients. BL edited and reviewed the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by research grants from Shanghai Shen Kang Hospital Development Center (16CR3110B) and the National Natural Science Foundation of China (81201145). The study sponsor have no roles in the study design, collection, analysis, and interpretation of data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00198/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* (2018) 68:7–30. doi: 10.3322/caac.21442
2. Liu X, AUID- Oho, Fu Y, Chen Q, Wu J, Gao W, et al. Predictors of distant metastasis on exploration in patients with potentially resectable pancreatic cancer. *BMC Gastroenterol.* (2018) 18:168. doi: 10.1186/s12876-018-0891-y
3. Gemenetzis G, AUID- Oho, Groot VP, Blair AB, AUID- Oho, Ding D, et al. Incidence and risk factors for abdominal occult metastatic disease in patients with pancreatic adenocarcinoma. *J Surg Oncol.* (2018) 118:1277–84. doi: 10.1002/jso.25288
4. Lee DH, Lee JM, Hur BY, Joo I, Yi NJ, Suh KS, et al. Colorectal cancer liver metastases: diagnostic performance and prognostic value of PET/MR imaging. *Radiology.* (2016) 280:782–92. doi: 10.1148/radiol.2016151975
5. Fraum TJ, Fowler KJ, Crandall JP, Laforest RA, Salter A, An H, et al. Measurement repeatability of 18F-FDG-PET/CT versus 18F-FDG-PET/MRI in solid tumors of the pelvis. *J Nucl Med.* (2019) 60:1080–6. doi: 10.2967/jnumed.118.218735
6. Beiderwellen K, Grueneisen J, Ruhlmann V, Buderath P, Aktas B, Heusch P, et al. [(18)F]FDG PET/MRI vs. PET/CT for whole-body staging in patients with recurrent malignancies of the female pelvis: initial results. *Eur J Nucl Med Mol Imaging.* (2015) 42:56–65. doi: 10.1007/s00259-014-2902-8
7. Mayerhoefer ME, Raderer M, Jaeger U, Staber P, Kiesewetter B, Senn D, et al. Ultra-early response assessment in lymphoma treatment: [18F]FDG PET/MR captures changes in glucose metabolism and cell density within the first 72 hours of treatment. *Eur J Nucl Med Mol Imaging.* (2018) 45:931–40. doi: 10.1007/s00259-018-3937-z
8. Mayerhoefer ME, Prosch H, Beer L, Tamandl D, Beyer T, Hoeller C, et al. PET/MRI versus PET/CT in oncology: a prospective single-center study of 330 examinations focusing on implications for patient management and cost considerations. *Eur J Nucl Med Mol Imaging.* (2019) 47:51–60. doi: 10.1007/s00259-019-04452-y
9. Joo I, Lee JM, Lee DH, Lee ES, Paeng JC, Lee SJ, et al. Preoperative assessment of pancreatic cancer with FDG PET/MR imaging versus FDG PET/CT plus contrast-enhanced multidetector CT: a prospective preliminary study. *Radiology.* (2017) 282:149–59. doi: 10.1148/radiol.2016152798
10. Satoh Y, Ichikawa T, Motosugi U, Kimura K, Sou H, Sano K, et al. Diagnosis of peritoneal dissemination: comparison of 18F-FDG PET/CT, diffusion-weighted MRI, and contrast-enhanced MDCT. *AJR Am J Roentgenol.* (2011) 196:447–53. doi: 10.2214/AJR.10.4687
11. Soussan M, Des Guetz G, Barrau V, Aflalo-Hazan V, Pop G, Mehanna Z, et al. Comparison of FDG-PET/CT and MR with diffusion-weighted imaging for assessing peritoneal carcinomatosis from gastrointestinal malignancy. *Eur Radiol.* (2012) 22:1479–87. doi: 10.1007/s00330-012-2397-2
12. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* (2018) 78:4786–9. doi: 10.1158/0008-5472.CAN-18-0125
13. Chakraborty J, Langdon-Embry L, Cunanan KM, Escalon JG, Allen PJ, Lowery MA, et al. Preliminary study of tumor heterogeneity in imaging predicts two year survival in pancreatic cancer patients. *PLoS ONE.* (2017) 12:e0188022. doi: 10.1371/journal.pone.0188022
14. Cassinotto C, Chong J, Zogopoulos G, Reinhold C, Chiche L, Lafourcade JP, et al. Resectable pancreatic adenocarcinoma: Role of CT quantitative imaging biomarkers for predicting pathology and patient outcomes. *Eur J Radiol.* (2017) 90:152–8. doi: 10.1016/j.ejrad.2017.02.033
15. Cozzi L, AUID- Oho, Comito T, Fogliata A, Franzese C, Franceschini D, et al. Computed tomography based radiomic signature as predictive of survival and local control after stereotactic body radiation therapy in pancreatic carcinoma. *PLoS ONE.* (2019) 14:e0210758. doi: 10.1371/journal.pone.0210758
16. Zukotynski KA, Vajapeyam S, Fahey FH, Kocak M, Brown D, Ricci KI, et al. Correlation of (18)F-FDG PET and MRI apparent diffusion coefficient histogram metrics with survival in diffuse intrinsic pontine glioma: a report from the pediatric brain tumor consortium. *J Nucl Med.* (2017) 58:1264–9. doi: 10.2967/jnumed.116.185389
17. Enkhbaatar NE, Inoue S, Yamamuro H, Kawada S, Miyaoka M, Nakamura N, et al. MR imaging with apparent diffusion coefficient histogram analysis: evaluation of locally advanced rectal cancer after chemotherapy and radiation therapy. *Radiology.* (2018) 288:129–37. doi: 10.1148/radiol.2018171804
18. Liu L, Liu Y, Xu L, Li Z, Lv H, Dong N, et al. Application of texture analysis based on apparent diffusion coefficient maps in discriminating different stages of rectal cancer. *J Magn Reson Imaging.* (2017) 45:1798–808. doi: 10.1002/jmri.25460
19. Shindo T, Fukukura Y, Umanodan T, Takumi K, Hakamada H, Nakajo M, et al. Histogram analysis of apparent diffusion coefficient in differentiating

pancreatic adenocarcinoma and neuroendocrine tumor. *Medicine.* (2016) 95:e2574. doi: 10.1097/MD.0000000000002574
20. Chen SW, Shen WC, Chen WT, Hsieh TC, Yen KY, Chang JG, et al. Metabolic imaging phenotype using radiomics of [(18)F]FDG PET/CT associated with genetic alterations of colorectal cancer. *Mol Imaging Biol.* (2019) 21:183–90. doi: 10.1007/s11307-018-1225-8
21. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. (1988) 44:837–45. doi: 10.2307/2531595
22. Karabicak I, Satoi S, Yanagimoto H, Yamamoto T, Hirooka S, Yamaki S, et al. Risk factors for latent distant organ metastasis detected by staging laparoscopy in patients with radiologically defined locally advanced pancreatic ductal adenocarcinoma. *J Hepatobiliary Pancreat Sci.* (2016) 23:750–5. doi: 10.1002/jhbp.408
23. Mackay TM, van Erning FN, van der Geest L, de Groot J, Haj Mohammad N, Lemmens VE, et al. Association between primary origin (head, body and tail) of metastasised pancreatic ductal adenocarcinoma and oncologic outcome: a population-based analysis. *Eur J Cancer.* (2019) 106:99–105. doi: 10.1016/j.ejca.2018.10.008
24. Lee SH, Hwang HK, Lee WJ, Yun M, Kang CM. Preoperative metabolic tumor volume2.5 associated with early systemic metastasis in resected pancreatic cancer: a transcriptome-wide analysis. *Gut Liver.* (2019) 13:356–65. doi: 10.5009/gnl18242
25. Wilson JM, Mukherjee S, Brunner TB, Partridge M, Hawkins MA. Correlation of (18)F-fluorodeoxyglucose positron emission tomography parameters with patterns of disease progression in locally advanced pancreatic cancer after definitive chemoradiotherapy. *Clin Oncol.* (2017) 29:370–7. doi: 10.1016/j.clon.2017.01.038
26. Lee JW, Kang CM, Choi HJ, Lee WJ, Song SY, Lee JH, et al. Prognostic value of metabolic tumor volume and total lesion glycolysis on preoperative (1)(8)F-FDG PET/CT in patients with pancreatic cancer. *J Nucl Med.* (2014) 55:898–904. doi: 10.2967/jnumed.113.131847
27. Im HJ, Oo S, Jung W, Jang JY, Kim SW, Cheon GJ, et al. Prognostic value of metabolic and volumetric parameters of preoperative FDG-PET/CT in patients with resectable pancreatic cancer. *Medicine.* (2016) 95:e3686. doi: 10.1097/MD.0000000000003686
28. Sheikhbahaei S, Wray R, Young B, Mena E, Taghipour M, Rahmim A, et al. 18F-FDG-PET/CT therapy assessment of locally advanced pancreatic adenocarcinoma: impact on management and utilization of quantitative parameters for patient survival prediction. *Nucl Med Commun.* (2016) 37:231–8. doi: 10.1097/MNM.0000000000000436
29. Chirindel A, Alluri KC, Chaudhry MA, Wahl RL, Pawlik TM, Herman JM, et al. Prognostic value of FDG PET/CT-derived parameters in pancreatic adenocarcinoma at initial PET/CT staging. *AJR Am J Roentgenol.* (2015) 204:1093–9. doi: 10.2214/AJR.14.13156
30. Chen BB, Tien YW, Chang MC, Cheng MF, Chang YT, Wu CH, et al. PET/MRI in pancreatic and periampullary cancer: correlating diffusion-weighted imaging, MR spectroscopy and glucose metabolic activity with clinical stage and prognosis. *Eur J Nucl Med Mol Imaging.* (2016) 43:1753–64. doi: 10.1007/s00259-016-3356-y
31. Chen BB, Tien YW, Chang MC, Cheng MF, Chang YT, Yang SH, et al. Multiparametric PET/MR imaging biomarkers are associated with overall survival in patients with pancreatic cancer. *Eur J Nucl Med Mol Imaging.* (2018) 45:1205–17. doi: 10.1007/s00259-018-3960-0
32. Schob S, Meyer HJ, Dieckow J, Pervinder B, Pazaitis N, Hohn AK, et al. histogram analysis of diffusion weighted imaging at 3T is useful for prediction of lymphatic metastatic spread, proliferative activity, and cellularity in thyroid cancer. *Int J Mol Sci.* (2017) 18:E821. doi: 10.3390/ijms18040821
33. Becker AS, Ghafoor S, Marcon M, Perucho JA, Wurnig MC, Wagner MW, et al. MRI texture features may predict differentiation and nodal stage of cervical cancer: a pilot study. *Acta Radiol Open.* (2017) 6:2058460117729574. doi: 10.1177/2058460117729574
34. Tsuchiya N, Doai M, Usuda K, Uramoto H, Tonami H. Non-small cell lung cancer: Whole-lesion histogram analysis of the apparent diffusion coefficient for assessment of tumor grade, lymphovascular invasion and pleural invasion. *PLoS ONE.* (2017) 12:e0172433. doi: 10.1371/journal.pone.0172433
35. Cheng Z, Zhang J, He N, Li Y, Wen Y, Xu H, et al. Radiomic features of the nigrosome-1 region of the substantia nigra: using quantitative susceptibility mapping to assist the diagnosis of idiopathic Parkinson's disease. *Front Aging Neurosci.* (2019) 11:167. doi: 10.3389/fnagi.2019.00167
36. Liang B, Yan H, Tian Y, Chen X, Yan L, Zhang T, et al. Dosiomics: extracting 3D spatial features from dose distribution to predict incidence of radiation pneumonitis. *Front Oncol.* (2019) 9:269. doi: 10.3389/fonc.2019.00269
37. Aide N, Talbot M, Fruchart C, Damaj G, Lasnon C. Diagnostic and prognostic value of baseline FDG PET/CT skeletal textural features in diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imaging.* (2018) 45:699–711. doi: 10.1007/s00259-017-3899-6
38. Soussan M, Orlhac F, Boubaya M, Zelek L, Ziol M, Eder V, et al. Relationship between tumor heterogeneity measured on FDG-PET/CT and pathological prognostic factors in invasive breast cancer. *PLoS ONE.* (2014) 9:e94017. doi: 10.1371/journal.pone.0094017
39. Baiocco S, Sah BR, Mallia A, Kelly-Morland C, Neji R, Stirling JJ, et al. Exploratory radiomic features from integrated 18F-fluorodeoxyglucose positron emission tomography/magnetic resonance imaging are associated with contemporaneous metastases in oesophageal/gastroesophageal cancer. *Eur J Nucl Med Mol Imaging.* (2019) 46:1478–84. doi: 10.1007/s00259-019-04306-7
40. Kirienko M, Cozzi L, Antunovic L, Lozza L, Fogliata A, Voulaz E, et al. Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery. *Eur J Nucl Med Mol Imaging.* (2018) 45:207–17. doi: 10.1007/s00259-017-3837-7
41. Antunovic L, De Sanctis R, Cozzi L, Kirienko M, Sagona A, Torrisi R, et al. PET/CT radiomics in breast cancer: promising tool for prediction of pathological response to neoadjuvant chemotherapy. *Eur J Nucl Med Mol Imaging.* (2019) 46:1468–77. doi: 10.1007/s00259-019-04313-8
42. Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur J Nucl Med Mol Imaging.* (2019) 46:2638–55. doi: 10.1007/s00259-019-04391-8
43. Vuong D, Tanadini-Lang S, Huellner MW, Veit-Haibach P, Unkelbach J, Andratschke N, et al. Interchangeability of radiomic features between [18F]-FDG PET/CT and [18F]-FDG PET/MR. *Med Phys.* (2019) 46:1677–85. doi: 10.1016/S0167-8140(18)32417-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 07 May 2020
doi: 10.3389/fonc.2020.00572

Feasibility of Automated Volumetric Assessment of Large Hepatocellular Carcinomas' Responses to Transarterial Chemoembolization

*Ahmed W. Moawad[1], David Fuentes[1], Ahmed M. Khalaf[2], Katherine J. Blair[2], Janio Szklaruk[2], Aliya Qayyum[2], John D. Hazle[1] and Khaled M. Elsayes[2]**

[1] Imaging Physics Department, University of Texas MD Anderson Cancer Center, Houston, TX, United States, [2] Diagnostic Radiology Department, University of Texas MD Anderson Cancer Center, Houston, TX, United States

Background: Hepatocellular carcinoma (HCC) is the most common liver malignancy and the leading cause of death in patients with cirrhosis. Various treatments for HCC are available, including transarterial chemoembolization (TACE), which is the commonest intervention performed in HCC. Radiologic tumor response following TACE is an important prognostic factor for patients with HCC. We hypothesized that, for large HCC tumors, assessment of treatment response made with automated volumetric response evaluation criteria in solid tumors (RECIST) might correlate with the assessment made with the more time- and labor-intensive unidimensional modified RECIST (mRECIST) and manual volumetric RECIST (M-vRECIST) criteria. Accordingly, we undertook this retrospective study to compare automated volumetric RECIST (A-vRECIST) with M-vRECIST and mRESIST for the assessment of large HCC tumors' responses to TACE.

Methods: We selected 42 pairs of contrast-enhanced computed tomography (CT) images of large HCCs. Images were taken before and after TACE, and in each of the images, the HCC was segmented using both a manual contouring tool and a convolutional neural network. Three experienced radiologists assessed tumor response to TACE using mRECIST criteria. The intra-class correlation coefficient was used to assess inter-reader reliability in the mRECIST measurements, while the Pearson correlation coefficient was used to assess correlation between the volumetric and mRECIST measurements.

Results: Volumetric tumor assessment using automated and manual segmentation tools showed good correlation with mRECIST measurements. For A-vRECIST and M-vRECIST, respectively, $r = 0.597$ vs. 0.622 in the baseline studies; 0.648 vs. 0.748 in the follow-up studies; and 0.774 vs. 0.766 in the response assessment ($P < 0.001$ for all). The A-vRECIST evaluation showed high correlation with the M-vRECIST evaluation ($r = 0.967$, 0.937, and 0.826 in baseline studies, follow-up studies, and response assessment, respectively, $P < 0.001$ for all).

Conclusion: Volumetric RECIST measurements are likely to provide an early marker for TACE monitoring, and automated measurements made with a convolutional neural network may be good substitutes for manual volumetric measurements.

Keywords: volumetric RECIST, hepatocellular carcinoma, TACE, automated segmentation, tumor response

OPEN ACCESS

Edited by:
Lei Deng,
Jacobi Medical Center, United States

Reviewed by:
Guolin Ma,
China-Japan Friendship Hospital, China
Yu Jiang,
University of Memphis, United States
Su Jong Yu,
Seoul National University Hospital, South Korea
Yaojun Zhang,
Sun Yat-sen University Cancer Center (SYSUCC), China

***Correspondence:**
Khaled M. Elsayes
kmelsayes@mdanderson.org

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *28 August 2019*
Accepted: *30 March 2020*
Published: *07 May 2020*

Citation:
Moawad AW, Fuentes D, Khalaf AM, Blair KJ, Szklaruk J, Qayyum A, Hazle JD and Elsayes KM (2020) Feasibility of Automated Volumetric Assessment of Large Hepatocellular Carcinomas' Responses to Transarterial Chemoembolization. Front. Oncol. 10:572. doi: 10.3389/fonc.2020.00572

INTRODUCTION

Hepatocellular carcinoma (HCC) is the most common liver malignancy and the leading cause of death in patients with cirrhosis. Despite advances in various treatment modalities over the past several years, the prognosis for HCC remains poor, with 5-year overall survival ranging from 24 to 41% (1, 2). Efforts have been made to improve early detection of HCC by the performance of frequent screening in high-risk populations. However, most cases are still diagnosed at intermediate to advanced stages (3). These patients are not candidates for curative therapies, such as surgical resection or liver transplant. As a result, treatment options for this patient population are limited to loco-regional treatments, including local radiofrequency ablation, radio and chemoembolization, and systemic chemotherapy with Sorafenib (4–6).

In the United States, transarterial chemoembolization (TACE) is the most common intervention for HCC (7). It is the standard of care for patients with intermediate-stage HCC (according to Barcelona clinic liver cancer (BCLC) staging, whether it is large tumor or multi-nodular). In addition, it may be used in advanced HCC prior to systemic therapy or as a bridging therapy prior to surgery (8). There is evidence that repeat TACE may also be beneficial in patients with advanced HCC (9–11). Radiologic tumor response following initial TACE has been shown to be an important prognostic factor for patients with HCC. Baseline imaging is usually obtained 2–3 weeks before therapy and follow-up imaging is performed 4–6 weeks after therapy. The most commonly used criteria for tumor response following HCC is mRECIST (1-dimension) or EASL (2-dimensions).

The recent attempts to improve the accuracy of radiologic response criteria to predict overall survival and Progression-free survival have focused on using quantitative volumetric analysis. This has resulted in the development of the volumetric RECIST (vRECIST) and quantitative EASL (qEASL) methods with better results in predicting patient's outcome than the currently used mRECIST (12–15).

The volumetric assessment of tumor response depends on manual segmentation of tumor and needs contouring of the lesion in every single slice of the study. This process is time consuming and tedious. The manual volumetric assessment show high inter- and intra-observer variability which make it impractical in daily practice (16–18). On the other hand, automated volumetric segmentation has the potential to reduce time for this process and make volumetric assessment of tumors more practical, with lower inter and intra-observer variability than both mRECIST and manual volumetric assessment (12, 18, 19).

Abbreviations: A-vRECIST, automated volumetric response evaluation criteria in solid tumors; CNN, convolutional neural network; CT, computed tomography; HCC, hepatocellular carcinoma; ICC, intra-class correlation coefficient; mRECIST, modified response evaluation criteria in solid tumors; MRI, magnetic resonance imaging; M-vRECIST, manual volumetric response evaluation criteria in solid tumors; RECIST, response evaluation criteria in solid tumors; TACE, transarterial chemo embolization; TNM, tumor-node-metastasis; vRECIST, volumetric response evaluation criteria in solid tumors.

Convolutional neural network (CNN) shows promise to achieve automated segmentation of liver and liver masses. These are, however, computationally demanding (12, 13). CNN in tumor segmentation has been found to be more accurate and closer to the manual volumetric segmentation in larger tumors, with far lower accuracy in smaller liver masses (14, 15).

The purpose of this study is to assess the feasibility of volumetric assessment of pre- and post-TACE HCC using fully automated segmentation and to evaluate the correlation of automated volumetric assessment with both manual volumetric assessment and mRECIST measurements.

MATERIALS AND METHODS

Study Cohort

This is a retrospective, single-institution, IRB approved study. This study included patients with large HCC tumors ($\geq$5 cm) diagnosed and treated at our institution between November 2002 and June 2012. Patients were included in the study if (1) they had undergone TACE as their sole first-line or initial bridging therapy; (2) their medical records included multiphasic, contrast-enhanced CT images that were obtained at baseline and that included no image artifacts (e.g., surgical clips); and (3) their tumor was diagnosed as tumor-node-metastasis (TNM) stage III or IV HCC based on the American Joint Committee on Cancer (AJCC). Although there are numerous HCC scoring systems that incorporate liver functional reserve, patient performance, and gross tumor characteristics (e.g., size, vascular invasion, number of lesions), we chose the TNM staging system because it is the only HCC staging system that considers tumor characteristics (including size) without taking any other factor into consideration (20, 21).

TACE

Briefly, TACE is delivery method of chemotherapy delivery to the tumor through its feeding arteries using trans-arterial approach. The hepatic artery was selected and injected with chemotherapy by super selection of the feeding vasculature using advanced micro-catheters. There are two main chemotherapeutic regimens that may be delivered. In conventional TACE, a mixture of radio-opaque ethiodized emulsion oil (Lipiodol; Guerbet, Villepinte, France) and doxorubicin or cisplatin was injected followed by embolization of the feeding vessels. While in TACE with drug eluting beads (DEB-TACE), a mixture of micro-sphere particles, doxorubicin and soluble non-ionic contrast was injected (22).

Our patients received one of the following chemotherapeutic regimens (Drug regimen details are missed from 2 cases): (1) embolic microsphere beads (Biocompatibles, UK) loaded with doxorubicin (DEBDOX; drug-eluting bead doxorubicin) (15 lesions), (2) cisplatin, doxorubicin, and mitomycin C (22 lesions), or (3) cisplatin and mitomycin C only (3 lesions) (such information is missed from 2 cases).

CT Imaging Technique

All patients underwent dynamic, contrast-enhanced CT scans of the abdomen on 4-, 16-, or 64-slice multidetector CT LightSpeed

scanners (GE Healthcare, Chicago, IL) pre-TACE and post-TACE. Liver protocol was used in all studies (the arterial, porto-venous, and delayed phases were captured 17, 60 s, and ∼5 min, respectively, after peak enhancement of the descending aorta). Injection was done with an automated contrast injector using a bolus tracking technique and an injection rate of about 3–5 mL/s. The image reconstruction thicknesses were 2.5 mm and 5 mm.

Assessment of Tumor Response

Tumor TACE response was assessed using mRECIST, M-vRECIST, and A-vRECIST (**Figure 1**). Three different board certified radiologists (KE, JS, and AQ), each with more than 20 years of experience in abdominal imaging, independently measured tumors using mRECIST criteria. The changes in measurements between the follow-up and baseline CT scans were reported, and tumor viability and enhancement in the late arterial phase were taken into consideration. Volume assessment using M-vRECIST and A-vRECIST at baseline and follow-up studies was also done. The Convert3D medical image processing tool was used to extract the segmentation volumes according to the voxel extensions (23).

Tumor Segmentation

The porto-venous phase of CT (both baseline and follow-up) were used to simplify lesion assessment, they were exported in DICOM format from our institution's picture archiving and communication system to a separate research server. Subsequently, the images were converted into the format recommended by the Neuroimaging Informatics Technology Initiative (Nifti) to preserve the orientation information for further data processing. Then the files were compressed and the images were reoriented into right-anterior-inferior orientation with Convert3D.

Manual segmentations were performed in the portal-venous phase of contrast administration. This was performed for the baseline and follow-up CT studies. A semi-automated segmentation tool available in Amira Software (Thermo Fisher Scientific, Waltham, MA) was used to delineate the tumors, including the (i) enhancing portions, (ii) non-enhancing portions, and Lipiodol containing portions of the tumors. Enhancing tumor tissue was defined as a "region with uptake of contrast agent in the arterial phase of dynamic contrast CT" while non-enhancing tumor tissue was defined as a "region of no enhancement within HCC on the arterial phase images," while Lipidol was defined as a "portions of high attenuation in pre-contrast images." Manual segmentation of the whole tumor in the baseline study was done by one author (AM) to ensure the consistency of the segmentation throughout the dataset. The segmentation was done for all axial CT images. These manual segmentations provided the training data ($n = 42$ pairs) used to develop a neural network classifier for segmentation of tumors from the background liver tissue. Automated segmentation, performed using a CNN approach (U-Net), was used to segment the liver and tumor in two steps (24). To determine the correlation between mRECIST and MvRECIST and mRECIST and A-vRECIST, we compared the average uni-dimensional mRECIST measurements of the 3 readers to the M-vRECIST and A-vRECIST volumetric assessments.

Statistical Analysis

Statistical analysis was performed using IBM SPSS Statistics V.24.0 software (IBM, Armonk, NY). Inter-reader reliability for mRECIST measurement was assessed using the intraclass correlation coefficient (ICC). The Pearson correlation coefficient (r) was used to measure the correlation between the diameter change (the average reading from mRECIST) and (i) the automated and (ii) manual volume changes after TACE. A *P*-value of 0.05 was used to determine the statistical significance of the measurements.

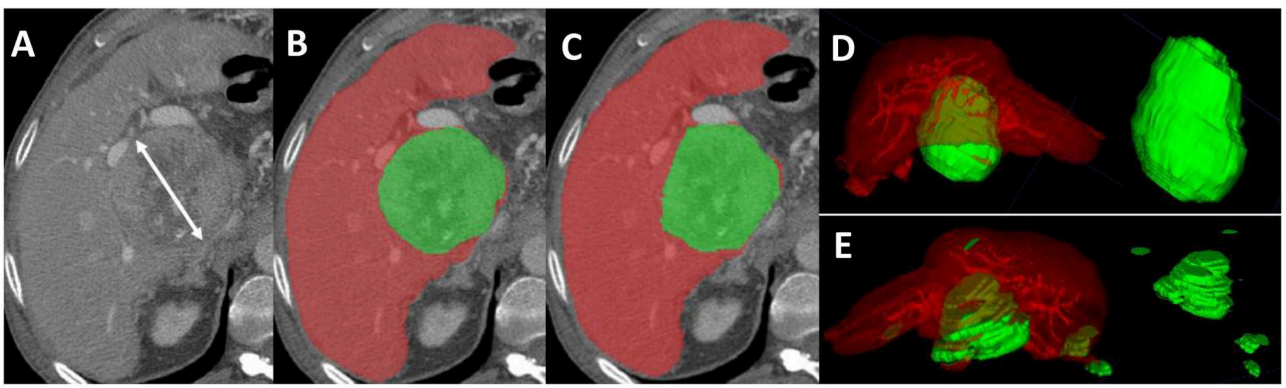


FIGURE 1 | Axial sections from a pre-TACE abdominal CT scan of a 64-year old male patient with advanced HCC. **(A)** Shows an 8.5 cm lesion measured with mRECIST criteria. **(B)** Shows the manual segmentation of the hepatic parenchyma (red) and the HCC tissue (green); the tumor volume, measured with M-vRECIST, was 377.25 cm^3. **(C)** Shows the automated segmentation of the parenchyma (red) and HCC tissue (green); the tumor volume measured with A-vRECIST was 187.9 cm^3. **(D,E)** Show the 3-dimensional voxel rendering according to M-vRECIST and A-vRECIST, respectively. A-vRECIST, automated volumetric response evaluation criteria in solid tumors; CT, computed tomography; HCC, hepatocellular carcinoma; mRECIST, modified response evaluation criteria in solid tumors; M-vRECIST, manual volumetric response evaluation criteria in solid tumors; RECIST, response evaluation criteria in solid tumors; TACE, transarterial chemo embolization.

RESULTS

Patient Characteristics

There were 320 patients (with complete medical and survival data) found in our institutional database with HCC patients underwent TACE, we excluded 8 patients due to difference in their treatment plan (either TACE was used as second line or combined with other form of therapy). After thorough review of patients imaging studies, another 209 patients were excluded for different reasons (**Figure 2**). The final 103 patients were categorized according to their TNM stage into stage I, II, III, and IV with 36, 25, 24, and 18 patients. A total of 42 patients met our inclusion criteria for the study (TNM stage III and IV). On average, patients' baseline CT scans were performed 4 weeks (range between 2 and 7 weeks) before the first TACE session, and their follow-up scans were performed 11 weeks (range between 8 and 13 weeks) after the first TACE session. Patients' demographic data and baseline tumor characteristics are provided in **Table 1**.

Manual and Automated Volumetric Assessment

The mRECIST response comparing the pre- and post-TACE images is listed in **Table 2** for the three radiologists'.

The manual vRECIST and automated vRECIST response comparing the pre- and post TACE are presented in **Table 3**, A-vRECIST was obtained with CNN. Assessment with the ICC showed statistically significant correlation between the three readers (ICC = 0.824; 95% CI = 0.70–0.89; $P < 0.001$).

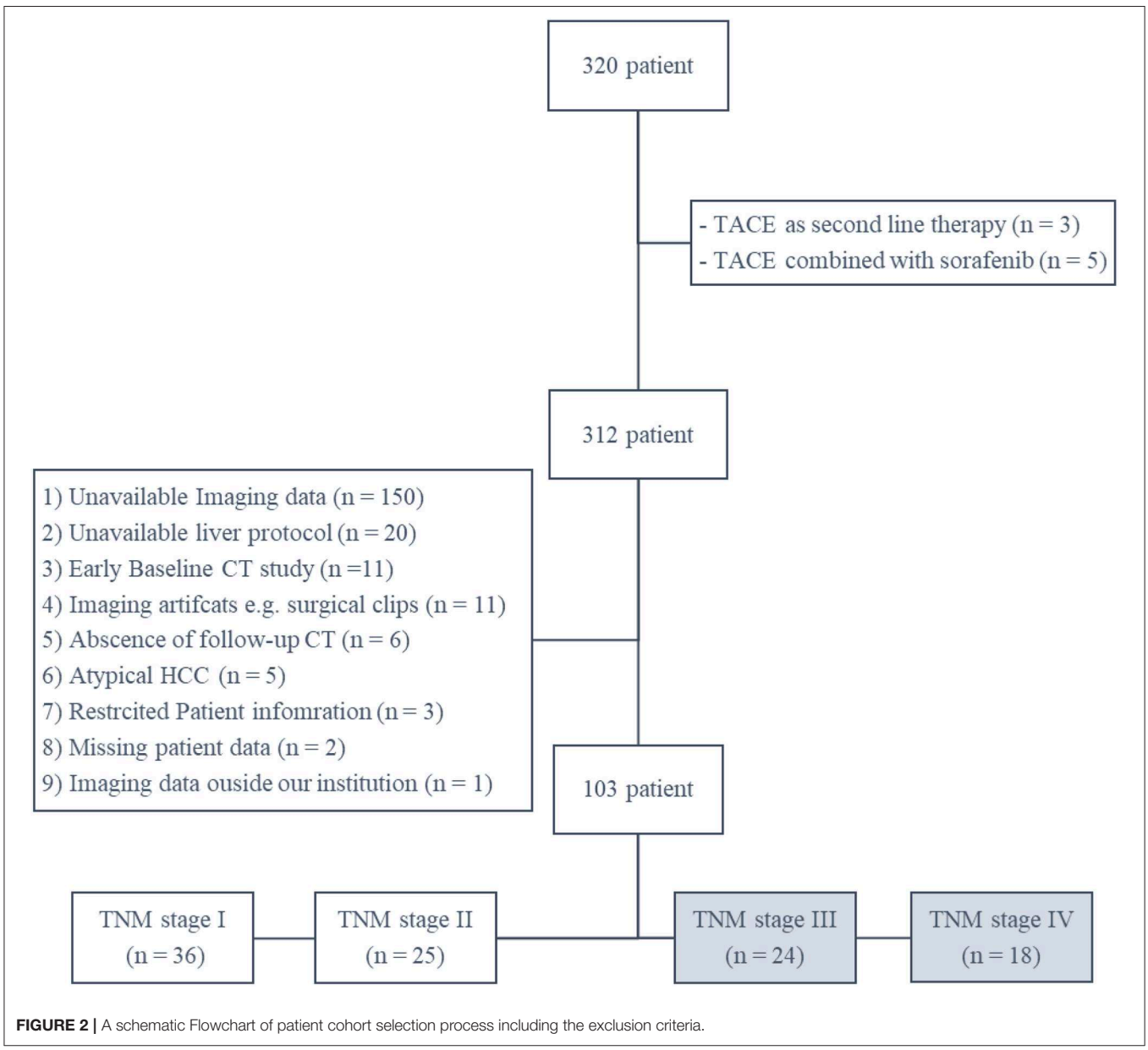


FIGURE 2 | A schematic Flowchart of patient cohort selection process including the exclusion criteria.

TABLE 1 | Patient demographic data and baseline tumor characteristics.

Baseline characteristics	Value (*n* or mean ± SD) *n* = 42, 100%
Demographics	
Age, years	67 ± 7
Sex ratio (male/female)	29 (69%)/13 (31%)
Cirrhosis (yes/no)	31 (74%)/11 (26%)
Etiology	
Tobacco use (yes/no)	28 (67%)/14 (33%)
Alcohol (yes/no)	25 (60%)/17 (40%)
Diabetes mellitus (yes/no)	12 (29%)/30 (71%)
Family history of cancer (yes/no)	24 (57%)/18 (43%)
Hepatitis (HBV/HCV/both/none)	3/7/8/24
Tumor extension	
Overall tumor size (cm)*	10 ± 5
Vascular invasion present	17 (40%)
Diffuse/infiltrative pattern	10 (24%)
Tumor involvement (≤50%/>50% of liver volume)	31/11
Alpha fetoprotein (ng/ml)**	38.5 ± 1567.3
Distant metastasis present	7 (17%)
Nodal metastasis present	14 (33%)
Portal vein thrombosis present	13 (31%)
Tumor nodularity (uni-/multilobular)	14/28
mRECIST category (CR/PR/SD/PD)	10/20/7/5
HCC scoring	
CLIP staging	
Stage 0	4 (9%)
Stage 1	17 (40%)
Stage 2	10 (24%)
Stage 3	7 (17%)
Stage 4	2 (5%)
Stage 5	2 (5%)
Okuda staging	
Stage I	26 (62%)
Stage II	16 (38%)
TNM staging	
Stage III	24 (57%)
Stage IV	18 (43%)
BCLC staging	
Stage B	7 (17%)
Stage C	35 (83%)

**Overall tumor size was determined based on RECIST measurements.*
***Alpha fetoprotein is reported using median ± interquartile range.*
BCLC, Barcelona clinic liver cancer; CLIP, Cancer of the Liver Italian Program; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; RECIST, response evaluation criteria in solid tumors; SD, standard deviation; TNM, tumor-node-metastasis; CR, Complete Response; PR, Partial Response; SD, Stable disease; PD, Progressive Disease.

TABLE 2 | Radiologists' unidimensional mRECIST measurements of HCC tumors and average readings.

	Baseline mRECIST (cm) Mean ± SD; 95% CI	Follow-up mRECIST (cm) Mean ± SD; 95% CI	Diameter change (cm) Mean ± SD; 95% CI
Reader 1	7.9 ± 4.8 (6.4–9.4)	5.9 ± 4.7 (4.4–7.4)	−1.9 ± 4.1 (−3.3 to −0.7)
Reader 2	8.6 ± 5.3 (6.9–10.2)	6 ± 4.9 (4.4–7.5)	−2.6 ± 3.2 (−3.3 to −1.6)
Reader 3	7.8 ± 4.7 (6.4–9.4)	5.6 ± 4.9 (4.1–7.2)	−2.2 ± 2.9 (−3.2 to −1.3)
Average	8.1 ± 4.7 (6.6– 9.6)	5.9 ± 4.6 (4.4–7.3)	−2.3 ± 2.9 (−3.2 to −1.3)

CI, confidence interval; HCC, hepatocellular carcinoma; mRECIST, modified response evaluation criteria in solid tumors; SD, standard deviation.

TABLE 3 | Voxel-based volumetric measurements of HCC tumors made using M-vRECIST and A-vRECIST with CNN.

	M-vRECIST Mean ± SD; 95% CI (cm^3)	A-vRECIST with CNN Mean ± SD; 95% CI (cm^3)
Baseline study	466.8 ± 600 (279.8–653.9)	438.6 ± 552.9 (266.4–611)
Follow-up study	537.9 ± 772.8 (297.1–778.8)	426.2 ± 673.9 (216.3–636.3)
Volume change	71 ± 322.9 (−29.5–171.7)	−12.39 ± 395.4 (−135.6–110.8)

A-vRECIST, automated volumetric response evaluation criteria in solid tumors; CI, confidence interval; CNN, convolutional neural network; HCC, hepatocellular carcinoma; M-vRECIST, manual volumetric response evaluation criteria in solid tumors; SD, standard deviation.

To determine the correlation between mRECIST and vRECIST (both manual and automated), we compared the average unidimensional mRECIST measurements to the M-vRECIST and A-vRECIST volumetric assessments (**Figure 1**) and compared the diameter changes determined through the manual and automated volumetric assessments. The correlation between mRECIST and M-vRECIST was moderate for both the baseline and the follow-up studies ($r = 0.622$ and 0.748, respectively). The correlation between tumor diameter measurement changes was higher ($r = 0.766$). The differences between these measurements were statistically significant ($P < 0.001$ for all). The correlation between mRECIST and A-vRECIST was similar: $r = 0.597$ for the baseline and $r = 0.648$ for the follow-up studies ($P < 0.001$ for all). For the correlation between the baseline and follow-up tumor measurements, $r = 0.774$.

We used the Pearson correlation coefficient to compare M-vRECIST and A-vRECIST and found strong linear correlation between the two approaches ($r = 0.967$ for the baseline studies, $r = 0.937$ for the follow-up studies, and $r = 0.826$ for the tumor volume change after TACE [$P < 0.001$ for all]) (**Figures 3**, **4**).

DISCUSSION

In this study, we hypothesized that, for large HCC tumors, assessment of volume changes before and after TACE using A-vRECIST would correlate with the measurement changes using uni-dimentional mRECIST and that, accordingly, A-vRECIST can be used to assess tumor response to TACE therapy.

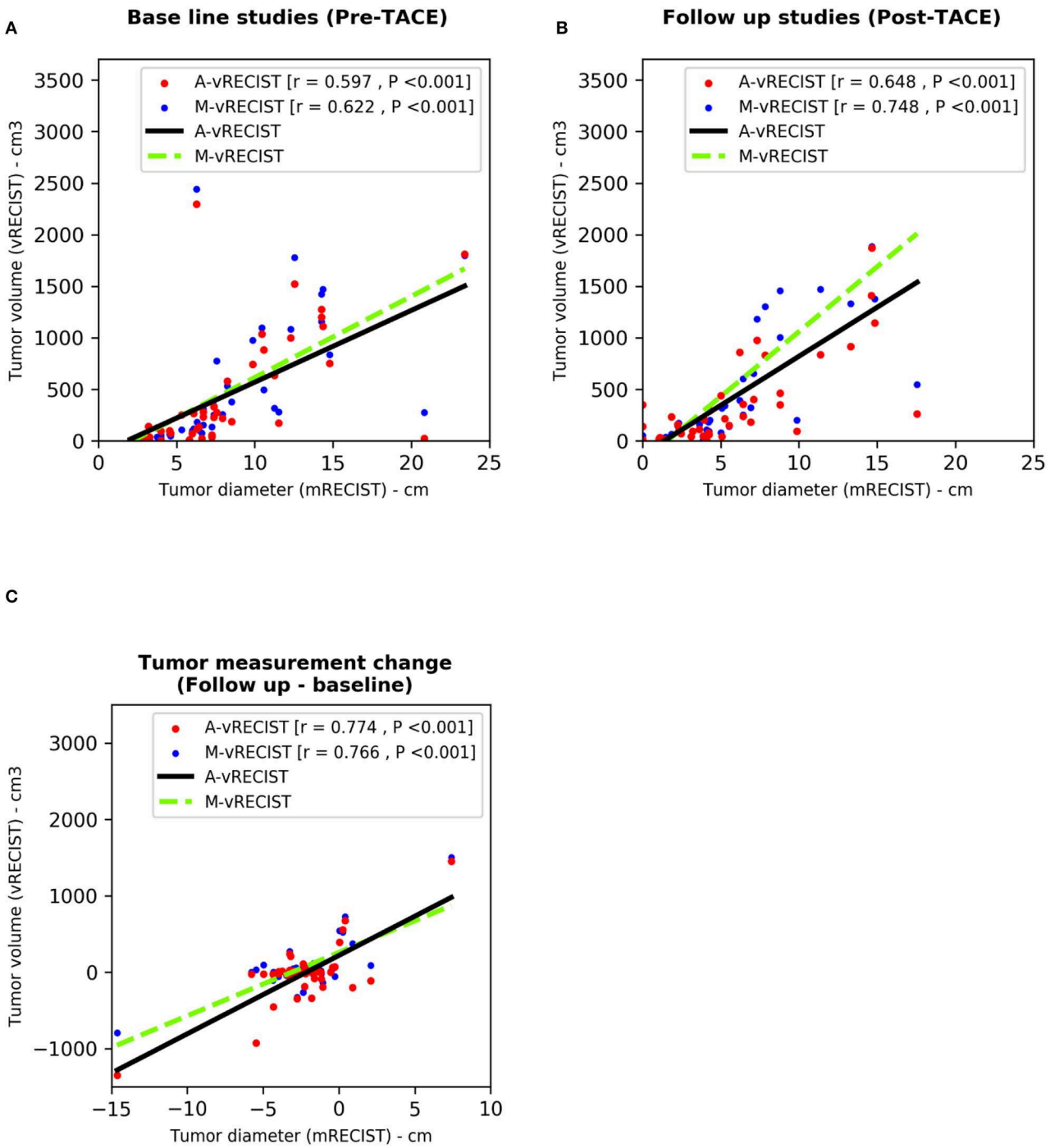


FIGURE 3 | Scatter plots compare tumor assessments from unidimensional mRECIST with those from A-vRECIST and M-vRECIST in pre- and post-TACE studies [graphs **(A,B)**, respectively]. Graph **(C)** shows the difference between the A-vRECIST and M-vRECIST measurements. The best-fit lines and Pearson coefficients (r) are shown. A-vRECIST, automated volumetric response evaluation criteria in solid tumors; mRECIST, modified response evaluation criteria in solid tumors; M-vRECIST, manual volumetric response evaluation criteria in solid tumors; RECIST, response evaluation criteria in solid tumors; TACE, transarterial chemo embolization.

Volumetric assessment of HCC have been emerged as recent tool for assessing of HCC response to treatment. Although assessment of treatment response of is some patterns of HCC is challenging (especially diffuse/infiltrative type) due to its indistinct borders. Previous studies showed that both uni-dimensional and volumetric measurement highly correlated with the actual pathological tumor volume, with volumetric assessment was similar to pathological volume while

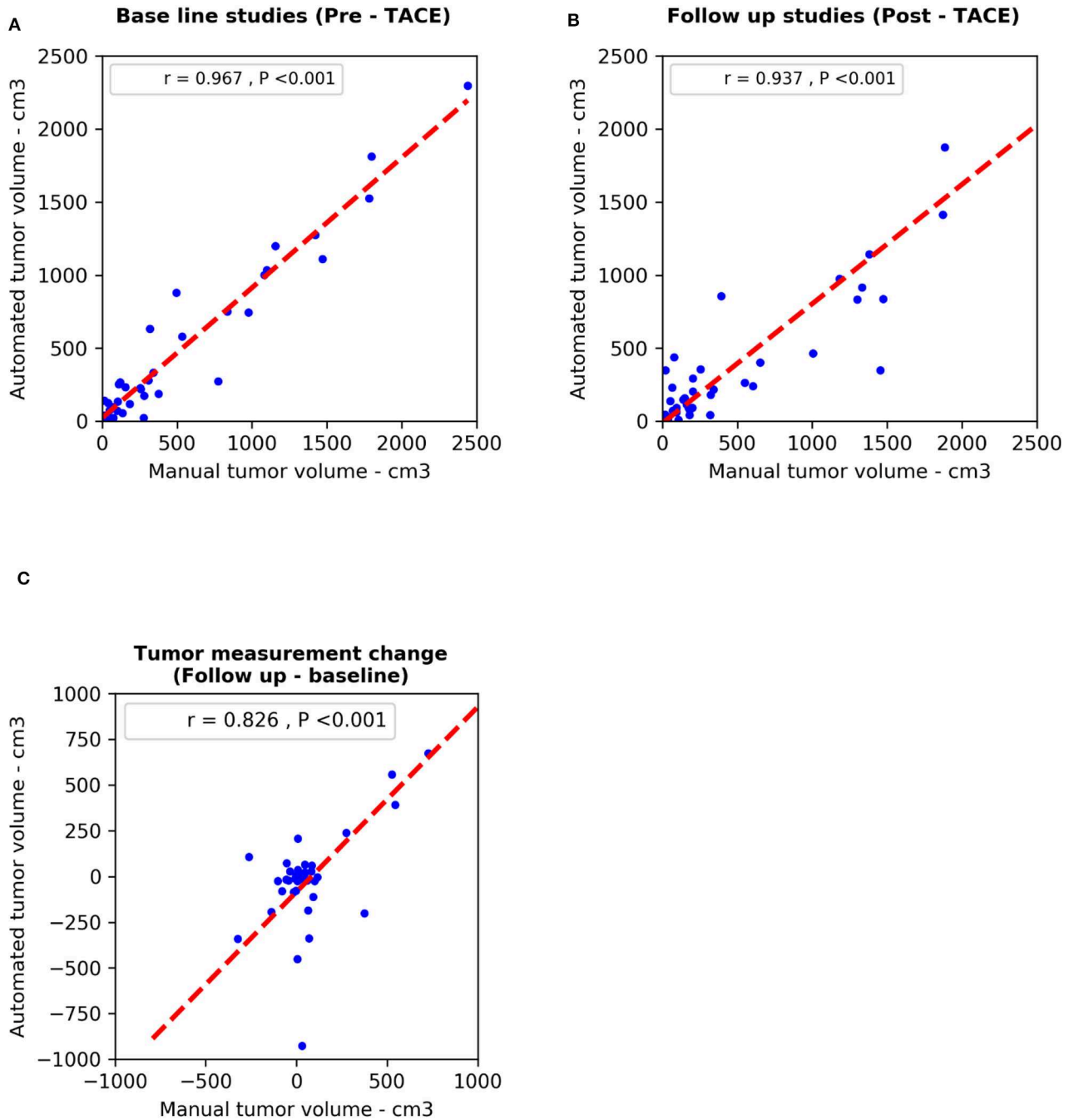


FIGURE 4 | Scatter plots compare diameter changes from M-vRECIST vs. A-vRECIST assessments in the **(A)** baseline and **(B)** follow-up studies. Graph **(C)** shows the tumor measurement changes. The best-fit lines (dashed line) and Pearson correlation coefficients (r) are shown. A-vRECIST, automated volumetric response evaluation criteria in solid tumors; M-vRECIST, manual volumetric response evaluation criteria in solid tumors.

uni-dimensional measurement overestimate the volume by 28% (25, 26). Such studies demonstrate the superiority of volumetric assessment to estimate the real tumor volume, which is more important during assessment og treatment response.

Another study showed that HCC response to Sorafenib using volumetric assessment can be used an alternative tool for monitoring therapy better than mRECIST measurement (27). Another study used functional MRI volumetric analysis of HCC tumors to separate patients into responders and non-responders following treatment with combination TACE and Sorafenib (23, 28). Both of these studies made 3D measurements of HCC tumor volumes. However, manual

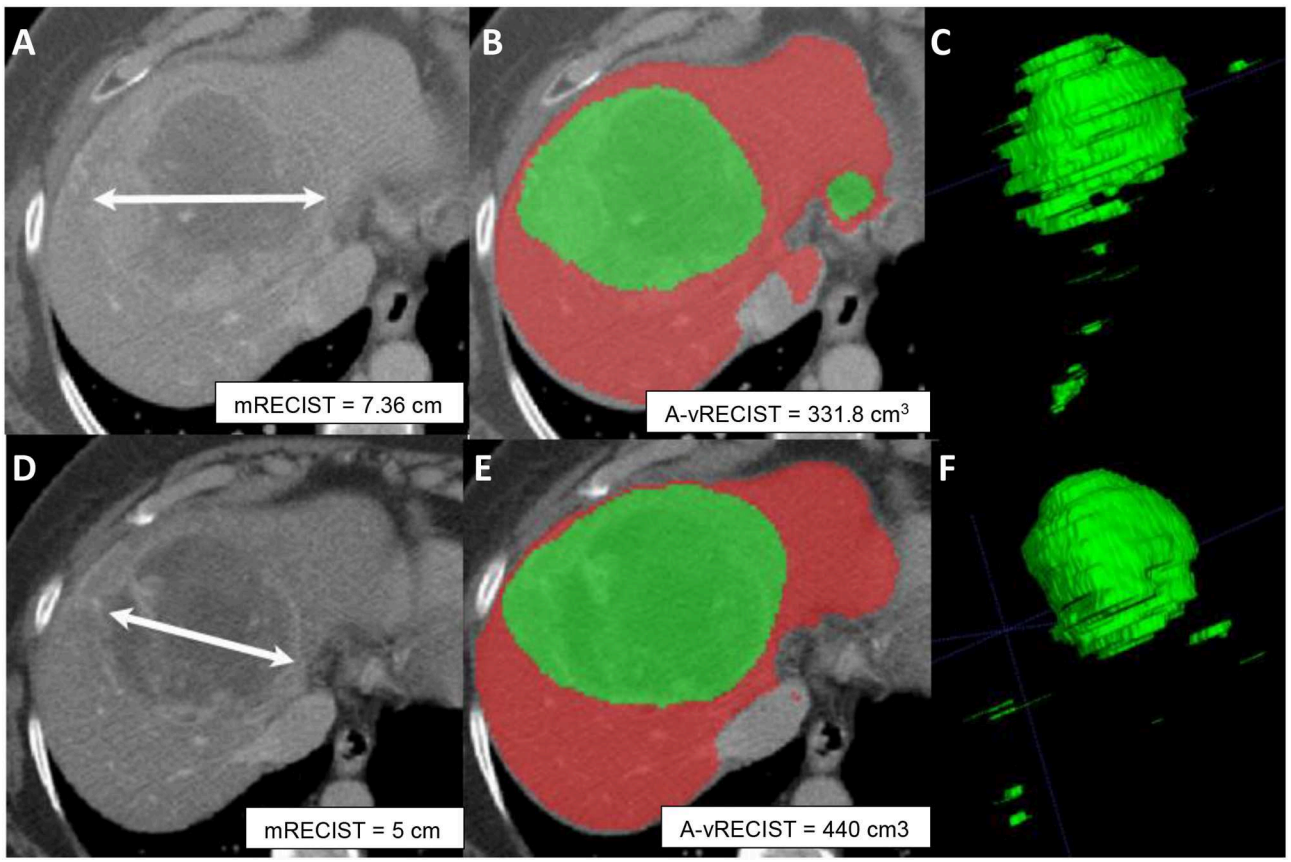


FIGURE 5 | Axial sections from pre-TACE **(A–C)** and 4-week post-TACE **(D–F)** scans from a 70-year old male patient with advanced HCC. **(A,D)** Show changes in tumor size made with mRECIST measurements and indicative of tumor shrinkage. **(B,E)** Show A-vRECIST segmentation of the hepatic parenchyma (red) and the HCC tissue (green) and indicate tumor size increase. **(C,F)** Show the 3-dimensional voxel renderings of the A-vRECIST measurements.

volumetric assessment is time consuming and highly variable leading to motivate scientists to automate this process. Automated tumor volume and enhancement measurements using cross-sectional images are proven to be both reproducible and feasible in clinical application (29). In addition, it have been demonstrated that automated quantitative tumor volume assessment can become part of monitoring response to TACE (29).

There are challenges to the automated segmentation of the liver on CT. Among these is that the attenuation of adjacent organs and tissues that may be very similar to the liver tissue itself. In addition, model-based approaches to segmentation are challenging due to the liver's widely varying shape (30). Also, automated segmentation of small liver tumors showed lower accuracy compared to both manual segmentation and automated segmentation of larger tumors (12, 24). As a result, there have been multiple attempts to develop methods of automated liver segmentation using CT (31, 32).

In our study, we found that A-vRECIST measurements highly correlated with both M-vRECIST and unidimensional mRECIST measurements in large HCCs from patients who had undergone TACE.

Because mRECIST is currently the preferred method of monitoring TACE therapy, we used it as the standard to which vRECIST was compared. Our results showed that, between our experienced radiologists, there was moderate to high inter-reader agreement for monitoring therapy using mRECIST measurements (ICC = 0.824). Our study also showed that correlation between unidimensional mRECIST and the vRECIST measurements was good $r = 0.766$ for M-vRECIST and $r = 0.774$ for A-vRECIST.

Results of our vRECIST measurements are likely to provide an early marker for TACE monitoring (23, 28, 32). We found that A-vRECIST measurements made using our neural network model could be a good substitute for M-vRECIST measurements and mRECIST (**Figure 5**). It also can improve the workflow as an alternate measure of response assessment because the measurements were highly correlated to each other in the baseline study, follow-up study, and volume change results ($r =$ 0.967, 0.937, and 0.826, respectively).

Our study had some limitations. First, the small cohort (42 patients) may have masked variability in the automated segmentation results. However, this study was a pilot, and we were aware from the outset that its findings would need

to be confirmed prospectively in a larger population with more variable tumor sizes and stages. Second, we did not examine correlation between A-vRECIST and patient's outcome. However, this study serves as a step for further evaluation of clinical importance of A-vRECIST and its relation to patient's survival endpoints. The small differences observed between A-vRECIST and M-vRECIST in the follow-up images (r = 0.648 vs. 0.748, respectively) may have been due to differences in TACE techniques, as the Lipiodol used in conventional TACE can distort CNNs, leading to differences in automated tumor segmentation.

Our next step is to confirm our findings with a larger sample size offering higher variability in tumor sizes and stages. We plan to use A-vRECIST results to classify patients according to their responses to TACE (partial response vs. no response). Also, we will thoroughly study the confounding factors, such as chronic parenchymal liver disease, that may affect the performance of neural networks.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The study was approved by the Institutional Review Board (IRB) at MD Anderson Cancer Center, office of protocol research. The informed consent was waived and approved by our institutional IRB.

AUTHOR CONTRIBUTIONS

AM contributed to the planning of the project, data collection, curation, data analysis, preparing figures, and write the first draft of the manuscript. KB helped in data curation, planning of the project, and reviewed the final draft of manuscript. AK contributed to the planning of the project, data collection, and shared in writing the first draft of the manuscript. DF and JH technical leaders of the project, construction of the neural network we used, planning of the project, and reviewing the first draft of the project. JS, AQ, and KE three radiologists who read the RECIST of the tumors, helped in planning the project, and also reviewed the final draft of the manuscript. KE provided supervision, support, conceptualization, and guidance throughout the project.

FUNDING

This article was submitted based on an invitation by the top editor Chunxiao Guo, we are invited to submit this in Frontiers in Oncology-Cancer Imaging and Image-directed Interventions under the special topic Novel Methods for Oncologic Imaging Analysis: Radiomics, Machine Learning, and Artificial Intelligence.

ACKNOWLEDGMENTS

I would like to acknowledge Scientific Publications Service, Research Medical Library, MD Anderson Cancer Center for their help in editing this article.

REFERENCES

1. Pesapane F, Nezami N, Patella F, Geschwind JF. New concepts in embolotherapy of HCC. *Med Oncol*. (2017) 34:58. doi: 10.1007/s12032-017-0917-2
2. Forner A, Hessheimer AJ, Isabel Real M, Bruix J. Treatment of hepatocellular carcinoma. *Crit Rev Oncol Hematol*. (2006) 60:89–98. doi: 10.1016/j.critrevonc.2006.06.001
3. Amit S, Jorge AM. Screening for hepatocellular carcinoma. *Gastroenterol Hepatol*. (2008) 4:201–8.
4. Facciorusso A, Serviddio G, Muscatiello N. Local ablative treatments for hepatocellular carcinoma: an updated review. *World J Gastrointest Pharmacol Ther*. (2016) 7:477–89. doi: 10.4292/wjgpt.v7.i4.477
5. Keating GM. Sorafenib: a review in hepatocellular carcinoma. *Target Oncol*. (2017) 12:243–53. doi: 10.1007/s11523-017-0484-7
6. Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*. (2018) 67:358–80. doi: 10.1002/hep.29086
7. Shah SR, Riordan SM, Karani J, Williams R. Tumour ablation and hepatic decompensation rates in multi-agent chemoembolization of hepatocellular carcinoma. *QJM*. (1998) 91:821–8. doi: 10.1093/qjmed/91.12.821
8. Raoul J-L, Forner A, Bolondi L, Cheung TT, Kloeckner R, de Baere T. Updated use of TACE for hepatocellular carcinoma treatment: how and when to use it based on clinical evidence. *Cancer Treat Rev*. (2019) 72:28–36. doi: 10.1016/j.ctrv.2018.11.002
9. Crocetti L, Bargellini I, Cioni R. Loco-regional treatment of HCC: current status. *Clin Radiol*. (2017) 72:626–35. doi: 10.1016/j.crad.2017.01.013
10. Prajapati HJ, Kim HS. Treatment algorithm based on the multivariate survival analyses in patients with advanced hepatocellular carcinoma treated with trans-arterial chemoembolization. *PLoS ONE*. (2017) 12:e0170750. doi: 10.1371/journal.pone.0170750
11. Han K, Kim JH. Transarterial chemoembolization in hepatocellular carcinoma treatment: barcelona clinic liver cancer staging system. *World J Gastroenterol*. (2015) 21:10327–35. doi: 10.3748/wjg.v21.i36.10327
12. Vorontsov E, Cerny M, Régnier P, Jorio LD, Pal CJ, Lapointe R, et al. Deep learning for automated segmentation of liver lesions at CT in patients with colorectal cancer liver metastases. *Radiol Artif Intell*. (2019) 1:180014. doi: 10.1148/ryai.2019180014
13. Gruber N, Antholzer S, Jaschke W, Kremser C, Haltmeier M. A joint deep learning approach for automated liver and tumor segmentation. *arXiv preprint arXiv:190207971*. (2019). doi: 10.1109/SampTA45681.2019.9030909
14. Li X, Chen H, Qi X, Dou Q, Fu C, Heng P. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans Med Imag*. (2018) 37:2663–74. doi: 10.1109/TMI.2018.2845918
15. Yu Q, Shi Y, Sun J, Gao Y, Zhu J, Dai Y. Crossbar-Net: a novel convolutional neural network for kidney tumor segmentation in CT images. *IEEE Trans Image Process*. (2019) 28:4060–74. doi: 10.1109/TIP.2019.2905537
16. Odland A, Server A, Saxhaug C, Breivik B, Groote R, Vardal J, et al. Volumetric glioma quantification: comparison of manual and semi-automatic tumor segmentation for the quantification of tumor growth. *Acta Radiol*. (2015) 56:1396–403. doi: 10.1177/0284185114554822
17. van Heeswijk MM, Lambregts DM, van Griethuysen JJ, Oei S, Rao S-X, de Graaff CA, et al. Automated and semiautomated segmentation of rectal tumor volumes on diffusion-weighted MRI: can it replace manual volumetry? *Int J Radiat Oncol Biol Phys*. (2016) 94:824–31. doi: 10.1016/j.ijrobp.2015.12.017
18. Soltaninejad M, Yang G, Lambrou T, Allinson N, Jones TL, Barrick TR, et al. Automated brain tumour detection and segmentation using superpixel-based

extremely randomized trees in FLAIR MRI. *Int J Comp Assist Radiol Surg*. (2017) 12:183–203. doi: 10.1007/s11548-016-1483-3
19. Kaus MR, Warfield SK, Nabavi A, Black PM, Jolesz FA, Kikinis R. Automated segmentation of MR images of brain tumors. *Radiology*. (2001) 218:586–91. doi: 10.1148/radiology.218.2.r01fe44586
20. Tu R, Xia L-P, Yu A-L, Wu L. Assessment of hepatic functional reserve by cirrhosis grading and liver volume measurement using CT. *World J Gastroenterol*. (2007) 13:3956–61. doi: 10.3748/wjg.v13.i29.3956
21. Karademir S. Staging of hepatocellular carcinoma. *HPB*. (2018) 7:35–41. doi: 10.20517/2394-5079.2018.40
22. Lencioni R, Petruzzi P, Crocetti L. Chemoembolization of hepatocellular carcinoma. *Semin Intervent Radiol*. (2013) 30:3–11. doi: 10.1055/s-0033-1333648
23. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015
24. Morshid A, Elsayes KM, Khalaf AM, Elmohr MM, Yu J, Kaseb AO, et al. A machine learning model to predict hepatocellular carcinoma response to transcatheter arterial chemoembolization. *Radiol Artif Intell*. (2019) 1:e180021. doi: 10.1148/ryai.2019180021
25. Welsh JL, Bodeker K, Fallon E, Bhatia SK, Buatti JM, Cullen JJ. Comparison of response evaluation criteria in solid tumors with volumetric measurements for estimation of tumor burden in pancreatic adenocarcinoma and hepatocellular carcinoma. *Am J Surg*. (2012) 204:580–5. doi: 10.1016/j.amjsurg.2012.07.007
26. Reynolds AR, Furlan A, Fetzer DT, Sasatomi E, Borhani AA, Heller MT, et al. Infiltrative hepatocellular carcinoma: what radiologists need to know. *Radiographics*. (2015) 35:371–86. doi: 10.1148/rg.352140114
27. Fuentes D, Ahmed K, Lin JS, Abdel-Wahab R, Kaseb AO, Hassan M, et al. Automated volumetric assessment of hepatocellular carcinoma response to sorafenib: a pilot study. *J Comp Assist Tomograp*. (2019) 43:499–506. doi: 10.1097/RCT.0000000000000866
28. Chapiro J, Duran R, Lin M, Schernthaner RE, Wang Z, Gorodetski B, et al. Identifying staging markers for hepatocellular carcinoma before transarterial chemoembolization: comparison of three-dimensional quantitative versus non–three-dimensional imaging markers. *Radiology*. (2014) 275:438–47. doi: 10.1148/radiol.14141180
29. Lin M, Pellerin O, Bhagat N, Rao PP, Loffroy R, Ardon R, et al. Quantitative and volumetric European association for the study of the liver and response evaluation criteria in solid tumors measurements: feasibility of a semiautomated software method to assess tumor response after transcatheter arterial chemoembolization. *J Vasc Interv Radiol*. (2012) 23:1629–37. doi: 10.1016/j.jvir.2012.08.028
30. Heimann T, Wolf I, Williams T, Meinzer HP. 3D active shape models using gradient descent optimization of description length. In: Christensen GE, Sonka M, editors. *Information Processing in Medical Imaging: Proceedings of the Conference*. Berlin: Springer (2005) 19:566–77. doi: 10.1007/11505730_47
31. Heimann T, Van Ginneken B, Styner MA, Arzhaeva Y, Aurich V, Bauer C, et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans Med Imag*. (2009) 28:1251–65. doi: 10.1109/TMI.2009.2013851
32. Lim S-J, Jeong Y-Y, Ho Y-S. Automatic liver segmentation for volume measurement in CT images. *J Visual Commun Image Rep*. (2006) 17:860–75. doi: 10.1016/j.jvcir.2005.07.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 07 May 2020
doi: 10.3389/fonc.2020.00551

Repeatability of Quantitative Imaging Features in Prostate Magnetic Resonance Imaging

Hong Lu[1,2], *Nestor A. Parra*[2], *Jin Qi*[2], *Kenneth Gage*[3], *Qian Li*[1], *Shuxuan Fan*[1,2], *Sebastian Feuerlein*[3], *Julio Pow-Sang*[4], *Robert Gillies*[2,3], *Jung W. Choi*[3]* and *Yoganand Balagurunathan*[3,4,5]*

[1] Department of Radiology, Tianjin Medical and Cancer Hospital, Tianjin, China, [2] Departments of Cancer Physiology, H. Lee Moffitt Cancer Center, Tampa, FL, United States, [3] Departments of Diagnostic Imaging, H. Lee Moffitt Cancer Center, Tampa, FL, United States, [4] Departments of Genitourinary Oncology, H. Lee Moffitt Cancer Center, Tampa, FL, United States, [5] Departments of Bioinformatics & Biostatistics, H. Lee Moffitt Cancer Center, Tampa, FL, United States

OPEN ACCESS

Edited by:
Lei Deng,
Jacobi Medical Center, United States

Reviewed by:
Naranamangalam Raghunathan Jagannathan,
Chettinad University, India
Satish E. Viswanath,
Case Western Reserve University, United States

***Correspondence:**
Jung W. Choi
Jung.Choi@moffitt.org
Yoganand Balagurunathan
Yoganand.Balagurunathan@moffitt.org

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 14 October 2019
Accepted: 27 March 2020
Published: 07 May 2020

Citation:
Lu H, Parra NA, Qi J, Gage K, Li Q, Fan S, Feuerlein S, Pow-Sang J, Gillies R, Choi JW and Balagurunathan Y (2020) Repeatability of Quantitative Imaging Features in Prostate Magnetic Resonance Imaging. Front. Oncol. 10:551. doi: 10.3389/fonc.2020.00551

Background: Multiparametric magnetic resonance imaging (mpMRI) has emerged as a non-invasive modality to diagnose and monitor prostate cancer. Quantitative metrics on the regions of abnormality have shown to be useful descriptors to discriminate clinically significant cancers. In this study, we evaluate the reproducibility of quantitative imaging features using repeated mpMRI on the same patients.

Methods: We retrospectively obtained the deidentified records of patients, who underwent two mpMRI scans within 2 weeks of the first baseline scan. The patient records were obtained as deidentified data (including imaging), obtained through the TCIA (The Cancer Imaging Archive) repository and analyzed in our institution with an institutional review board–approved Health Insurance Portability and Accountability Act–compliant retrospective study protocol. Indicated biopsied regions were used as a marker for our study radiologists to delineate the regions of interest. We extracted 307 quantitative features in each mpMRI modality [T2-weighted MR sequence image (T2w) and apparent diffusion coefficient (ADC) with b values of 0 and 1,400 mm/s^2] across the two sequential scans. Concordance correlation coefficients (CCCs) were computed on the features extracted from sequential scans. Redundant features were removed by computing the coefficient of determination (R^2) among them and replaced with a feature that had the highest dynamic range within intercorrelated groups.

Results: We have assessed the reproducibility of quantitative imaging features among sequential scans and found that habitat region characterization improves repeatability in ADC maps. There were 19 T2w features and two ADC features in radiologist drawn regions (native raw image), compared to 18 T2w and 15 ADC features in habitat regions (sphere), which were reproducible (CCC $\geq$0.65) and non-redundant ($R^2 \geq 0.99$). We also found that z-transformation of the images prior to feature extraction reduced the number of reproducible features with no detrimental effect.

Conclusion: We have shown that there are quantitative imaging features that are reproducible across sequential prostate mpMRI acquisition at a preset level of filters.

We also found that a habitat approach improves feature repeatability in ADC. A validated set of reproducible image features in mpMRI will allow us to develop clinically useful disease risk stratification, enabling the possibility of using imaging as a surrogate to invasive biopsies.

Keywords: radiomics, mpMRI, prostate cancer, test–retest in mpMRI, prostate TRUS-MRI, repeatable MRI features

INTRODUCTION

Prostate cancer detection using multiparametric magnetic resonance imaging (mpMRI) has been gaining consensus in the community for disease detection due to superior lesion sensitivity compared to transrectal ultrasound (TRUS) imaging (1, 2). Multiparametric MRI modalities have been useful in estimating size, volume, and relation to the underlying pathology of prostate cancer (3). Improvements in imaging technologies coupled with advances in mpMRI have led to its combined use with TRUS to guide prostate biopsies that improve detection of clinically aggressive cancers (4). Most clinical diagnoses follow a consensus reporting standard with the adoption of Prostate Imaging Reporting and Data Systems (PI-RADS v2) (5), which provides qualitative guidelines for clinical assessment. Variability in mpMRI scan interpretations among radiologists can in part be attributed to the steep learning curve required to interpret the scans (6). Quantitative imaging metrics or radiomics has been used to distinguish clinical abnormalities found in medical imaging (7, 8). For example, radiomics has been shown to be both reproducible in lung cancer computed tomography imaging and prognostic of lung cancer patient survival (9, 10). Recently, quantitative imaging features obtained from tumor regions on prostate mpMRI scans have been shown to be both predictive of clinically aggressive disease (11) and improve PI-RADS performance (12). In a recent survey on the role of imaging biomarkers in clinical decision making, the European Organization for Research and Treatment of Cancer and Cancer Research UK released a consensus statement with key recommendations to accelerate clinical biomarker translation (13). The key component of the consensus statement emphasizes the importance of validating the repeatability and reproducibility of these biomarkers for.

Information extraction (as part of a technical assay) and for proper downstream clinical utilization. Repeatability and reproducibility are necessary, but not sufficient, conditions for clinical usage of imaging biomarkers, as there is a higher relevance requirement such as accurate cancer prediction and prognosis (14, 15). As mpMRI has no biological reference for derived image intensity values, there are studies that have proposed standardizing these values (16–18). Recently, there have been efforts to find repeatable quantitative (radiomic) features in mpMRI scan of various cancers, such as rectal (19), cervix (20), lacrimal gland (21), and prostate (22, 23). Notably in prostate (23) and cervical studies (20), enrolled patients were scanned in a test–retest setting. Quantification of regions of interest has been accomplished in various ways, either through the use of a few open source tools (24) or more commonly through custom implementation methods. Recently, there has been an initiative to standardize definitions of these quantitative metrics, as recommended by the Image Biomarker Standardization Initiative (IBSI) (25). In our study, we obtained test–retest deidentified prostate mpMRI studies from patients enrolled at the Brigham and Women's Hospital, shared in a public repository (26). Patients with pathologically verified lesions were scored by a clinical pathologist (Gleason score). Independently marked regions of interest were standardized and quantified using custom radiomic features that followed the IBSI consensus criteria (25). We investigated the feasibility of reproducing these features across the cohort for a diverse set of prostate lesions. We also propose a habitat-based approach that converges regions of interest, followed by lesion characterization to improve repeatability of image features. This work will provide the basis for using repeatable quantitative features in prognostic evaluation of prostate cancer patients.

MATERIALS AND METHODS

We obtained deidentified mpMRI patient images along with segmentation masks (Dicom-Seg) through The Cancer Imaging Archive (TCIA) collection titled "QIN-PROSTATE-Repeatability" with detailed descriptions summarized about the cohort (26).

The patients were accrued for a research study at Brigham and Women's Hospital, Harvard Medical School. The patients waived informed consent, and their deidentified records were analyzed through our institutional review board–approved Health Insurance Portability and Accountability Act–compliant retrospective study protocol. The original study collection had 15 treatment-naive men who had mpMRI scans and biopsy-confirmed pathology and were scanned again within 2 weeks of their first baseline scan, during which patients did not receive any interim treatment. The cohort had 11 patients with a standard template biopsy and four patients who had suspicion of prostate cancer based on their clinical record. The mpMRI scans had T2w axial images [repetition time (TR) 3,350–5,109 ms, time to echo (TE) 84–107 ms, field of view (FOV) 140–200 nm] and ADC map derived from the diffusion-weighted MRI (b = 0, 1,400 s/mm, TR 2,500–8,250 ms, TE 76–80 ms, FOV 760–280mm). **Figure 1** illustrates sample lesions delineated on the test and retest T2w axial images.

Our study radiologist (H.L.) read the patient scans and localized lesions within the regions on patient scans identified by the prior study and provided consensus region segmentation in consultation with the second study radiologist (J.Q.). A third radiologist (K.G.) provided a random overread. Our radiologists in consensus agreed to use 13 of the 15 patient mpMRIs; scans

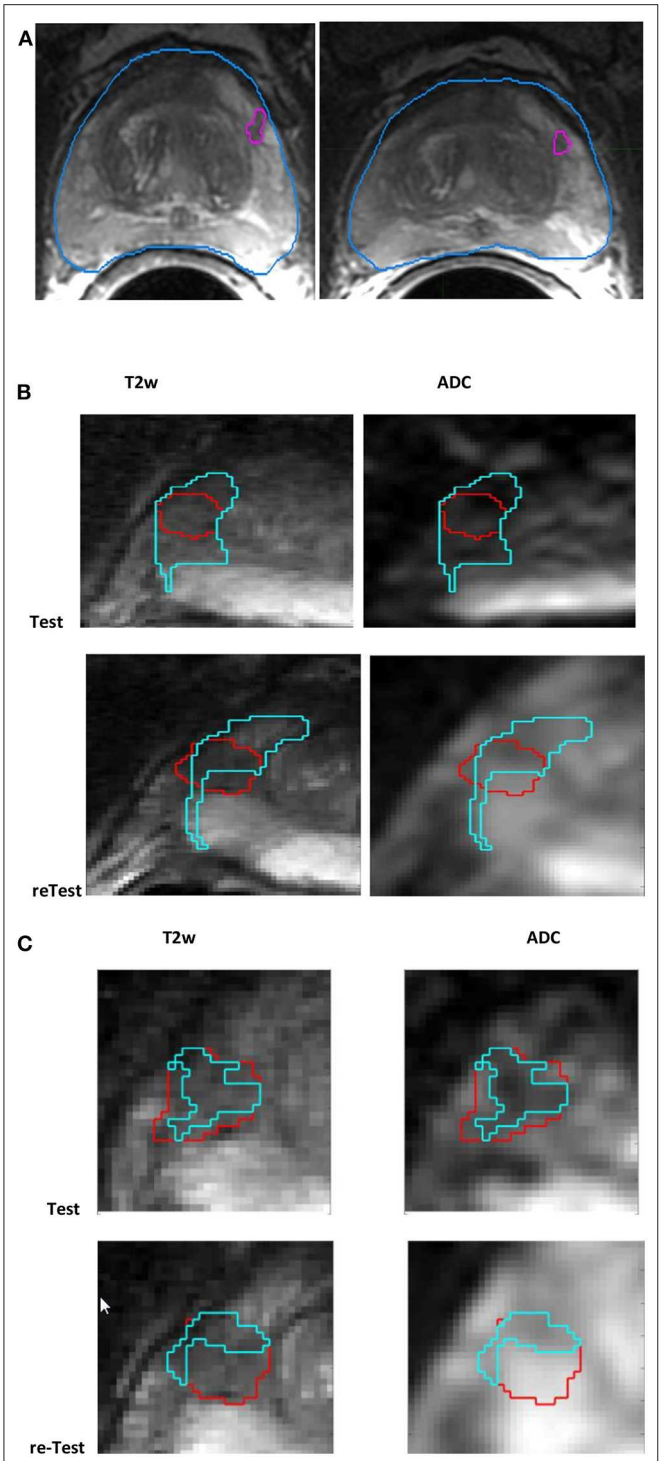


FIGURE 1 | Screen capture of mpMRI prostate scan **(A)** with radiologist-marked lesion shown for the baseline and follow-up scans in T2w, **(B)** habitat converged with a sphere (15mm), and **(C)** habitat (≤median) (in cyan) for a lesion (in red) shown for test and retest (along the rows) in T2w, ADC (along the column).

from two patients were dropped because of disagreements in identifying the abnormality and suboptimal quality of the scans. Among the converged patients, our radiologists in consensus identified 15 tumor lesions that were anatomically matched, longitudinally, across the test and retest time images. Of these, 11 identified lesion boundaries were verified to match with the prior study at the same anatomical location that had been pathologically verified (Gleason score); in addition, four additional lesion abnormalities were identified by our study radiologist(s), matched longitudinally. **Table 1** provides patient clinical details, including subject identifier, lesions anatomical location, prostate-specific antigen (PSA) value, and pathological diagnostic (Gleason) score. Newly identified lesions without corroborating pathological findings were marked as not available (NA), by our study radiologists.

Segmentation and Feature Extraction

Our study radiologists used MIM™ PACS [MIM Software Inc. (Cleveland, OH, USA)] to delineate regions of interest three-dimensionally (3D) on the prostate mpMRI scans, using T2w images as the reference sequence. Lesion boundaries were independently marked on the test and retest scans, whose cancer status was pathological identified by prior study. Four additional abnormalities that appeared radiologically malignant were identified and anatomically matched in sequential time points by our radiologists, but these lesions did not have pathological assessment. All lesion boundary segmentation was carried out as consensus reads by the study radiologists. Independent boundary delineation between lesions in the test and retest time point scan not only depicts the real clinical situation, but also introduces boundary variations, which can increase variability in the computed quantitative features.

Using the MIM libraries, T2w and ADC sequences were coregistered to avoid any motion artifacts in acquisition between the modalities. The registered multimodal image sequences were exported as 3D image matrices along with segmentation masks. We developed custom radiomic feature extraction tools, whose feature definition and formulation followed the IBSI consensus recommendations (25). We extracted 307 quantitative imaging features in the converged region of interest, which could be broadly categorized into three broad groups: C1: size and shape (45 features), C2: intensity, co-occurrence, run length (107 features), and C3: texture—laws and wavelets (155 features); see **Supplemental Tables S1–S3**.

Standardization of Image Regions

To assess the role of standardization procedures on feature stability in test–retest imaging, we propose to use conventional z-score standardization. We started by segmenting the prostate gland in 3D, and the region voxels were standardized by subtracting the mean and dividing by the deviation obtained at the gland level. The standardization was carried out independently for each modality (T2w and ADC) at a patient level. The lesion region of interest is standardized at the scan level and tend to have relative intensity with respect to the entire gland for a patient scan.

Habitat Image Region

We intend to find aggressive tumor-like regions in a marked lesion boundary of interest, which we call a *habitat* region. We define this region as one with restricted diffusion, whose

TABLE 1 | Summary of patient scans with clinical diagnosis for the biopsies.

#	Subject	MRI exam	PSA, ng/mL	Gleason(Bx)	Gland	PI-RADS v2 (test)	PI_RADS v2 (retest)	Comment
1	1-b1	Known Pca, staging	5.4	3 + 4	PZ	4	4	Identified
2	1-b2			NA	PZ	4	4	Additional
3	2-b1	Known Pca, assess change	7.5	3 + 4	PZ	2	3	Identified
4	2-b2			NA	PZ	2	3	Additional
5	3	Known Pca, staging	8.2	3 + 3	PZ	4	4	Identified
6	4	Known Pca, staging	4.3	3 + 3	PZ/TZ	2	2	Identified
7	5-b2	Suspected Pca, staging	NA	NA	TZ	2	2	Additional
8	6-b2	Elevated PSA, staging	5	Benign	TZ	1	1	Additional
9	7	Elevated PSA, staging	6.2	4 + 5	PZ	4	4	Identified
10	8	Known Pca, assess change	4.8	4.8	PZ	4	4	Identified
11	9	Elevated PSA, staging	9.4	Benign	PZ	4	4	Identified
12	10	Known Pca, assess change	3.15	3 + 3	PZ	4	4	Identified
13	11	Known Pca, assess change	9.7	3 + 3	PZ	4	4	Identified
14	12	Elevated Pca, staging	5.5	Benign	PZ	3	4	Identified
15	13	Known Pca, assess change	4.16	3 + 4	PZ	4	3	Identified
16	14	No biopsy performed	7	NA	NA	2	2	Ignored
17	15	Benign	9.5	Benign	NA	1	2	Ignored

NA, biopsies pathological score not available or cannot be determined; Identified, identified abnormality as stated; Additional, additional abnormality; Ignored, unable to locate abnormality.

characteristics resemble malignancy. We converge on a habitat region in two different ways: (a) sphere around lesion and (b) converge region within lesion. To find such a region, we first consider the entire lesion in 3D and on a colocalized volume across modalities. In the first case (a), we increase the search space to a 3D sphere with a fixed diameter of 15 mm and converge on a restricted diffusion region based on ADC values using a threshold defined by the distributional deviation (27, 28) and conforming regions to within the prostate gland structure. In the second case (b), we find the most contiguous lower median cutoff that is spanned in the ADC map within the radiologist-marked lesion region of interest. Converged habitat region will be mapped back to each modality of interest (T2w and ADC), and quantitative features are computed on a newly defined boundary. In the first case, it is possible to obtain a region larger than that marked by the radiologists. In the second case, the habitat region will always be contained within the marked lesion.

Concordant Features

Quantitative features that are reproducible in repeated experiments and can describe differential physiology are a necessary step for consideration as biomarkers. The feature values that are consistent between the test and retest experiment were evaluated. For each image feature, the concordance correlation coefficient (CCC) was computed to quantify reproducibility between the two scans for a patient across the cohort and independently computed in each modality (T2w, ADC). The CCC measures deviation from the diagonal line averaged over samples in the cohort and is commonly used to measure fidelity in repeated experiments (25). On this set of highly reproducible features, the next step was to select the features with a large interpatient variability, measured using the *dynamic range* (DR) metric. The normalized DR for a feature was defined by the inverse of the ratio of the average difference between measurements to the observed interpatient variability or biological range:

$$\mathrm{DR}=\left(1\text{-}{}^{1}\!/\!_{n}\ \sum_{i=1}^{n}\frac{\left|\mathrm{Test}(i)\text{-}\mathrm{Retest}(i)\right|}{\text{Max-Min}}\right) \tag{1}$$

where n is the total number of patient case; the DR varies from 0 to 1. Values close to 1 are preferred and imply that the feature has a large relative biological range, limited by the diversity in the cohort. As the variation between test–retest features increases, the DR values will show a reduction. Screening for a large DR will eliminate features that show greater variability in the repeat scans compared to the range of coverage. It is critical that a clinically relevant feature have a large DR to adequately distinguish the variations with tumor types, but show minimal variability in describing the same tumor type.

Redundancy Reduction

We propose to eliminate redundancies in features that are found to be reproducible. We computed the coefficient of determination (R^2) between the features that are considered to be reproducible, which measures the level of dependency between features. The R^2 has a range of 0 to 1 and is a ratio of the known variance as measured by linear model to the total variance between two variables or features, where one is the outcome, and the other is used to form the predictor. Values close to 1 would mean that the data points are close to the fitted line (i.e., closer to dependency) (24, 25). The coefficient of determination of simple regression is equal to the square of the Pearson correlation coefficient (29, 30). The features were grouped based on the R^2 values between them; in this subset, one representative was picked that had the highest

TABLE 2 | Distribution of quantitative imaging features at various levels of concordance and redundancy limits (Rsq at ≥0.99 and ≥0.95) for regions identified by **(A)** radiologist marked and **(B)** habitats converged (sphere), **(C)** habitat converged (≤median, ADC map).

(A). Radiologist marked

Concordance and dynamic range with redundancy reduction (Rsq ≥ 0.99) Test–retest mpMRI: number of features (radiologist)

CCC and DR and (Rsq ≥ 0.99)	ADC	ADCz	T2	T2z
≥0.95	0	0	0	0
≥0.90	0	0	3	0
≥0.85	0	0	4	1
≥0.80	0	0	5	3
≥0.75	0	0	9	6
≥0.70	1	3	13	10
≥0.65	2	3	19	12

(B). Habitat converged (Sphere)

Concordance and dynamic range with redundancy reduction (Rsq ≥0.99) Test–retest mpMRI: number of features (habitats)

CCC and DR and (Rsq ≥0.95)	ADC	ADCz	T2	T2z
≥0.95	0	0	0	1
≥0.90	0	0	1	2
≥0.85	1	0	3	6
≥0.80	4	3	6	11
≥0.75	6	5	8	14
≥0.70	9	8	13	18
≥0.65	15	15	18	23

(C). Habitat (≤median ADC map)

Concordance and dynamic range ≥ 0.65 and Rsq ≥0.99

CCC and DR	ADC	ADCz	T2	T2z
≥0.95	0	0	3	1
≥0.90	0	0	4	1
≥0.85	0	0	4	1
≥0.80	0	0	4	3
≥0.75	0	0	7	3
≥0.70	0	0	10	5
≥0.65	1	1	12	10

DR. The procedure was repeated recursively to cover all the features. We implemented different cutoff values for R^2 that assess linear dependence with any of the other features in the list. The purpose of this filter step is to eliminate redundancies, but not necessarily identify independence. The test–retest values were averaged before computing the R^2. We set different cutoff limits to reduce redundancy and combine features that are over the cutoff range. We repeated this process for a range of cutoffs (0.95–0.99), in our study.

RESULTS

As described in the *Materials and Methods*, the lesion was independently delineated in test and retest mpMRI scan,

TABLE 3 | Radiomic features that show concordance and non-redundancy in the test–retest cohort **(**CCC and DR ≥ 0.65;Rsq ≥ 0.99) for **(A)** radiologist-marked region, **(B)** habitat region (sphere, 15 mm), **(C)** habitat (≤median, ADC).

(A). Radiologist regions

Radiologist marked region (T2): CCC and DR >=0.65; Rsq ≥ 0.99.

T2 (raw): 19 features	T2 (*z*-normalized): 12 features
F138:GLSZM_Large-zone-low-gray-level-emphasis-F149:NGTDM_Contrast-F10:Stat-Max-gray-level	F19:Stat-Root-Mn-Sq-F149:NGTDM_Contrast-F10:Stat-Max-gray-level-F12:Stat-range-
F150:NGTDM_Busyness-F107:avgCooc_3D_Inv-diff-mom-norm	F107:avgCooc_3D_Inv-diff-mom-norm-
F284:3D-Wave-P1-L2-C4-F115:avgCooc_3D_Second-measure-of-information-correlation-	F171:3D-LawsF-L5-R5-R5-F284:3D-Wave-P1-L2-C4-
F113:avgCooc_3D_Cluster-prominence-F93:avgCooc_3D_Joint-var	F3:Stat-SD-F152:NGTDM_Strength-F8:Stat-10th-percentile-
F99:avgCooc_3D_Sum-var-F151:NGTDM_Complexity	
F302:3D-Wave-P1-L2-C13	
F300:3D-Wave-P1-L2-C12	
F27:Int-hist-90th-percentile	
F294:3D-Wave-P1-L2-C9	
F304:3D-Wave-P1-L2-C14	
F152:NGTDM_Strength	
Shape and Size	**Shape and size**
F47:Vol-at-Int-fraction-diff	F47:Vol-at-Int-fraction-diff-
F43:Vol-at-Int-Fraction-10-	F43:Vol-at-Int-Fraction-10-

Radiologist marked region (ADC): CCC **and DR ≥ 0.65; Rsq ≥ 0.99**

ADC (raw): two features	ADC (*z*-normalized): three features
F9:Stat-90th percentile	F96:avgCooc_3D_Difference-var
F8:Stat-10th percentile	F120:avg_3D_SRLGE-(Short-run-low-gray-level-emphasis)
	Shape and size
	F88:Center-of-mass-shift-(mm)

(B). Habitat regions

Habitat using sphere (ADC): CCC and DR ≥ 0.65; Rsq ≥ 0.99

ADC (raw): 15 features	ADC (*z*-score): 15 features
F228:3D-LawsF-R5-L5-L5-ADC-Auto	F302:3D-Wave-P1-L2-C13-ADC
F126:avg_3D_RLN-(Run-length-non-uniformity)-ADC	F294:3D-Wave-P1-L2-C9-ADC
F157:3D-LawsF-L5-L5-W5-ADC	F218:3D-LawsF-S5-R5-L5-ADC
F231:3D-LawsF-R5-L5-R5-ADC	F296:3D-Wave-P1-L2-C10-ADC
F246:3D-LawsF-R5-R5-R5-ADC	F304:3D-Wave-P1-L2-C14-ADC
F170:3D-LawsF-L5-R5-S5-ADC	F228:3D-LawsF-R5-L5-L5-ADC
F154:3D-LawsF-L5-L5-E5-ADC	F243:3D-LawsF-R5-R5-L5-ADC
F155:3D-LawsF-L5-L5-S5-ADC	F198:3D-LawsF-E5-W5-L5-ADC
F169:3D-LawsF-L5-R5-E5-ADC	F151:NGTDM_Complexity-ADC
F171:3D-LawsF-L5-R5-R5-ADC	F140:GLSZM_Gray-level-non-uniformity-ADC
F140:GLSZM_Gray-level-non-uniformity	F7:Stat-Min-gray-level-ADC
F124:avg_3D_GLN-(Gray-level-non-uniformity)	F115:avgCooc_3D_Second-measure-of-information-correlation-ADC
F148:NGTDM_Coarseness	F148:NGTDM_Coarseness-ADC

(Continued)

TABLE 3 | Continued

Shape and size:	**Shape and size:**
F90:Border-length-(mm)	F52:Vol-(mm^3)-ADC
F54:Surface-area-(mm^2)-ADC	F54:Surface-area-(mm^2)-ADC
Habitats using sphere (T2): CCC and DR >=0.65; Rsq ≥0.99	
T2 (raw): 18 features	**T2 (*z*-normalized): 23 features**
F117:avg_3D_LRE-(Long-runs-emphasis)-ADC	F115:avgCooc_3D_Second-measure-of-information-correlation-ADC
F126:avg_3D_RLN-(Run-length-non-uniformity)-ADC	F149:NGTDM_Contrast-ADC
F144:GLSZM_Zone-percentage-ADC	F29:Int-hist-mode-ADC
F27:Int-hist-90th-percentile-ADC	F19:Stat-Root-Mn-Sq-ADC
F100:avgCooc_3D_Sum-entropy-ADC	F159:3D-LawsF-L5-E5-E5-ADC
F161:3D-LawsF-L5-E5-R5-ADC	F122:avg_3D_LRLGE-(Long-run-low-gray-level-emphasis)-ADC
F141:GLSZM_Gray-level-non-uniformity-normalized-ADC-	F298:3D-Wave-P1-L2-C11-ADC
F124:avg_3D_GLN-(Gray-level-non-uniformity)-ADC	F306:3D-Wave-P1-L2-C15-ADC
F150:NGTDM_Busyness-ADC	F40:Max-hist-Gradient-gray-level-ADC
F243:3D-LawsF-R5-R5-L5-ADC	F42:Min-hist-Gradient-gray-level-ADC
F149:NGTDM_Contrast-ADC	F8:Stat-10th-percentile-ADC
F231:3D-LawsF-R5-L5-R5-ADC	F136:GLSZM_Small-zone-low-gray-level-emphasis-ADC
F300:3D-Wave-P1-L2-C12-ADC	F152:NGTDM_Strength-ADC
F140:GLSZM_Gray-level-non-uniformity-ADC	F139:GLSZM_Large-zone-high-gray-level-emphasis-ADC
F115:avgCooc_3D_Second-measure-of-information-correlation-ADC	F97:avgCooc_3D_Difference-entropy-ADC-
F114:avgCooc_3D_First-measure-of-information-correlation-ADC	F41:Min-hist-Gradient-ADC-
F148:NGTDM_Coarseness-ADC	F140:GLSZM_Gray-level-non-uniformity-ADC
	F231:3D-LawsF-R5-L5-R5-ADC
	F54:Surface-area-(mm^2)-ADC
	F114:avgCooc_3D_First-measure-of-information-correlation-ADC
	F148:NGTDM_Coarseness-ADC
Shape and Size:	**Shape and Size:**
F54:Surface-area-(mm^2)-	F88:Center-of-mass-shift-(mm)-ADC
	F87:Weighted-CoM_z-(mm)-ADC
(C). Habitat within lesion	
Habitat within lesion (≤ Median): ADC: CCC and DR ≥ 0.65; Rsq ≥ 0.99	
ADC (raw): 1 feature	**ADC (*z*-normalized): 1 feature**
F10:Stat-Max-gray-level-ADC	F143:GLSZM_Zone-size-non-uniformity-normalized-ADCz-Auto
Habitat within lesion (≤ Median): T2: CCC and DR ≥ 0.65; Rsq ≥0.99	
T2 (raw): 12 features	**T2 (*z*-normalized): 10 features**
F93:avgCooc_3D_Joint-var-ADC-T2	F3:Stat-SD-T2z
F151:NGTDM_Complexity-ADC-T2	F9:Stat-90th-percentile-T2z
F152:NGTDM_Strength-ADC-T2	F149:NGTDM_Contrast-T2z
F27:Int-hist-90th-percentile-ADC-T2	F93:avgCooc_3D_Joint-var-T2z
F131:avg_3D_RE-(Run-entropy)-ADC-T2	F27:Int-hist-90th-percentile-T2z
F113:avgCooc_3D_Cluster-prominence-ADC-T2	F113:avgCooc_3D_Cluster-prominence-T2z
F282:3D-Wave-P1-L2-C3-ADC-T2	F151:NGTDM_Complexity-T2z
F290:3D-Wave-P1-L2-C7-ADC-T2	F152:NGTDM_Strength-T2z
F149:NGTDM_Contrast-ADC-T2	F19:Stat-Root-Mn-Sq-T2z
F18:Stat-ENERGY-ADC-T2	F30:Int-hist-interquartile-range-T2z
F40:Max-hist-Gradient-gray-level-ADC-T2	
F99:avgCooc_3D_Sum-var-ADC-T2	

with each delineation done in consensus between the study radiologists. Using the lesion boundary as reference, the habitat region was converged automatically. We define habitat as a contiguous region colocalized to a low diffusion region defined by the ADC map. We then standardize the image voxels using *z*-score prior to any computations; in addition, we contrasted our findings with a non-standardized (raw) image region. In total, four image regions were investigated (raw-radiologist, *z*-score radiologist, raw-habitat, *z*-score habitat) by computing 307 quantitative image features in each of the regions, independently in test and retest images. We computed CCC to find repeatable image features, followed by application of a DR filter. Additionally, redundant features were removed based on coefficient of determination between the feature sets, repeated at different cutoffs. Distribution of features with different level settings in concordance correlation (CCC) and DR across the patient cohort is shown in **Table 2**. The imaging features that were extracted for respective modalities are listed in **Table 3**, obtained with $R^2 \geq 0.99$ (CCC and DR ≥0.65) and **Supplemental Tables 4,5**, obtained with $R^2 \geq 0.95$ (CCC and DR ≥0.65). **Figure 2** shows the distribution of CCC and DR for features extracted using different boundary regions; radiologist delineated (R), habitat converged (H), and habitat within the manually delineated region (H_{50}).

In our analysis, we find there are similar distributions of features between T2w- raw (native intensity values), radiologist-marked regions (19 features, CCC ≥0.65), and T2w-habitat with sphere regions (18 features, CCC ≥0.65), with standardized T2w *z*-score habitat (23 features, CCC ≥0.65) regions showing more stable features compared to T2w *z*-score raw regions. There were 12 stable features in T2w and 10 in T2w *z*-normalized regions, both evaluated at CCC ≥0.65 and with redundancy $R^2 \geq 0.99$ (**Tables 2B,C**, **3B,C**). Of the 19 features that are stable in T2w radiologist-marked regions, there are two volume features that measure within a certain intensity range and 17 others that are texture features.

In ADC map images, there were two features found within radiologist-marked regions compared to three features in ADC *z*-score regions, both evaluated at CCC ≥0.65 and with redundancy $R^2 \geq 0.99$. Using ADC-sphere–based habitats, we find the number of stable features increased to 15 in the ADC-habitat, seen in both radiologist-marked and *z*-scored normalized regions. While using habitat region within lesion approach, the new region was restricted to be within the lesion. We find there was one stable feature in ADC and ADC *z*-normalized region; in these regions, five and one feature were concordant, respectively (see **Supplemental Tables**). It seems *z*-score standardization moderately helps to improve the number of repeatable features in ADC maps.

Figures 3, **4** show the distribution of concordance coefficient and DR metric values, computed on features, respectively. They are grouped into the following broad categories: size and shape (C1), intensity and co-occurrence (C2), and laws and wavelets (C3). Texture features in the C2 intensity and co-occurrence category show higher concordance compared to other categories of features in T2w. The features computed in ADC

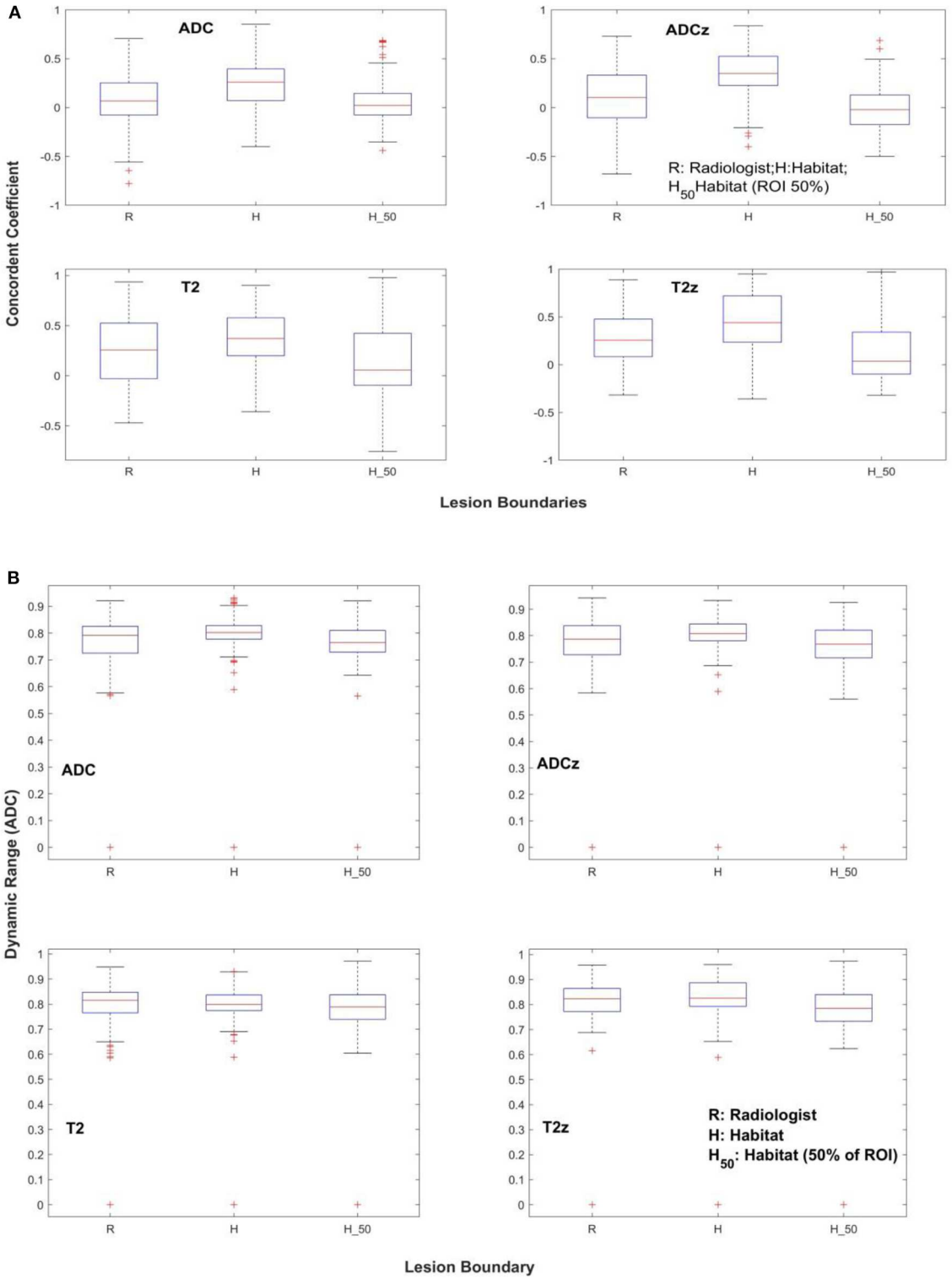


FIGURE 2 | Repeatability of quantitative features across different lesion boundaries. **(A)** Concordance coefficient, **(B)** dynamic range.

map do not show any consistent trend. It is also interesting to note that features in size and shape categories show lower concordance values.

While the ADC map shows intensity statistics to be reproducible, the *z*-score region shows reproducible co-occurrence matrix. The habitat region using sphere approach

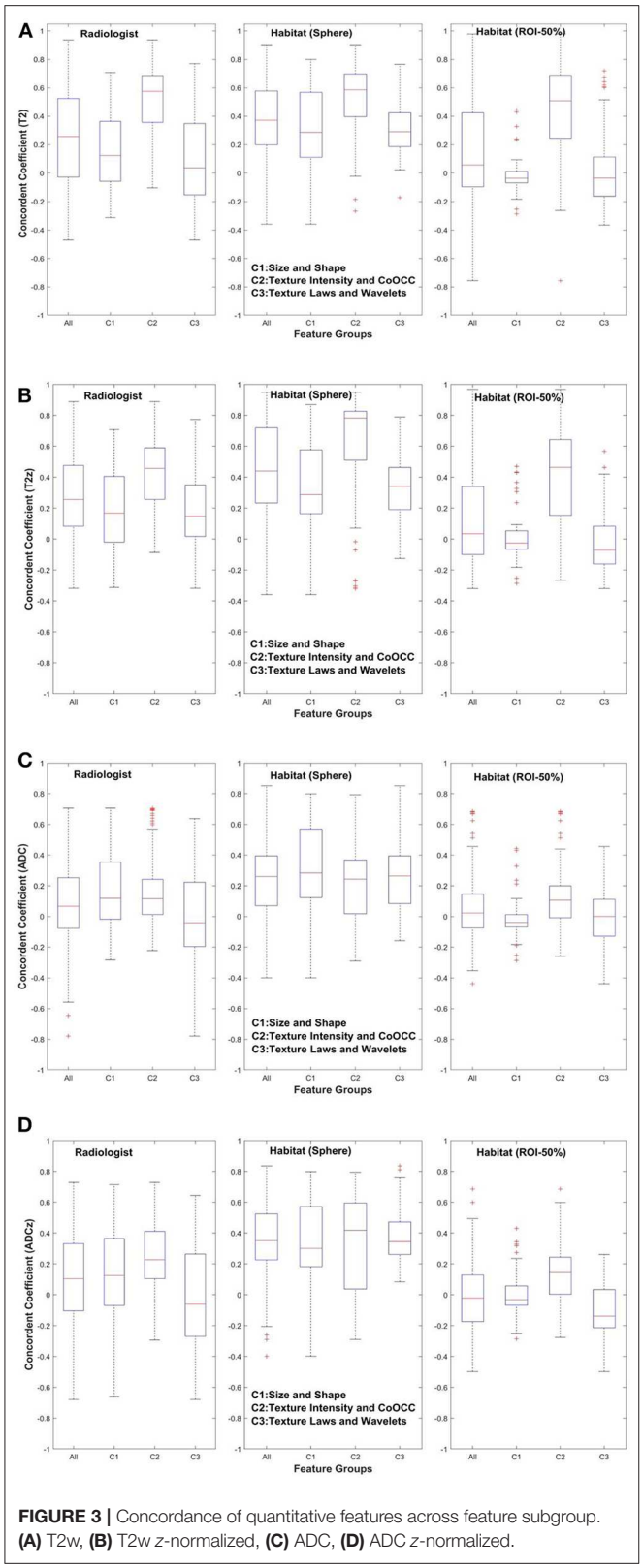


FIGURE 3 | Concordance of quantitative features across feature subgroup. **(A)** T2w, **(B)** T2w *z*-normalized, **(C)** ADC, **(D)** ADC *z*-normalized.

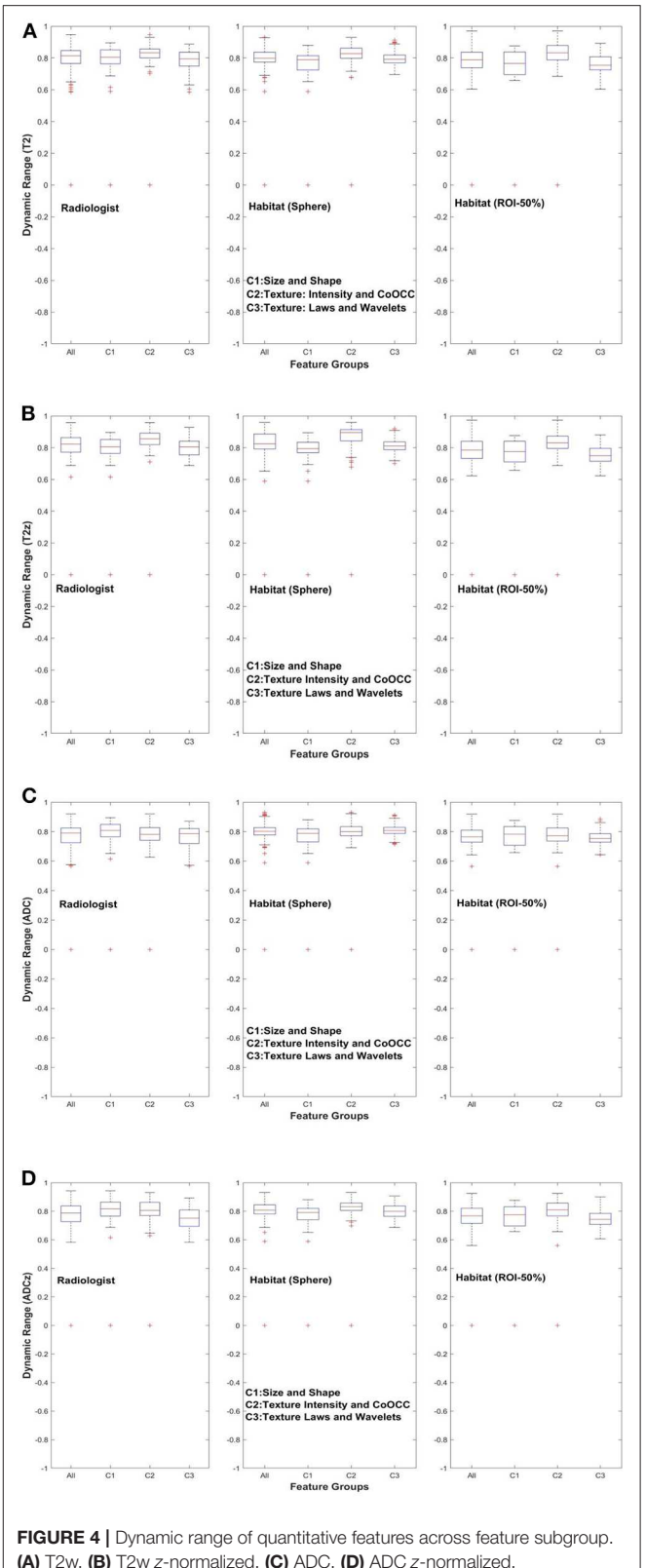


FIGURE 4 | Dynamic range of quantitative features across feature subgroup. **(A)** T2w, **(B)** T2w *z*-normalized, **(C)** ADC, **(D)** ADC *z*-normalized.

shows more than eight features related to fine texture (Laws) and two features related to shape category. While the T2-habitat (sphere) shows more features from co-occurrence, neighborhood gray tone difference categories. In region converged by habitat within lesions, ADC map shows one feature related to gray level that is stable and non-redundant. The T2-habitat (within lesion)

shows 12 features that are related to texture–neighborhood gray tone, co-occurrence, and wavelet based. The *z*-score standardization in T2-habitat (within lesion) region shows features related to gray-level intensities, co-occurrence, and neighborhood gray tone features that are reproducible and stable.

DISCUSSION

Clinically relevant imaging biomarkers are expected to be repeatable in a test–retest patient cohort, reproducible across centers, and relevant to describing the tumor physiology across different conditions. It is essential for imaging features to be used as a biomarker to be repeatable, at an acceptable level, which is dependent on the current imaging technology. In our study, we obtained prostate patient mpMRI scans within 2 weeks of the baseline time point and believe that the cohort is a unique public data set in prostate cancer. While we understand that the cohort size may be small for obtaining elaborate inferences, the methods applied by our study nonetheless allow us to assess feature stability and generate potential biomarkers in prostate mpMRI. We analyzed the repeatability in four different regions: (a) raw, radiologist-drawn; (b) *Z* score, radiologist drawn; (c) raw, habitats; (d) *z*-score, habitat regions. The study allowed us to contrast the reproducible features under these constraints.

The sphere-based habitat tends to increase the capture region that may provide a larger lesion boundary. This certainly helps to find stable and reproducible features in the ADC map and T2w image region. In comparison to habitat region formed within the lesion, it seems to restrict ADC intensity gray level that helps to find stable features in T2w, with more than 21 features with high concordance (CCC $\geq$0.75), of which 18 features are stable and reproducible (CCC $\geq$0.85, $R^2 > 0.099$).

We believe that the habitat approach reduces variability in T2w and rather highly variable ADC map images, which typically have lower resolution. There are a number of automated and semiautomated segmentation procedures that could be used in mpMRI, but we restricted our approach to manual, expert radiologist–drawn boundaries to initially delineate the lesions. We used the manual segmented region as an initial seed point for habitat region delineation, which is automatically converged using multimodal sequences (T2w, ADC).

In a prior study (23), they used an interclass correlation with a cutoff of 0.85 and reported features related to entropy, inverse difference moments to be highly repeatable. In our study, we find that co-occurrence and neighborhood gray tone difference matrices (NGTDMs) are two feature categories that are repeatable in T2w and T2w z. In ADC maps, the statistics of intensity-type features seem to show up as stable even in raw intensities (without any standardization), whereas average co-occurrence, short run length gray level emphasis-type features are stable in *z*-normalized ADC maps. We also find habitat (sphere) approach seems to improve the number of repeatable features in ADC maps and in T2w (**Figure 2**).

In the previously mentioned study, the authors claimed neither standardization nor prefiltering improved repeatability of image features. In our study, we used CCC with additional criteria such as DR and redundancy reduction to filter the features. We also find that most size and shape–based features show lower concordance in T2w/T2w *z*, but a larger spread on ADC map in comparison to two categories of features (**Figure 3**). This is probably due to the use of different regional convergence methods coupled with independently defined, delineated lesion boundaries in the test and retest scans. In comparison, the prior study (23) claimed high concordance for features in the size and shape–based category.

Because of scan quality limitations, some of the prior marked regions could not be ascertained by our radiologist, and additional regions of abnormality were located in consensus by the study radiologists. Additionally, prior studies (22, 23) restricted lesions to the peripheral zone, while our study radiologists identified lesions without any zonal restrictions. These differences have certainly increased the feature variability, which could be one cause for a lower number of repeatable features. Nevertheless, our cohort of patients provides a diverse set of lesions that are spread across the gland. The habitat approach proposed in the study shows promise in increasing the number of repeatable imaging features.

Study Limitations

This study provides a unique patient cohort with test–retest scans obtained within 2 weeks between scans; the cohort size is certainly a limited factor for a broader inference. The methodology used in the study with endorectal coils introduced artifacts that could have altered the voxel intensities and influenced the image feature reproducibility. We have taken effort to remove patient scans that show large artifacts and regions that could not be converged in a consensus read. Despite our efforts, there could be a certain level of variation in features value due to voxel level changes.

CONCLUSIONS

In the current study, we demonstrate that there are quantitative imaging features that can be obtained repeatedly in prostate mpMRI. We show that sublocalized regions or habitats can improve repeatability of imaging features, possibly by restricting the range of variations in the voxel intensity levels in these MRI scan modalities. We also find that *z*-score normalization of the image intensities had minimal effect on the feature reproducibility. Current findings allow us to obtain reproducible and non-redundant sets of image features that could be used for predictive and prognostic purpose.

DATA AVAILABILITY STATEMENT

Dataset used in this study can be accessed using following URL: https://wiki.cancerimagingarchive.net/display/Public/QIN-PROSTATE-Repeatability and additional analysis, study information for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by USF IRB. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

HL and YB: Hypothesis, methods development. HL, JQ, QL, KG, and SFa: lesions identification and marking. YB, NP, and HL: results inference, implementation of methods. YB, HL, and JC: manuscript writing. HL, YB, RG, JP-S, JC, JQ, QL, KG, SFa, and SFe: manuscript/results proof read and approval.

FUNDING

We acknowledged research grant that support this work NIH/NCI 1R01CA190105-01, U01-CA200464 and Cohen's Family Donation 2018-19 for Moffit's Prostate Cancer research Program.

ACKNOWLEDGMENTS

We would like to extend our gratitude to Dr. Federov and his team for reviewing the lesion markings in the shared patient data records. We would like to thank the Moffitt Radiomics team and IRAT core (Dr. Abdalah and his team) for the invaluable discussion and for extending support to this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00551/full#supplementary-material

REFERENCES

1. Ahmed HU, El-Shater Bosaily A, Brown LC, Gabe R, Kaplan R, Parmar MK, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet*. (2017) 389:815–22. doi: 10.1016/S0140-6736(16)32401-1
2. Brown LC, Ahmed HU, Faria R, El-Shater Bosaily A, Gabe R, Kaplan RS, et al. Multiparametric MRI to improve detection of prostate cancer compared with transrectal ultrasound-guided prostate biopsy alone: the PROMIS study. *Health Technol Assess*. (2018) 22:1–176. doi: 10.3310/hta22390
3. Kwak JT, Sankineni S, Xu S, Turkbey B, Choyke PL, Pinto PA, et al. Prostate cancer: a correlative study of multiparametric mr imaging and digital histopathology. *Radiology*. (2017) 285:147–56. doi: 10.1148/radiol.2017160906
4. Siddiqui MM, Rais-Bahrami S, Turkbey B, George AK, Rothwax J, Shakir N, et al. Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer. *Jama*. (2015) 313:390–7. doi: 10.1001/jama.2014.17942
5. Radiology ACo. *Prostate Imaging Reporting and Data System (PIRADS) version 2*. (2015). Available online at: https://www.acr.org/-/media/ACR/Files/RADS/Pi-RADS/PIRADS-V2.pdf (accessed January 10, 2020).
6. Rosenkrantz AB, Ginocchio LA, Cornfeld D, Froemming AT, Gupta RT, Turkbey B, et al. Interobserver reproducibility of the PI-RADS Version 2 lexicon: a multicenter study of six experienced prostate radiologists. *Radiology*. (2016) 280:152542. doi: 10.1148/radiol.2016152542
7. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
8. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
9. Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep*. (2016) 6:23428. doi: 10.1038/srep23428
10. Balagurunathan Y, Gu Y, Wang H, Kumar V, Grove O, Hawkins S, et al. Reproducibility and prognosis of quantitative features extracted from CT images. *Trans Oncol*. (2014) 7:72–87. doi: 10.1593/tlo.13844
11. Algohary A, Viswanath S, Shiradkar R, Ghose S, Pahwa S, Moses D, et al. Radiomic features on MRI enable risk categorization of prostate cancer patients on active surveillance: preliminary findings. *J Magn Reson Imaging*. (2018) 48:818–28. doi: 10.1002/jmri.25983
12. Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang YD. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol*. (2017) 27:4082–90. doi: 10.1007/s00330-017-4800-5
13. O'Connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol*. (2017) 14:169–86. doi: 10.1038/nrclinonc.2016.162
14. Liu Y, deSouza NM, Shankar LK, Kauczor H-U, Trattnig S, Collette S, et al. A risk management approach for imaging biomarker-driven clinical trials in oncology. *Lancet Oncol*. (2015) 16:e622–8. doi: 10.1016/S1470-2045(15)00164-3
15. European Society of Radiology (ESR). White paper on imaging biomarkers. *Insights Imaging*. (2010) 1:42–5. doi: 10.1007/s13244-010-0025-8
16. Nyul L, Udupa. J. On standardizing the mr image intensity scale. *Magn Reson Imaging*. (1999) 42:1072–81. doi: 10.1002/(SICI)1522-2594(199912)42:6<1072::AID-MRM11>3.0.CO;2-M
17. De Nunzio G, Cataldo R, Carla A. Robust intensity standardization in brain magnetic resonance images. *J Digital Imaging*. (2015) 28:727–37. doi: 10.1007/s10278-015-9782-8
18. Robitaille N, Mouiha A, Crépeault B, Valdivia F, Duchesne S, The Alzheimer's Disease Neuroimaging I. Tissue-based MRI intensity standardization: application to multicentric datasets. *Int J Biomed Imaging*. (2012) 2012:347120. doi: 10.1155/2012/347120
19. Traverso A, Kazmierski M, Shi Z, Kalendralis P, Welch M, Nissen HD, et al. Stability of radiomic features of apparent diffusion coefficient (ADC) maps for locally advanced rectal cancer in response to image pre-processing. *Physica Medica*. (2019) 61:44–51. doi: 10.1016/j.ejmp.2019.04.009
20. Fiset S, Welch ML, Weiss J, Pintilie M, Conway JL, Milosevic M, et al. Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. *Radiother Oncol*. (2019) 135:107–14. doi: 10.1016/j.radonc.2019.03.001
21. Duron L, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik J-C, et al. Gray-level discretization impacts reproducible MRI radiomics texture features. *PLoS ONE*. (2019) 14:e0213459. doi: 10.1371/journal.pone.0213459
22. Fedorov A, Vangel MG, Tempany CM, Fennessy FM. Multiparametric magnetic resonance imaging of the prostate: repeatability of volume and apparent diffusion coefficient quantification. *Invest Radiol*. (2017) 52:538–46. doi: 10.1097/RLI.0000000000000382
23. Schwier M, van Griethuysen J, Vangel MG, Pieper S, Peled S, Tempany C, et al. Repeatability of multiparametric prostate mri radiomics features. *Sci Rep*. (2019) 9:9441. doi: 10.1038/s41598-019-45766-z
24. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77:e104–e7. doi: 10.1158/0008-5472.CAN-17-0339

25. IBSI. *Image Biomarker Standardisation Initiative*. (2019). Available online at: https://arxiv.org/abs/1612.07003 (accessed January 10, 2020).
26. Fedorov A, Schwier M, Clunie D, Herz C, Pieper S, Kikinis R, et al. An annotated test-retest collection of prostate multiparametric MRI. *Sci Data.* (2018) 5:180281. doi: 10.1038/sdata.2018.281
27. Otsu N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics.* (1979) 9:62–6. doi: 10.1109/TSMC.1979.4310076
28. Sezgin M, Sankur B. Survey over image thresholding techniques and quantitative performance evaluation. *J Electron Imaging.* (2004) 13:146-65 doi: 10.1117/1.1631315
29. RGD Steel JT. *Principles and Procedures of Statistics*. New York, NY: McGraw-Hill. (1960)
30. Cameron AC, Windmeijer FAG. An *R*-sqaure measured of goodness of fit for some common nonlinear regression models. *J Econom.* (1997) 77:1790–2. doi: 10.1016/S0304-4076(96)01818-0

Conflict of Interest: HL, QL, and SFa received research scholarship to partial support their salary from Tianjin Medical and Cancer Hospital, Tianjin, China during their tenured research time at Moffitt Cancer Center. RG is an investor and consultant in Health Myne.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 11 May 2020
doi: 10.3389/fonc.2020.00604

Radiomics-Based Preoperative Prediction of Lymph Node Status Following Neoadjuvant Therapy in Locally Advanced Rectal Cancer

Xuezhi Zhou[1,2†], Yongju Yi[3,4†], Zhenyu Liu[2,5†], Zhiyang Zhou[6†], Bingjia Lai[7], Kai Sun[1], Longfei Li[8], Liyu Huang[1], Yanqiu Feng[3], Wuteng Cao[6*] and Jie Tian[1,2,5,9*]*

[1] *Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University, Xi'an, China,* [2] *CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Science, Beijing, China,* [3] *Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering, Southern Medical University, Guangzhou, China,* [4] *Network Information Center, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China,* [5] *University of Chinese Academy of Science, Beijing, China,* [6] *Department of Radiology, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China,* [7] *Department of Radiology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China,* [8] *Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou, China,* [9] *Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing, China*

Background and Purpose: Lymph node status is a key factor for the recommendation of organ preservation for patients with locally advanced rectal cancer (LARC) following neoadjuvant therapy but generally confirmed post-operation. This study aimed to preoperatively predict the lymph node status following neoadjuvant therapy using multiparametric magnetic resonance imaging (MRI)-based radiomic signature.

Materials and Methods: A total of 391 patients with LARC who underwent neoadjuvant therapy and TME were included, of which 261 and 130 patients were allocated to the primary cohort and the validation cohort, respectively. The tumor area, as determined by preoperative MRI, underwent radiomics analysis to build a radiomic signature related to lymph node status. Two radiologists reassessed the lymph node status on MRI. The radiomic signature and restaging results were included in a multivariate analysis to build a combined model for predicting the lymph node status. Stratified analyses were performed to test the predictive ability of the combined model in patients with post-therapeutic MRI T1-2 or T3-4 tumors, respectively.

Results: The combined model was built in the primary cohort, and predicted lymph node metastasis (LNM+) with an area under the curve of 0.818 and a negative predictive value (NPV) of 93.7% were considered in the validation cohort. Stratified analyses indicated that the combined model could predict LNM+ with a NPV of 100 and 87.8% in the post-therapeutic MRI T1-2 and T3-4 subgroups, respectively.

Conclusion: This study reveals the potential of radiomics as a predictor of lymph node status for patients with LARC following neoadjuvant therapy, especially for those with post-therapeutic MRI T1-2 tumors.

Keywords: lymph node metastasis, prediction, neoadjuvant therapy, locally advanced rectal cancer, radiomics

OPEN ACCESS

Edited by:
Xuelei Ma,
Sichuan University, China

Reviewed by:
Satish E. Viswanath,
Case Western Reserve University, United States
Takashi Akiyoshi,
Cancer Institute Hospital of Japanese Foundation for Cancer Research, Japan

****Correspondence:***
Yanqiu Feng
foree@163.com
Wuteng Cao
caowteng@163.com
Jie Tian
jie.tian@ia.ac.cn

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *28 October 2019*
Accepted: *02 April 2020*
Published: *11 May 2020*

Citation:
Zhou X, Yi Y, Liu Z, Zhou Z, Lai B, Sun K, Li L, Huang L, Feng Y, Cao W and Tian J (2020) Radiomics-Based Preoperative Prediction of Lymph Node Status Following Neoadjuvant Therapy in Locally Advanced Rectal Cancer. Front. Oncol. 10:604. doi: 10.3389/fonc.2020.00604

INTRODUCTION

Neoadjuvant therapy followed by total mesorectal excision (TME) is the standard treatment for patients with locally advanced rectal cancer (LARC) (1). After neoadjuvant therapy, ∼50–60% of patients are downstaged, and ∼20% show pathologic complete response (1–3). Although TME is effective at providing local tumor control, it is also associated with significant genitourinary and gastrointestinal morbidity and long-lasting complications such as sexual dysfunction and urinary or fecal problems (4–6). Hence, organ preservation strategies, such as watchful waiting and local excision (7) following neoadjuvant therapy, are becoming more popular for preserving organ function and improving the patients' quality of life (8–12).

One of the disadvantages of organ preservation is a lack of exact pathologic lymph node staging. Leaving lymph node metastasis (LNM+) unresected can potentially lead to local recurrence or distant spread. Magnetic resonance imaging (MRI) and computer tomography are the routine imaging modalities for restaging following neoadjuvant therapy for rectal cancer, but with limited accuracy and with no consensus regarding the standard definitions of LNM+ (13). Neoadjuvant therapy results in changes in shape, size, and texture of a positive lymph node, but these changes still cannot exactly indicate a positive node turning out to be negative. The remains of tumor cells in small nodes make nodal restaging a challenge, which makes patients to have to undergo TME to obtain the precise pathological nodal stage (14). Several studies have investigated the predictive factors for LNM+ but have not identified measures with sufficient predictive precision to enable clinical decisions. For example, a nomogram based on preoperatively available clinicopathologic features has been created to predict LNM+ following neoadjuvant treatment for LARC. If the threshold of 0.3 nomogram predicting the risk of positive nodes is used, almost 80% of the patients with LNM+ will be correctly identified (15). Azizian et al. found that changes of circulating miR-18b and miR-20a expression levels during neoadjuvant treatment could predict LNM+ with a NPV of 79 and 85%, respectively (16). A recent study reported that two factors (ypT stage <3 and lymphovascular invasion) were associated with ypN0 status in good responders following neoadjuvant therapy, indicating a high positive predictive value (PPV) for identifying ypN0 patients (17). However, this study had a small sample size and lacked validation, and the predictive factors were derived from resection specimens; this precluded desirable preoperative decision-making.

Radiomics is a rapid developing field of quantitative image analysis that may facilitate the prediction of lymph node status following neoadjuvant therapy (18, 19). The utility of radiomics is evident from clinical research, such as the prediction of therapeutic responses (20–23), survival analysis (24, 25), and prediction of clinical events (26, 27). Recently, two studies (28, 29) have attempted to detect the associations between local tumor region information on imaging and surrounding nodals and demonstrated the potential of preoperative tumor radiomic features in predicting LNM+ in rectal cancer; however, their analyses were limited to patients that were not administered with any preoperative treatment. Therefore, we hypothesize that local tumor region information following neoadjuvant therapy may also associate with regional nodal status.

Radiomics could quantitatively analyze image information, which may help to detect some associations between local tumor information on imaging and surrounding nodal status. This study aimed to assess if preoperative MRI-based radiomic features could reliably predict lymph node status following neoadjuvant therapy in LARC to improve patient management. Briefly, we first attempted to construct a multiparametric MRI-based radiomic signature. Then, we built and validated a prediction model incorporating the radiomic signature and radiologist's assessment results. Finally, we evaluated the prediction model's performance in two subgroups with different post-therapeutic MRI T (ymrT) stages to identify the ideal population in which this model would be applicable.

Abbreviations: ADC, apparent diffusion coefficient; AUC, area under curve; CI, confidence interval; LARC, locally advanced rectal cancer; LNM+, lymph node metastasis; LNM-, lymph node non-metastasis; LoG, Laplacian of Gaussian; MRI, magnetic resonance imaging; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic; TME, total mesorectal excision

MATERIALS AND METHODS

Patients

This retrospective study was approved by the institutional review board of the Sixth Affiliated Hospital of Sun Yat-sen University. The requirement for informed patient consent was waived. A total of 425 patients who were initially diagnosed with N+ or T3/T4 rectal cancer, also named as LARC, and received neoadjuvant therapy followed by TME surgery between November 2012 and May 2017 at the Sixth Affiliated Hospital of Sun Yat-sen University were included. The exclusion criteria were as follows: (i) lack of multiparametric MRI data including T1-weighted fast spin-echo imaging (T1w), T2 weighted fast spin-echo imaging (T2w), diffusion-weighted imaging (DWI), or contrast-enhanced T1-weighted fast spin-echo imaging (CE-T1w) 1 week before TME surgery; (ii) insufficient MRI quality due to bowel peristalsis-related artifacts; (iii) lack of clinical information including sex, age, and carcinoembryonic antigen (CEA) (cutoff: $\geq$ 5 ng/ml, $<$ 5 ng/ml) blood level; and (iv) lack of pathology reports, since the pathological lymph nodal status will be obtained from the pathology reports. The recruitment of patients is depicted in **Figure 1**. Patients were then randomly allocated to a primary cohort and a validation cohort in a ratio of 2:1.

Multiparametric MRI Acquisition

All patients were scanned with a 1.5-Tesla MR (Optima MR 360, GE Medical Systems, USA) using an eight-element body array coil with fixed image protocols. The scanning sequences consisted of T1w, T2w, DWI (two b-values including 0 and 800 s/mm^2), and CE-T1w. The technical MRI parameters are listed in **Supplementary Table A2**.

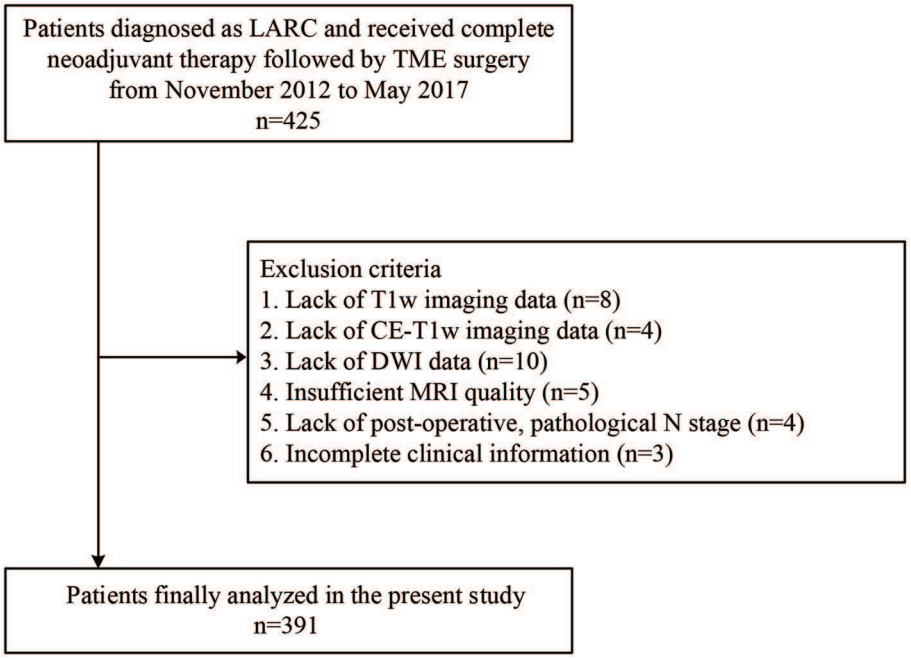


FIGURE 1 | Recruitment pathway for patients in this study. LARC, locally advanced rectal cancer; T1w, T1-weighted; CE-T1w, contrast-enhanced T1-weighted; DWI, diffusion-weighted imaging; MRI, magnetic resonance imaging.

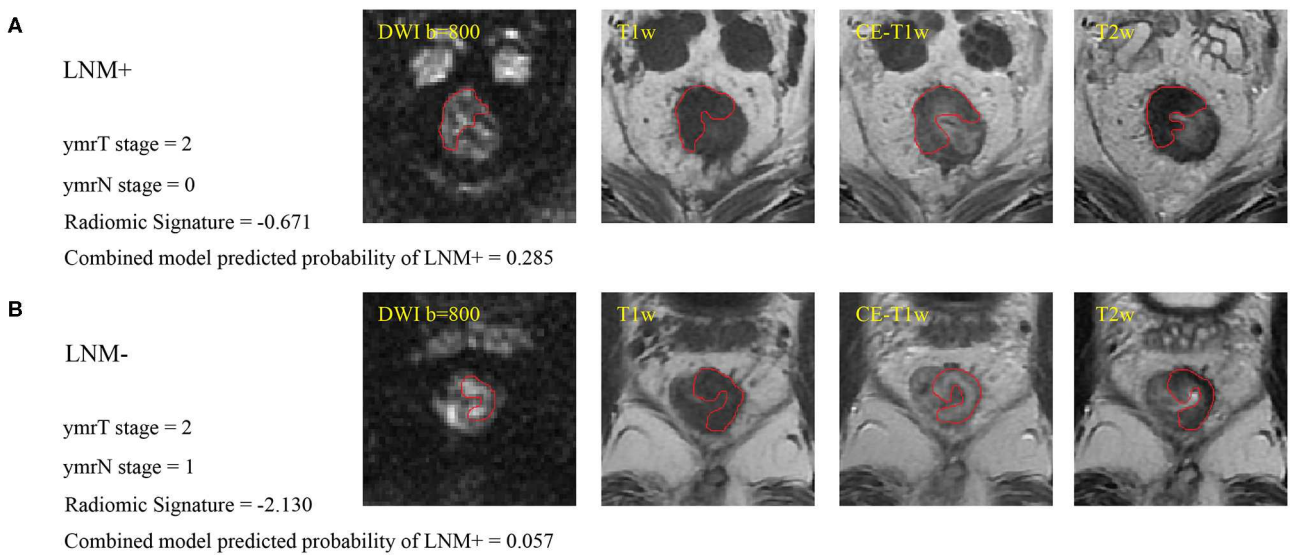


FIGURE 2 | (A) Sample images of LNM+, where the red line indicates the tumor margin. **(B)** Sample images of LNM–. Cutoffs for the radiomic signature and combined model are −1.4208 and 0.0897, respectively, for patients with ymrT1-2 tumors in this study. These two patients were misdiagnosed by the radiologist but were correctly assessed by radiomics analysis. LNM+, lymph node metastasis; LNM–, lymph node non-metastasis; DWI, diffusion-weighted imaging; T1w, T1-weighted; CE-T1w, contrast-enhanced T1-weighted; T2w, T2-weighted.

Tumor Masking and Radiomic Feature Extraction

Two gastrointestinal radiologists with 5 (radiologist #1) and 10 (radiologist #2) years of experience examined the MR images and independently defined the regions of interest by manually outlining the tumor margin using itk-SNAP software (www.itksnap.org) on axial slices containing the largest cross-sectional tumor area on each imaging sequence, as shown in **Figure 2**. At an intuitive level, the most reasonable way to predict the lymph node status is to perform radiomic analysis on each

node. However, doing so in this retrospective study is almost impossible as it needs to know every node's pathological status and needs to map every lymph node tissue on MRIs. In this study, we could only obtain the patient-level lymph node status from the post-operative pathology report, which was the number of positive nodes and all nodes from the resection specimens. There even existed some small nodes that could be identified under the microscope but are missed on MRIs following neoadjuvant therapy. In addition, one problem must be solved if we perform radiomic analysis on identifiable nodes on MRIs. The number of identifiable nodes can vary a lot between different patients. That means that we will obtain feature sets with different feature numbers between different patients. Transforming these feature sets into the same feature space is difficult to solve. Thus, we defined a local tumor area as a region of interest like most of the published study (28, 30, 31). Local tumor region information following neoadjuvant therapy may also associate with regional nodal status. Intraclass correlation coefficients were used to assess the agreement of extracted features by two radiologists. The regions of interest on DWI were delineated at a *b* value of 800 s/mm^2 and were then copied onto the corresponding apparent diffusion coefficient (ADC) maps.

The radiomic features extracted are listed in **Supplementary Table A3**. A total of 264 features were extracted from each of the T1w, T2w, and CE-T1w images and the ADC maps. These features could be divided into three categories, including first-order statistics, textural features, and Laplacian of Gaussian (LoG) filtration features. Radiomic feature extraction was conducted using an in-house software written in MATLAB (MathWorks, Inc., Natick, MA, USA). All features were linearly normalized into a range [0, 1] with the formula as follows:

$$X_i^{norm} = (X_i - X_i^{\min})/(X_i^{\max} - X_i^{\min}) \quad (1)$$

where X_i^{norm} was the *ith* normalized feature value, X_i was the *ith* raw feature value, and $X_i^{\min}$ and $X_i^{\max}$ were the minimum value and maximum value of the *ith* raw feature values in the primary cohort, respectively.

Feature Selection and Radiomic Signature Construction

We built a model for predicting LNM+ in the primary cohort and evaluated its generalizability in the validation cohort. Before modeling, a feature selection program consisting of three steps was executed in the primary cohort. First, the Wilcoxon rank-sum test was performed for every feature between the LNM+ and LNM- groups as a rough identification of features with $p \leq 0.1$ to be used in further processing. Second, the Spearman correlation coefficient was calculated between any two features, and the feature with the bigger Wilcoxon rank-sum test *p*-value was excluded when the absolute value of the correlation coefficient exceeded 0.9. Third, the least absolute shrinkage and selection operator (LASSO) method was applied to select the most predictive features (32). To avoid over-fitting, the best LASSO regularization parameter "lambda" was determined by a 10-fold cross-validation. Features with one standard error from the minimum criterion were selected for modeling. Then, a multivariate logistic regression model was built based on the selected features. Summation of the selected features multiplied by the corresponding coefficients was performed for each patient as a radiomic signature, which was mathematically represented as follows:

$$\text{radiomic signature} = \sum_{i=1}^{n} C_i^* X_i + b \quad (2)$$

$$Y = 1/\left(1 + \exp\left(-\left(\sum_{i=1}^{n} C_i^* X_i + b\right)\right)\right) \quad (3)$$

where Y was the probability of LNM+ predicted by this model, b was the intercept, X_i was the *ith* selected feature, and C_i was the coefficient of the *i*th selected feature. Receiver operating characteristic (ROC) curve analysis was performed in both cohorts to evaluate the predictive ability of radiomic signatures in differentiating LNM+ from LNM-. All steps were performed with R version 3.5.2 (www.r-project.org) using the "glmnet," "glm2," and "pROC" packages.

Comparison of Radiomic Signature and Radiologists' Diagnostic Performance

Radiologists #1 and #2, who were blinded to any medical record information, independently reviewed the MRIs and independently determined the post-therapeutic ymrT stage and post-therapeutic MRI N (ymrN) stage. The ymrT stage was based on the depth of tumor penetration (mucin or soft components) relative to the muscularis propria as T1 (limited to the mucosa and submucosa), T2 (invasion but no penetration of the muscularis propria), T3 (penetration beyond the muscularis propria), or T4 (involvement of other organs). The ymrN status was defined as positive metastasis if the regional lymph node manifested with a small diameter ($\geq$ 6 mm), irregular border, mixed signal intensity (SI), or high SI assumed to represent mucin. The N stage was based on the number of positive lymph nodes: N1 (at least one but less than three nodes) or N2 (more than or equal to three nodes). If the smallest diameter of the largest lymph node was <6 mm and had no features of irregular border and no mixed SI was observed, the N status was graded as N0 (33). Mcnemar test (34) and net reclassification improvement (NRI) test (35) were used for statistical analysis of the prediction results of the radiomic signature and radiologists' diagnosis. Univariate logistic regression analysis was performed in the primary cohort to select the clinical variables with a significant association. Finally, we established a combined model incorporating the radiomic signature and the associated clinical variables by multivariate logistic regression and evaluated this model in the validation cohort. A clinical model incorporating associated clinical variables without radiomic signature was also built through multivariate logistic regression for comparison purposes. To provide an easily used quantitative tool to predict the probability of LNM+, we converted the combined model to a nomogram. The calibration curves were plotted to assess the consistency between the predicted probability and the actual rate

of LNM+. Hosmer-Lemeshow test with p-value > 0.05 indicates a good fit of the model (36). Decision curve analysis was also conducted to assess the clinical use of this nomogram.

Unlike patients with ymrT3-4 tumors, patients with ymrT1-2 tumors usually exhibit a lower probability of LNM+ and a smaller depth of invasion (37); thus, they are more suitable candidates for local excision. The predictive ability of the model may differ in subgroups divided according to ymrT stage. Thus, we conducted stratified analyses in ymrT1-2 and ymrT3-4 groups, respectively.

Area under the curve (AUC), accuracy, sensitivity, specificity, PPV, and negative predictive value (NPV) according to the Youden cutoff (38) were calculated to quantize the predictive ability of the prediction models in both cohorts.

RESULTS

Demographic and Clinical Data

A total of 391 patients were enrolled in the study, as described in **Figure 1**; 231 of these patients underwent preoperative treatment with four to six cycles of mFOLFOX6 chemotherapy (infusional fluorouracil plus oxaliplatin of 85 mg/m^2 intravenously on day 1 of each chemotherapy cycle). Postoperative adjuvant chemotherapy was performed with seven cycles of mFOLFOX6; the rest of the 160 patients received preoperative treatment with five cycles of infusional fluorouracil (leucovorin 400 mg/m^2 intravenously followed by fluorouracil 400 mg/m^2 intravenously and fluorouracil 2.4 g/m^2 by 48-h continuous intravenous infusion) and concurrent radiation treatment. Radiotherapy was delivered at 1.8 to 2.0 Gy daily from Monday through Friday for a total of 23 to 28 fractions over 5 to 6 weeks and a total dose of 46.0 to 50.4 Gy. Radiation was delivered with a minimum energy of 6-MV photons through a three- or four-field box technique to the primary tumor and to mesorectal, presacral, and internal iliac lymph nodes (39). A post-operative pathological examination indicated that 87 patients were LNM+. The number of positive nodes ranges from 1 to 12, with a median number of 2. The other 304 patients were LNM-. The clinical characteristics of the patients enrolled are summarized in **Table 1** and in **Supplementary Table A1**. There were no significant

TABLE 1 | Clinical characteristics of patients in primary and validation cohorts.

Characteristic	Primary cohort ($n = 261$)		p	Validation cohort ($n = 130$)		p
	LNM+ ($n = 58$)	LNM- ($n = 203$)		LNM+ ($n = 29$)	LNM- ($n = 101$)	
Age, years	50.24 ± 11.76	54.61 ± 12.69	0.203	53.17 ± 12.95	53.88.06 ± 11.36	0.775
cT stage, *n* (%)			0.520			0.247
T2	2 (4)	14 (7)		0 (0)	9 (9)	
T3	42 (72)	149 (73)		24 (83)	75 (74)	
T4	14 (24)	40 (20)		5 (17)	17 (17)	
cN stage, *n* (%)			**0.003**			0.100
N0	6 (11)	42 (21)		5 (17)	30 (30)	
N1	17 (29)	88 (43)		10 (35)	43 (43)	
N2	35 (60)	73 (36)		14 (48)	28 (27)	
Concurrent radiation, *n* (%)			0.759			0.533
Yes	21 (36)	78 (38)		12 (41)	49 (48)	
No	37 (64)	125 (62)		17 (59)	52 (52)	
Sex, *n* (%)			0.717			0.293
Male	42 (72)	142 (70)		23 (79)	70 (69)	
Female	16 (28)	61 (30)		6 (21)	31 (31)	
CEA, *n* (%)			0.317			0.247
Positive	14 (24)	37 (18)		8 (28)	18 (18)	
Negative	44 (76)	166 (82)		21 (72)	83 (82)	
ymrT stage, *n* (%)			**0.018**			**0.032**
T1	1 (2)	25 (12)		0 (0)	7 (7)	
T2	11 (19)	55 (27)		3 (10)	32 (32)	
T3	36 (62)	105 (52)		23 (80)	57 (56)	
T4	10 (17)	18 (9)		3 (10)	5 (5)	
ymrN stage, *n* (%)			**<0.001**			**0.006**
N0	24 (41)	142 (70)		14 (49)	73 (72)	
N1	19 (33)	51 (25)		10 (34)	25 (25)	
N2	15 (26)	10 (5)		5 (17)	3 (3)	

Age is presented as mean ± standard deviation. The p-value for age was calculated using independent samples t-test analysis. The p-values for the categorical variables were calculated using Pearson's chi-square test analysis. ymrT stage and ymrN stage were restaged by radiologist #2 who has 10 years of experience. Bold font indicates $p < 0.05$. LNM+, lymph node metastasis; LNM–, lymph node non-metastasis; CEA, carcinoembryonic antigen; ymr, restaging MRI assessments.

differences in the clinical variables between the primary and the validation cohorts. **Table 2** exhibited the agreement of ymrT/N stage and ypT/N stage. The ymrT could predict ypT stage with an accuracy of 88.2%. The major predicted error was derived from overstaging of ypT0–2. However, in terms of node restaging, ymrN and ypN showed bad concordance.

Radiomic Signature Construction

The intraclass correlation coefficients calculated for features extracted by the two radiologists ranged from 0.725 to 0.942, reflecting a good agreement. The features extracted from the regions of interest delineated by the radiologist with 10 years of experience were used for further analysis. Thirteen features were selected to build a radiomic signature, as listed in **Supplementary Table A4**. None of the T1w feature was selected,

TABLE 2 | Confusion matrice for tumor restaging and node restaging.

	ypT0-2	ypT3-4		yN0	yN+
ymrT1–2	120	14	ymrN0	215	38
ymrT3–4	32	225	ymrN+	89	49

The ymrT/N stages were assessed by radiologist #2 who has 10 years of experience.

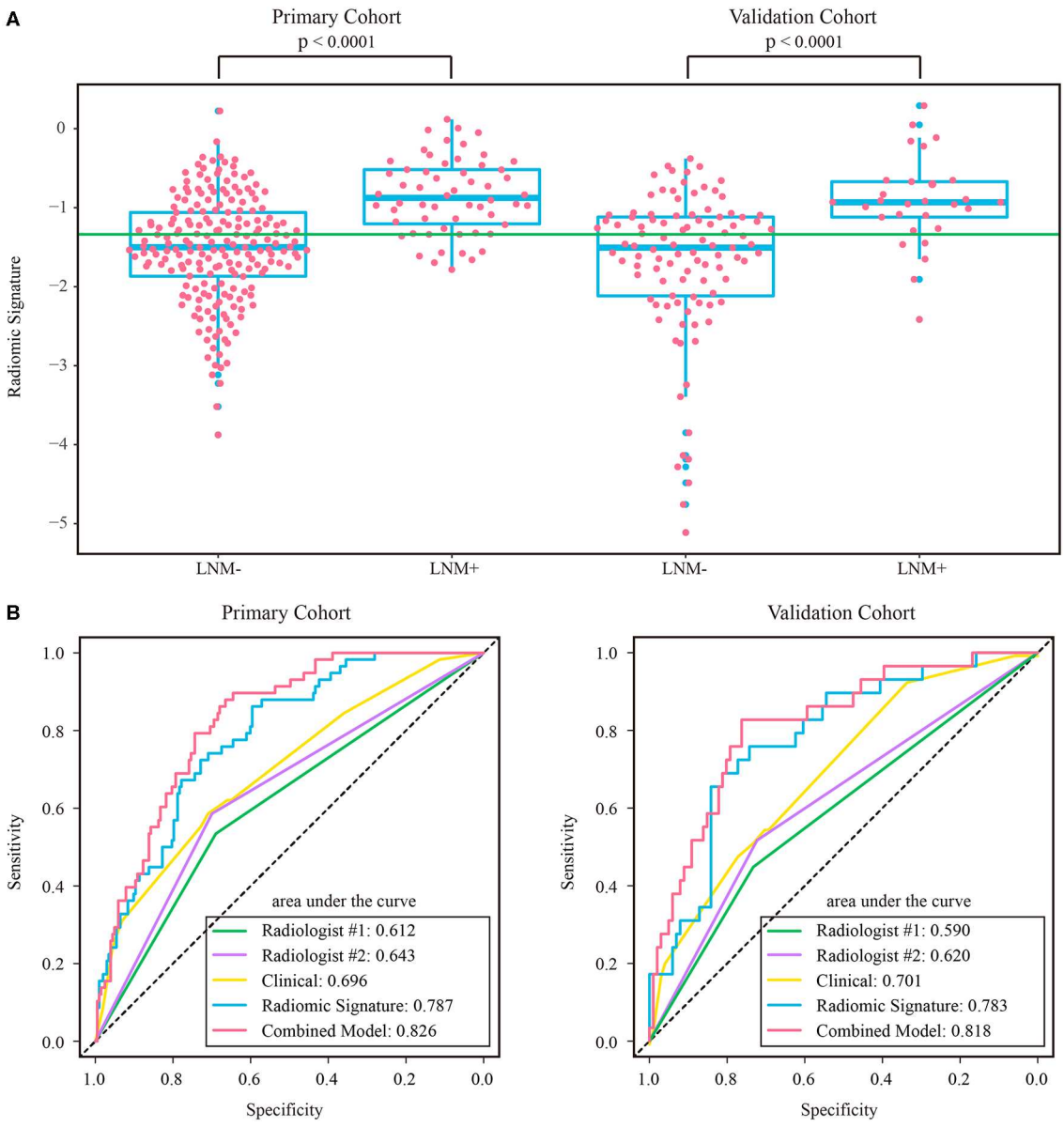


FIGURE 3 | (A) Distribution of radiomic signature in the primary and validation cohorts, where the green line indicates the Youden cutoff in the primary cohort, and the *p*-value was calculated using Wilcoxon rank-sum test. **(B)** ROC curves of radiologists' and prediction models. LNM+, lymph node metastasis; LNM-, lymph node non-metastasis; ROC, receiver operating characteristic.

indicating a poor predictive ability of T1w features. In both cohorts, the radiomic signature was significantly higher in the LNM+ group than in the LNM- group, as shown in **Figure 3**. The radiomic signature yielded an AUC of 0.787 [95% confidence interval (CI): 0.726–0.848] and 0.783 (95% CI: 0.690–0.875) in the primary and validation cohorts, respectively.

Comparison of Radiomic Signature and Radiologists' Diagnostic Performance

The assessment results of the two radiologists were highly consistent, yielding a Kappa value of 0.936 and 0.933 for ymrT stage and ymrN stage, respectively. The confusion matrices, as shown in **Supplementary Figure A1**, indicated that radiologist #1 and radiologist #2 yielded a sensitivity of 50.57% (95% CI: 35.6–64.7%) and 56.32% (95% CI: 42.3–70.8%), respectively. The AUC of radiologist #2 was 0.62 (95% CI: 0.518–0.722) in the validation cohort, which was significantly (Delong test *p*-value: 0.021) smaller than that of the radiomic signature. The sensitivity of the radiomic signature reached a score of 82.8% (95% CI: 68.8–96.6%), which was significantly (Mcnemar test *p*-value: 0.022) different from that of radiologist #2 in the validation cohort. The specificity values of radiomic signature and radiologist #2 were 58.4% (95% CI: 48.8–67.7%) vs. 72.2% (95% CI: 63.5–81.0%), which were also significantly different (Mcnemar test *p*-value: 0.044).

As the diagnostic accuracy of radiologist #2 was higher than that of radiologist #1, here we only reported the prediction results based on post-therapeutic restaging results from radiologist #2, and those based on the restaging results from radiologist #1 were provided in the **Supplementary File**. In univariate logistic regression analysis in the primary cohort, post-therapeutic ymrT stage, ymrN stage, and radiomic signature were statistically significant (**Table 3**). We built a combined model to integrate the staging results of radiologist #2 and the radiomic signature using multivariate logistic regression in the primary cohort and converted it into a nomogram, as shown in **Figure 4**. Compared to the radiologists' performance, the prediction accuracy in the validation cohort using the combined model was improved (NRI test *p*-value: 0.125) to 75.4% from the accuracy value of 63.8% of radiologist #2, yielding a sensitivity of 82.8% (95% CI: 68.5–82.8%), specificity of 73.3% (95% CI: 64.8–81.9%), PPV of 47.1% (33.2–60.8%), and NPV of 93.7% (88.3–99.0%). The clinical model incorporating ymrT and ymrN yielded an AUC value of 0.696 (95% CI: 0.619–0.773) and 0.701 (95% CI: 0.601–0.801) in the primary cohort and the validation cohort, respectively. The Delong test analysis showed that the clinical model performed significantly ($p < 0.05$) worse than the combined model but was comparable to the radiomic signature ($p > 0.05$) in both cohorts. All these results are listed in **Supplementary Table A5** and **Figure 2A**.

The stratified analyses indicated that radiologist #2 yielded a better prediction in the ymrT1-2 subgroup than that in the ymrT3-4 subgroup with NPV of 90.7 vs. 80%. The combined model also performed better in the ymrT1-2 subgroup of the validation cohort with an AUC of 0.915 and a NPV of 100%. In the ymrT3-4 subgroup of the validation cohort, the combined model yielded an AUC of 0.764 and a NPV of 87.8% according to the Youden cutoff. Detailed results are shown in **Figures 5**, **6** and in **Supplementary Tables A6, A7**. For comparison, the combined model based on radiomic signature and restaging results from radiologist #1 yielded a NPV of 100 and 86.7% in ymrT1–2 subgroup and ymrT3-4 subgroup, respectively (**Supplementary Table A8**).

DISCUSSION

In this study, a major finding was that radiomics is a promising approach for the preoperative prediction of LNM+ following neoadjuvant therapy in patients with LARC. Radiomic signature was a powerful predictor independent of the radiologists' diagnostic results, offering a NPV of 92.2% in the validation cohort. Combining radiomic signature with the radiologists' diagnostic results improved the NPV to 93.7%. In the post-therapeutic ymrT1-2 subgroup, the combined model yielded a NPV of 100% and specificity of 59%. However, in the post-therapeutic ymrT3-4 subgroup, the combined model did not achieve 100% NPV.

TABLE 3 | Univariate and multivariate logistic regression analysis for clinical characteristics and radiomic signature.

Parameter	Univariate			Multivariate			
	p	OR	95% CI	*p*	Coefficient	OR	95% CI
Sex	0.7169	0.89	0.46–1.69	–	–	–	–
Age	0.2029	0.78	0.53–1.14	–	–	–	–
CEA	0.3183	1.43	0.71–2.87	–	–	–	–
Concurrent radiation	0.7590	0.91	0.49–1.66	–	–	–	–
ymrT stage	**0.0020**	1.93	1.27–2.93	0.4210	0.2000	1.22	0.75–1.99
ymrN stage	**<0.0001**	7.72	3.31–18.02	**0.0063**	0.6963	4.03	1.48–10.94
Radiomic signature	**<0.0001**	6.31	3.45–11.55	**<0.0001**	1.7705	5.15	2.78–9.55
Intercept	–	–	–	0.8685	−0.1320	–	–

The p-values were calculated using Wald test analysis. Bold font indicates p < 0.05. OR, odds ratio; CI, confidence interval; CEA, carcinoembryonic antigen; ymr, restaging MRI assessments.

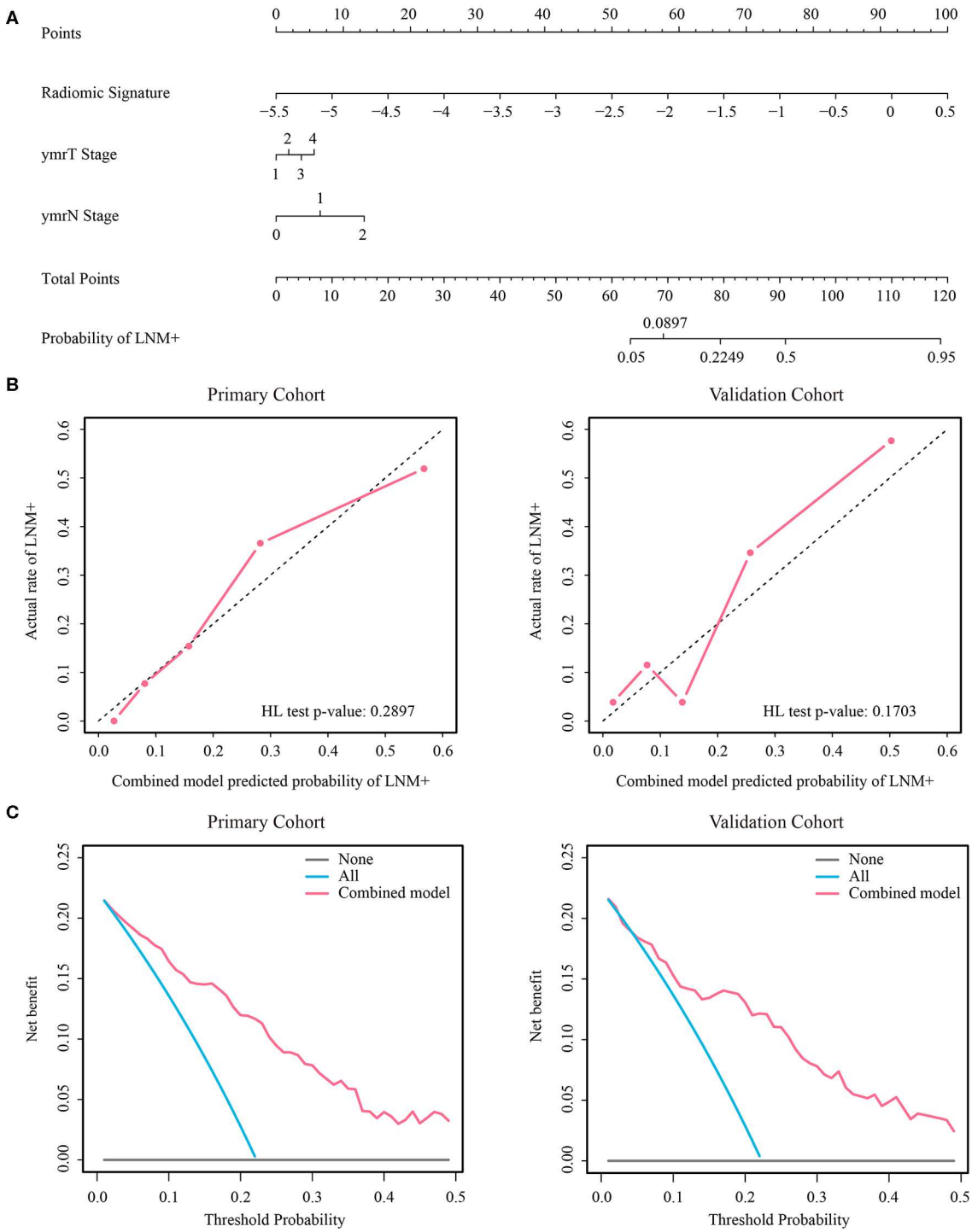


FIGURE 4 | (A) Nomogram of the combined model. **(B)** Calibration curves of the combined model in both cohorts. **(C)** Decision curve analysis of the combined model in both cohorts. LNM+, lymph node metastasis; HL, Hosmer-Lemeshow.

A major factor limiting the clinical application of organ preservation strategies is that the precise assessment of lymph node status is challenging (14, 40) since the completeness of tumor resection can be determined by pathological examination, but residual LNM+ has a high risk of leading to an adverse prognosis. Although the size and the morphological features (i.e., round shape, irregular border, and heterogeneous texture) have been proposed to define a clinically positive lymph node

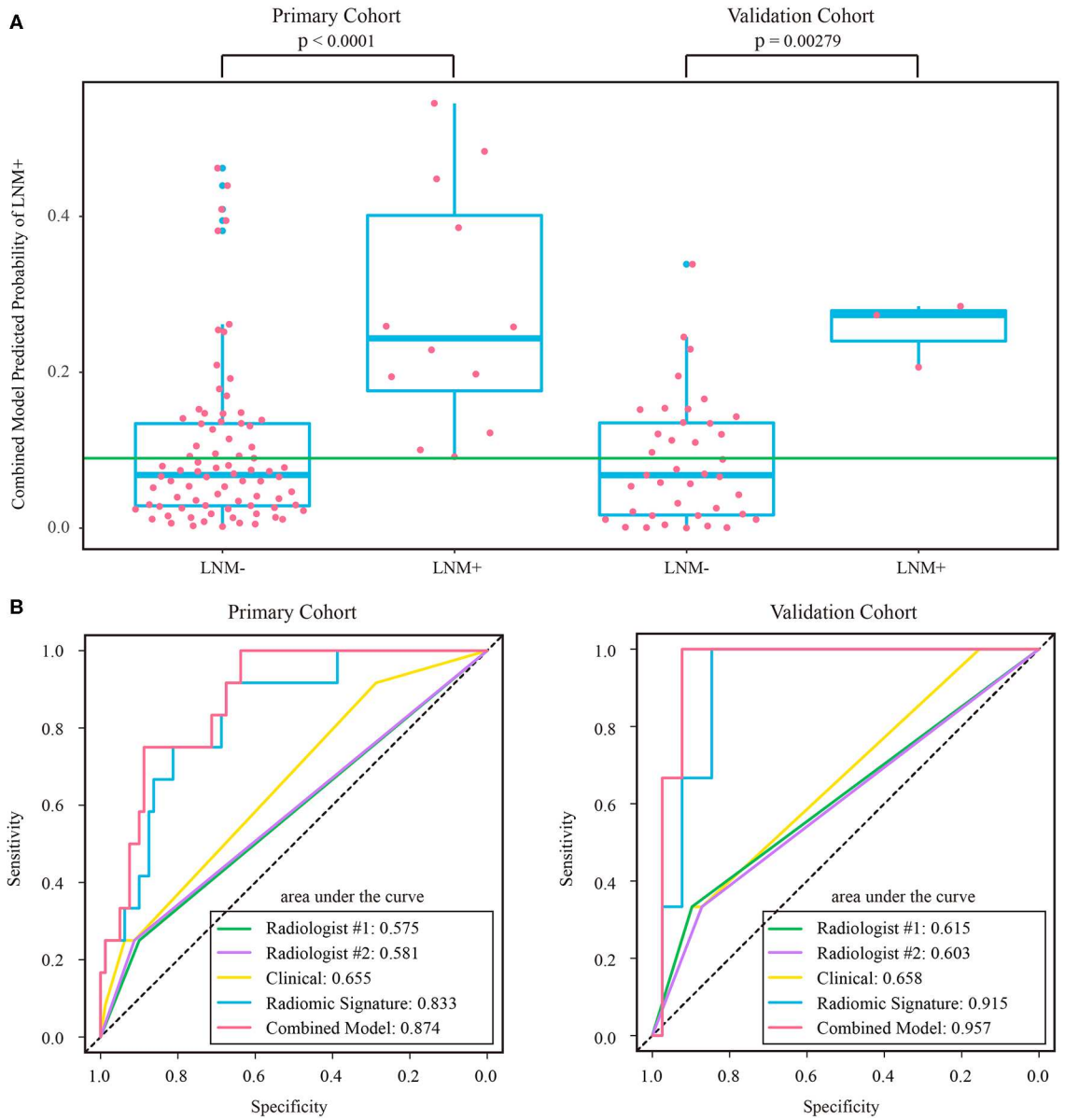


FIGURE 5 | (A) Distribution of combined model predicted probability of LNM+ in post-therapeutic ymrT1-2 subgroups of both cohorts, where the green line indicates the Youden cutoff in the primary cohort. **(B)** ROC curves of radiologists' and prediction models in post-therapeutic ymrT1-2 subgroups of both cohorts. LNM+, lymph node metastasis; LNM-, lymph node non-metastasis; ROC, receiver operating characteristic.

on MRI, the correspondence between post-therapeutic cN+ and pN+ is still poor. Recently, a large retrospective study from the Netherlands revealed that using post-therapeutic cN+ to predict pN+ yielded a sensitivity of 56%, specificity of 67%, PPV of 47%, and NPV of 75% for rectal cancer patients who received a short course of radiotherapy with short interval to surgery between 2011 and 2014 (41). Our study obtained similar results, whereby the more experienced radiologist's visual assessments could only accurately detect a small proportion of LNM+ with a sensitivity of 56.3%, specificity of 70.7%, PPV of 35.5%, and NPV of 84.9%.

Although receiver operating characteristic analysis indicated that the radiomic signature had superior predictive ability to that of the more experienced radiologist, the radiologist's assessment results should not be overlooked. Compared with the radiomic signature, the radiologist exhibited a lower sensitivity and a higher specificity. In the univariate logistic regression analysis, ymrT and ymrN stages were significantly associated with LNM+. In particular, ymrN stage was still an independent predictor for LNM+ even when considering radiomic signature and ymrT stage as covariates. Thus, based on the advantages of radiomic

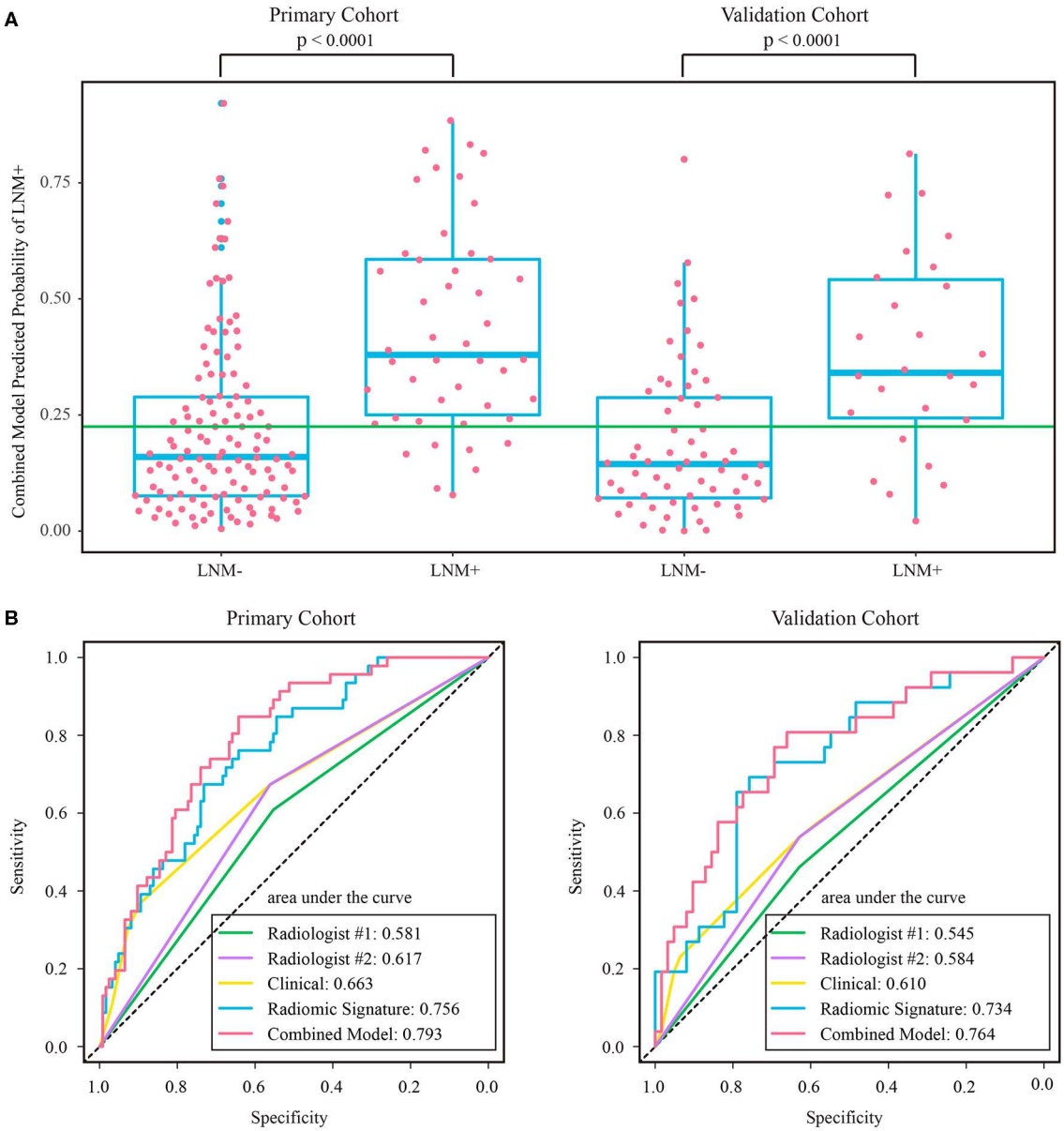


FIGURE 6 | (A) Distribution of combined model predicted probability of LNM+ in post-therapeutic ymrT3-4 subgroups of both cohorts, where the green line indicates the Youden cutoff in the primary cohort. **(B)** ROC curves of radiologists and prediction models in post-therapeutic ymrT3-4 subgroups of both cohorts. LNM+, lymph node metastasis; LNM–, lymph node non-metastasis; ROC, receiver operating characteristic.

signature and the radiologist's restaging results, the combined model was able to achieve a higher prediction accuracy.

The results of the stratified analyses highlight the potential of the combined model for clinical application. For patients with post-therapeutic ymrT1-2 tumors following neoadjuvant therapy, the combined model achieved a NPV of 100% and corresponding specificity of 63.8 and 59% in the primary cohort and validation cohort, respectively. This result indicates that approximately 60% of ymrT1-2 patients with LNM- would benefit from the model's prediction results. In practice, choosing less invasive treatment after neoadjuvant therapy for rectal cancer is a difficult and complex decision for both the doctor and the patient. Local excision or wait-and-watch is typically only considered for ypT0-2, lymph node-negative patients. However, we cannot obtain the ypT stage and lymph node status other than by pathologic evaluation after TME. This contradiction spurs us on to achieve a more precise clinical T/N staging. We believe that our combined model can serve as an important assistive tool for assessing the likelihood of node status following neoadjuvant therapy. Further research aiming at the simultaneous precise prediction of ypT stage and ypN before TME is indispensable to promote organ preservation strategies in the clinic.

Radiomics is a data-driven approach which has been successfully used to assess treatment response after neoadjuvant therapy (30) and to predict pathological features such as degree of differentiation, T stage, and N stage (31). It is an advanced framework which selects the most useful features from a high-throughput feature set to build a signature correlated to an object in a linear or a non-linear way. To the best of our knowledge, our study may be one of the first attempts to cope with this clinical problem by using radiomics. The selected radiomic features included understandable first-order statistics features such as LoG3-FOS_Mean, LoG2-FOS_Skewness, and so on, which reflect the strength information of tumor. The selected features also included textural features such as LoG3-GLCM_cshade, LoG2-GLSZM_LZLGE, and so on, which reflect a high-order statistical property among image elements and usually cannot be visually examined, but we believe that these features can be associated with an underlying pathology. Some published studies have mapped radiomic features to gene mutation (42, 43) and molecular pathway activation (44–46) by a radiogenomic method (47, 48). In the future, interpreting these selected features by specific genetic profiles may help to improve decision making in node restaging.

Several limitations existed in this study. It was a retrospective study with single-center samples in China. The chemoradiotherapy regimens usually are not the same in different hospitals, which may cause different lymph node responses. The imaging equipment parameters are usually different in multicenter research, which makes the reliability of the extracted features challenged. In order to control for confounders as much as possible, we conducted our study in a single hospital. Another limitation is that the enrolled sample size was relatively small, especially for the post-therapeutic ymrT1-2 subgroup. Thus, a prospective, international, multicenter clinical trial with a large sample size is needed to confirm our findings. In addition, only two radiologists were involved in the diagnosis in our study, and the more experienced radiologist provided a more accurate diagnosis. Thus, future research should include more experienced radiologists. Perirectal environment is another area that is worth to analyze, but blood vessels, muscles, nerves, and posttreatment edema may exist in this area. These confounding factors may affect the extracted features, causing negative effects to the accuracy of the prediction results. Although manually excluding these confounding factors on MRI is very time-consuming, it is worth trying to analyze the perirectal environment in a further study to get better prediction accuracy. Deep learning is an emerging field that surpasses radiomics in many tasks. Modeling with deep learning to correctly identify more LNM- patients may be a promising direction.

In summary, we demonstrated that combining a radiologist's staging results and radiomics analysis assists in the prediction of lymph node status in patients with LARC following neoadjuvant therapy, especially for patients with post-therapeutic ymrT1-2 tumors. An external validation of this study is warranted to guide the treatment recommendations for patients eligible for organ preservation strategies.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available as it is private data that belongs to the Sixth Affiliated Hospital of Sun Yat-sen University. Requests to access the datasets should be directed to corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by institutional review board of Sixth Affiliated Hospital of Sun Yat-sen University. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

JT, YF, and WC conceptualized and designed the study. Data were acquired by WC, YY, ZZ, and YF. XZ, YY, WC, BL, LL, KS, and LH analyze and interpreted the data. XZ, YY, ZL, and ZZ drafted the manuscript. JT, YF, and WC critically revised the paper.

FUNDING

This work was supported by the National Natural Science Foundation of China (grant numbers 81922040, 81772012, 81227901, and 81527805), the Beijing Natural Science Foundation (grant number 7182109), the National Key Research and Development Plan of China (grant numbers 2016YFA0100900, 2016YFA0100902, and 2017YFA0205200), the Youth Innovation Promotion Association CAS (grant number 2019136), the Chinese Academy of Sciences (grant numbers GJJSTD20170004, KFJ-STS-ZDTP-059, and YJKYYQ20180048), the Strategic Priority Research Program from CAS (grant number XDBS01030200), and the Key Research Projects in Frontier Science of CAS (grant number QYZDJ-SSW-JSC005).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00604/full#supplementary-material

REFERENCES

1. Benson AB III, Venook AP, Al-Hawary MM, Cederquist L, Chen YJ, Ciombor KK, et al. Rectal cancer, Version 2.2018, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Netw.* (2018) 16:874–901. doi: 10.6004/jnccn.2018.0061

2. Collette L, Bosset JF, den Dulk M, Nguyen F, Mineur L, Maingon P, et al. Patients with curative resection of cT3-4 rectal cancer after preoperative radiotherapy or radiochemotherapy: does anybody benefit from adjuvant fluorouracil-based chemotherapy? A trial of the European Organisation for Research and Treatment of Cancer Radiation Oncology Group. *J Clin Oncol.* (2007) 25:4379–86. doi: 10.1200/jco.2007.11.9685

3. Park IJ, You YN, Agarwal A, Skibber JM, Rodriguez-Bigas MA, Eng C, et al. Neoadjuvant treatment response as an early response indicator for patients with rectal cancer. *J Clin Oncol.* (2012) 30:1770–6. doi: 10.1200/jco.2011.39.7901
4. Marijnen CA, van de Velde CJ, Putter H, van den Brink M, Maas CP, Martijn H, et al. Impact of short-term preoperative radiotherapy on health-related quality of life and sexual functioning in primary rectal cancer: report of a multicenter randomized trial. *J Clin Oncol.* (2005) 23:1847–58. doi: 10.1200/jco.2005.05.256
5. Peeters KC, van de Velde CJ, Leer JW, Martijn H, Junggeburt JM, Kranenbarg EK, et al. Late side effects of short-course preoperative radiotherapy combined with total mesorectal excision for rectal cancer: increased bowel dysfunction in irradiated patients–a Dutch colorectal cancer group study. *J Clin Oncol.* (2005) 23:6199–206. doi: 10.1200/jco.2005.14.779
6. Pahlman L, Bohe M, Cedermark B, Dahlberg M, Lindmark G, Sjodahl R, et al. The Swedish rectal cancer registry. *Br J Surg.* (2007) 94:1285–92. doi: 10.1002/bjs.5679
7. Baxter NN, Garcia-Aguilar J. Organ preservation for rectal cancer. *J Clin Oncol.* (2007) 25:1014–20. doi: 10.1200/jco.2006.09.7840
8. Kim CJ, Yeatman TJ, Coppola D, Trotti A, Williams B, Barthel JS, et al. Local excision of T2 and T3 rectal cancers after downstaging chemoradiation. *Ann Surg.* (2001) 234:352–8; discussion 358–9. doi: 10.1097/00000658-200109000-00009
9. Schell SR, Zlotecki RA, Mendenhall WM, Marsh RW, Vauthey JN, Copeland EM. III. Transanal excision of locally advanced rectal cancers downstaged using neoadjuvant chemoradiotherapy. *J Am Coll Surg.* (2002) 194:584–90; discussion 590–1. doi: 10.1016/s1072-7515(02)01128-6
10. Callender GG, Das P, Rodriguez-Bigas MA, Skibber JM, Crane CH, Krishnan S, et al. Local excision after preoperative chemoradiation results in an equivalent outcome to total mesorectal excision in selected patients with T3 rectal cancer. *Ann Surg Oncol.* (2010) 17:441–7. doi: 10.1245/s10434-009-0735-7
11. Lezoche E, Baldarelli M, Lezoche G, Paganini AM, Gesuita R, Guerrieri M. Randomized clinical trial of endoluminal locoregional resection versus laparoscopic total mesorectal excision for T2 rectal cancer after neoadjuvant therapy. *Br J Surg.* (2012) 99:1211–8. doi: 10.1002/bjs.8821
12. Pucciarelli S, De Paoli A, Guerrieri M, La Torre G, Maretto I, De Marchi F, et al. Local excision after preoperative chemoradiotherapy for rectal cancer: results of a multicenter phase II clinical trial. *Dis Colon Rectum.* (2013) 56:1349–56. doi: 10.1097/DCR.0b013e3182a2303e
13. De Nardi P, Carvello M. How reliable is current imaging in restaging rectal cancer after neoadjuvant therapy? *World J Gastroenterol.* (2013) 19:5964–72. doi: 10.3748/wjg.v19.i36.5964
14. van der Paardt MP, Zagers MB, Beets-Tan RG, Stoker J, Bipat S. Patients who undergo preoperative chemoradiotherapy for locally advanced rectal cancer restaged by using diagnostic MR imaging: a systematic review and meta-analysis. *Radiology.* (2013) 269:101–12. doi: 10.1148/radiol.13122833
15. Newton AD, Li J, Jeganathan AN, Mahmoud NN, Epstein AJ, Paulson EC. A nomogram to predict lymph node positivity following neoadjuvant chemoradiation in locally advanced rectal Cancer. *Dis Colon Rectum.* (2016) 59:710–7. doi: 10.1097/DCR.0000000000000638
16. Azizian A, Kramer F, Jo P, Wolff HA, Beissbarth T, Skarupke R, et al. Preoperative prediction of lymph node status by circulating Mir-18b and Mir-20a during chemoradiotherapy in patients with rectal cancer. *World J Surg.* (2015) 39:2329–35. doi: 10.1007/s00268-015-3083-8
17. Dumont F, Dartigues P, Delga B, Thibaudeau E, Benhaim L, Campion L, et al. Model predicting the ypN0 status after good response to chemoradiotherapy in rectal cancer. *Am J Surg.* (2018) 216:438–43. doi: 10.1016/j.amjsurg.2018.03.025
18. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
19. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics.* (2019) 9:1303–22. doi: 10.7150/thno.30309
20. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res.* (2017) 23:7253–62. doi: 10.1158/1078-0432.Ccr-17-1038
21. Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, et al. Radiomics of multi-parametric MRI for pretreatment prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer: a multicenter study. *Clin Cancer Res.* (2019) 25:3538–47. doi: 10.1158/1078-0432.Ccr-18-3190
22. Tang Z, Zhang XY, Liu Z, Li XT, Shi YJ, Wang S, et al. Quantitative analysis of diffusion weighted imaging to predict pathological good response to neoadjuvant chemoradiation for locally advanced rectal cancer. *Radiother Oncol.* (2019) 132:100–8. doi: 10.1016/j.radonc.2018.11.007
23. Zhou X, Yi Y, Liu Z, Cao W, Lai B, Sun K, et al. Radiomics-based pretherapeutic prediction of non-response to neoadjuvant therapy in locally advanced rectal cancer. *Ann Surg Oncol.* (2019) 26:1676–84. doi: 10.1245/s10434-019-07300-3
24. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology.* (2016) 281:947–57. doi: 10.1148/radiol.2016152234
25. Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin Cancer Res.* (2017) 23:4259–69. doi: 10.1158/1078-0432.Ccr-16-2910
26. Zheng J, Chakraborty J, Chapman WC, Gerst S, Gonen M, Pak LM, et al. Preoperative prediction of microvascular invasion in hepatocellular carcinoma using quantitative image analysis. *J Am Coll Surg.* (2017) 225, 778–88.e771. doi: 10.1016/j.jamcollsurg.2017.09.003
27. Liu Z, Wang Y, Liu X, Du Y, Tang Z, Wang K, et al. Radiomics analysis allows for precise prediction of epilepsy in patients with low-grade gliomas. *Neuroimage Clin.* (2018) 19:271–8. doi: 10.1016/j.nicl.2018.04.024
28. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol.* (2016) 34:2157–64. doi: 10.1200/jco.2015.65.9128
29. Meng X, Xia W, Xie P, Zhang R, Li W, Wang M, et al. Preoperative radiomic signature based on multiparametric magnetic resonance imaging for noninvasive evaluation of biological characteristics in rectal cancer. *Eur Radiol.* (2018) 29:3200–9. doi: 10.1007/s00330-018-5763-x
30. Horvat N, Veeraraghavan H, Khan M, Blazic I, Zheng J, Capanu M, et al. MR imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy. *Radiology.* (2018) 287:833–43. doi: 10.1148/radiol.2018172300
31. Ma X, Shen F, Jia Y, Xia Y, Li Q, Lu J. MRI-based radiomics of rectal cancer: preoperative assessment of the pathological features. *BMC Med Imaging.* (2019) 19:86. doi: 10.1186/s12880-019-0392-7
32. Tibshirani RJ. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B.* (1996) 58:267–88.
33. Park SH, Lim JS, Lee J, Kim HY, Koom WS, Hur H, et al. Rectal mucinous adenocarcinoma: MR imaging assessment of response to concurrent chemotherapy and radiation therapy-A hypothesis-generating study. *Radiology.* (2017) 285:124–33. doi: 10.1148/radiol.2017162657
34. Mc NQ. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika.* (1947) 12:153–7. doi: 10.1007/bf02295996
35. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* (2008) 27:157–72; discussion 207–112. doi: 10.1002/sim.2929
36. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. *Crit Care Med.* (2007) 35:2052–6. doi: 10.1097/01.Ccm.0000275267.64078.B0
37. Gerard JP, Chamorey E, Gourgou-Bourgade S, Benezery K, de Laroche G, Mahe MA, et al. Clinical complete response (cCR) after neoadjuvant chemoradiotherapy and conservative treatment in rectal cancer. Findings from the ACCORD 12/PRODIGE 2 randomized trial. *Radiother Oncol.* (2015) 115:246–52. doi: 10.1016/j.radonc.2015.04.003
38. Youden WJ. Index for rating diagnostic tests. *Cancer.* (1950) 3:32–5.
39. Deng Y, Chi P, Lan P, Wang L, Chen W, Cui L, et al. Modified FOLFOX6 with or without radiation versus fluorouracil and leucovorin with radiation in

neoadjuvant treatment of locally advanced rectal cancer: initial results of the Chinese FOWARC multicenter, open-label, randomized three-arm phase III trial. *J Clin Oncol.* (2016) 34:3300–7. doi: 10.1200/jco.2016.66.6198
40. Lahaye MJ, Beets GL, Engelen SM, Kessels AG, de Bruine AP, Kwee HW, et al. Locally advanced rectal cancer: MR imaging for restaging after neoadjuvant radiation therapy with concomitant chemotherapy. Part II What are the criteria to predict involved lymph nodes? *Radiology.* (2009) 252:81–91. doi: 10.1148/radiol.2521081364
41. Brouwer NPM, Stijns RCH, Lemmens V, Nagtegaal ID, Beets-Tan RGH, Futterer JJ, et al. Clinical lymph node staging in colorectal cancer; a flip of the coin? *Eur J Surg Oncol.* (2018) 44:1241–6. doi: 10.1016/j.ejso.2018.04.008
42. Chen SW, Chiang HC, Chen WT, Hsieh TC, Yen KY, Chiang SF, et al. Correlation between PET/CT parameters and KRAS expression in colorectal cancer. *Clin Nucl Med.* (2014) 39:685–9. doi: 10.1097/rlu.0000000000000481
43. Cui Y, Liu H, Ren J, Du X, Xin L, Li D, et al. Development and validation of a MRI-based radiomics signature for prediction of KRAS mutation in rectal cancer. *Eur Radiol.* (2020) 30:1948–58. doi: 10.1007/s00330-019-06572-3
44. Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, et al. (2017). Defining the biological basis of radiomic phenotypes in lung cancer. *Elife*. 6: e23421. doi: 10.7554/eLife.23421
45. Wu J, Li B, Sun X, Cao G, Rubin DL, Napel S, et al. Heterogeneous enhancement patterns of tumor-adjacent parenchyma at MR imaging are associated with dysregulated signaling pathways and poor survival in breast cancer. *Radiology.* (2017) 285:401–13. doi: 10.1148/radiol.2017162823
46. Badic B, Hatt M. Radiogenomics-based cancer prognosis in colorectal cancer. *Sci Rep.* (2019) 9:9743. doi: 10.1038/s41598-019-46286-6
47. Pinker K, Shitano F, Sala E, Do RK, Young RJ, Wibmer AG, et al. Background, current role, and potential applications of radiogenomics. *J Magn Reson Imaging.* (2018) 47:604–20. doi: 10.1002/jmri.25870
48. Horvat N, Bates DDB, Petkovska I. Novel imaging techniques of rectal cancer: what do radiomics and radiogenomics have to offer? A literature review *Abdomin Radiol.* (2019) 44:3764–74. doi: 10.1007/s00261-019-02042-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 28 May 2020
doi: 10.3389/fonc.2020.00752

Machine-Learning Classifiers in Discrimination of Lesions Located in the Anterior Skull Base

Yang Zhang[1†], Lan Shang[2†], Chaoyue Chen[1†], Xuelei Ma[3,4], Xuejin Ou[5], Jian Wang[6], Fan Xia[1] and Jianguo Xu[1*]*

[1] Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu, China, [2] Department of Radiology, Sichuan Provincial People's Hospital, Sichuan Academy of Medical Sciences, Chengdu, China, [3] Department of Biotherapy, Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [4] State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University and Collaborative Innovation Center for Biotherapy, Chengdu, China, [5] West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China, [6] School of Computer Science, Nanjing University of Science and Technology, Nanjing, China

OPEN ACCESS

Edited by:
Andreia Vasconcellos Faria, Johns Hopkins University, United States

Reviewed by:
Ivan Jambor, University of Turku, Finland
Yanwei Miao, Dalian Medical University, China

***Correspondence:**
Xuelei Ma
drmaxuelei@gmail.com
Jianguo Xu
drjianguoxu@gmail.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *10 May 2019*
Accepted: *20 April 2020*
Published: *28 May 2020*

Citation:
Zhang Y, Shang L, Chen C, Ma X, Ou X, Wang J, Xia F and Xu J (2020) Machine-Learning Classifiers in Discrimination of Lesions Located in the Anterior Skull Base. Front. Oncol. 10:752. doi: 10.3389/fonc.2020.00752

Purpose: The aim of this study was to investigate the diagnostic value of machine-learning models with radiomic features and clinical features in preoperative differentiation of common lesions located in the anterior skull base.

Methods: A total of 235 patients diagnosed with pituitary adenoma, meningioma, craniopharyngioma, or Rathke cleft cyst were enrolled in the current study. The discrimination was divided into three groups: pituitary adenoma vs. craniopharyngioma, meningioma vs. craniopharyngioma, and pituitary adenoma vs. Rathke cleft cyst. In each group, five selection methods were adopted to select suitable features for the classifier, and nine machine-learning classifiers were employed to build discriminative models. The diagnostic performance of each combination was evaluated with area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity calculated for both the training group and the testing group.

Results: In each group, several classifiers combined with suitable selection methods represented feasible diagnostic performance with AUC of more than 0.80. Moreover, the combination of least absolute shrinkage and selection operator as the feature-selection method and linear discriminant analysis as the classification algorithm represented the best comprehensive discriminative ability.

Conclusion: Radiomics-based machine learning could potentially serve as a novel method to assist in discriminating common lesions in the anterior skull base prior to operation.

Keywords: pituitary adenoma, meningioma, craniopharyngioma, Rathke cleft cyst, anterior skull base, radiomics, machine learning

INTRODUCTION

A variety of lesions are present in the anterior skull base. The most common types of tumors in this area are pituitary adenoma, craniopharyngioma, and meningioma (1, 2). Rathke cleft cyst is also taken as the common differential diagnosis for the sellar mass as a congenital lesion (3). The importance of early diagnosis for lesions in this region has been highlighted because even these benign lesions may be progressive and unrelenting if situated in an area where growth cannot be controlled, and some of them could show aggressive behavior (4). Magnetic resonance (MR) scan is highly recommended for preoperative evaluation of the anterior skull base lesion owing to the advantage of excellent soft tissue resolution. Descriptions of the four types of lesions in MR imaging (MRI) are characteristic (5). However, the diagnostic accuracy of MR images depends on experiences of radiologists, and in some cases, these lesions with similar MRI patterns may mimic each other and complicate the radiological diagnosis (6, 7). Therefore, new methods that could assist in preoperative differentiation may be of clinical value.

Radiomics could extract high-dimensional features from medical images and provide information associated with the pathophysiology of lesions that is difficult to be assessed by visual inspection (8–10). Moreover, mineable radiomic features of lesions could be analyzed with the novel machine-learning technology that has shown promising prospects in the biomedical domain (11). Radiomics-based machine learning has been applied in differential diagnosis of various brain tumors in previous studies, representing the potential to be utilized in clinical practice to facilitate diagnosis, and offer guidance for decision making (12–16). In the present study, we evaluated the ability of machine-learning technology combined with MRI radiomic features and clinical parameters in differentiating the four common types of lesions in the anterior skull base. Considering the epidemiology and position of lesions, the differential analysis was divided into three groups: pituitary adenoma vs. craniopharyngioma (the most common tumors in the sellar/suprasellar region), meningioma vs. craniopharyngiomas (the most common tumors in the parasellar region), and pituitary adenoma vs. Rathke cleft cyst (the most common lesions in the intrasellar region).

METHOD

Patient Selection

Institution database was reviewed to search for patients treated at our neurosurgery department from November 2014 to June 2018. We initially selected the potentially qualified patients according to the following criteria: (a) with the pathological confirmation of pituitary adenoma, craniopharyngioma, meningioma, or Rathke cleft cyst; (b) the lesion was located at the anterior skull base; and (c) with preoperative sellar MR images. Exclusion criteria were as follows: (a) history of any other intracranial diseases, such as stroke and intracranial infection; (b) history of any anti-tumor treatment prior to MR scans, such as brain surgery, chemotherapy, or radiotherapy; and (c) incomplete electronic medical records. The flowchart of patient selection is shown in **Figure 1**. Clinical parameters were recorded, including age, gender, lesion size, and the time between MR scan and surgery. The lesion size was measured by the maximum diameter of the lesion that was collected from radiological reports. This retrospective study was approved by the institutional review board. The written informed consent was obtained from all participants (written informed consent for patients <16 years old was obtained from their parents or guardians).

Image Acquisition

All patients underwent MR scans *via* a 3.0-T GE scanner with an eight-channel phase array head coil. The parameters of the contrast-enhanced T1-weighted imaging were as follows: TR/TE = 552/10 ms, slice thickness = 5 mm, flip angle = 90°, field of view = 150 × 150 mm^2, data matrix = 256 × 256, and voxel size = 1.0 × 1.0 × 1.0 mm^3. The scanning was conducted within 200 s after injection of gadopentetate dimeglumine (0.1 mmol/kg) as the contrast agent. The preoperative MR images were collected from picture archiving and communication system (PACS) of our institutional radiology department (**Figure 2**).

Feature Extraction

Texture features were extracted from MR images as radiomic parameters by two neurosurgeons together with the assistance of senior radiologists using LIFEx software (http://www.lifexsoft.org) (17). Following protocols of the software, the region of interest (ROI) was manually drawn within the border of the lesion in each slice. Considering the clear depiction of the boundary of lesions, ROI delineation was performed on the contrast-enhanced T1-weighted imaging, in which lesions were carefully separated from adjacent brain tissues through different enhancement patterns and surrounding anatomic structures. Any disagreements regarding the border of lesions were recorded and solved by senior radiologists. After the whole lesion was contoured slice by slice, three-dimensional radiomic parameters could be automatically calculated by the software with established formulas (**Supplementary Material 1**). A total of 40 features were obtained from two orders, including the first-order features from shape-based matrix and histogram-based matrix, and the second-order/higher-order features from gray-level co-occurrence matrix (GLCM), gray-level zone length matrix (GLZLM), neighborhood gray-level dependence matrix (NGLDM), and gray-level run length matrix (GLRLM) (**Supplementary Material 2**). Examples of ROI delineation are shown in **Supplementary Material 3**. Combined with two clinical parameters (age and gender), a dataset was built for further analysis.

Machine-Learning Modeling

Given that there was a relatively large number of statistics and some parameters may not be associated with the differential process, optimal features should be selected first for the predictive model. The feature-selection method was important but complicated considering the sample size and efficiency in discrimination. Least absolute shrinkage and selection operator (LASSO) regression model was reported to be appropriate for high-dimensional data regression analysis (18, 19). Other

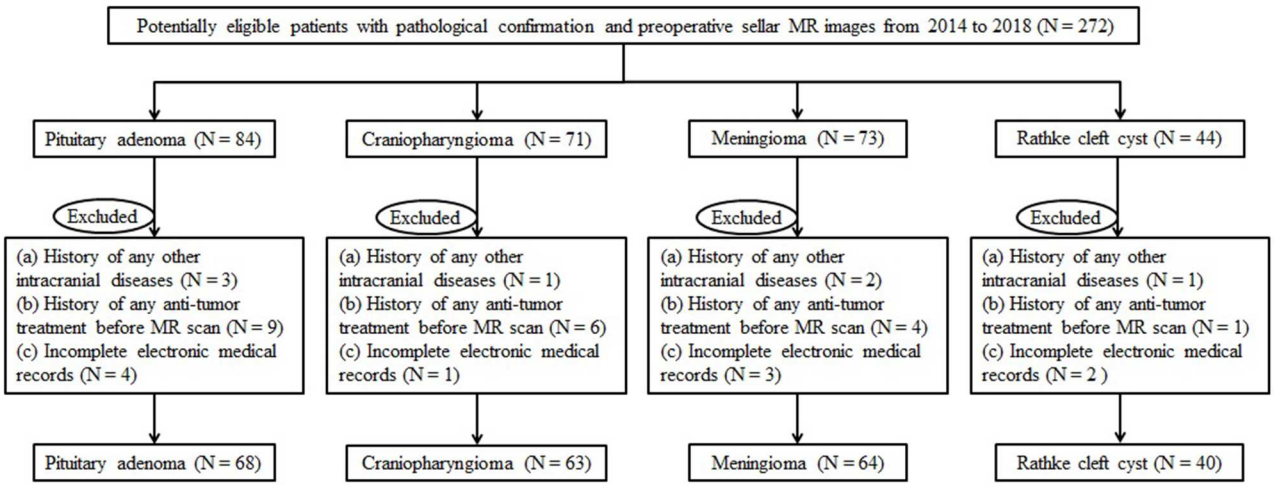


FIGURE 1 | Workflow chart of the patient enrollment process.

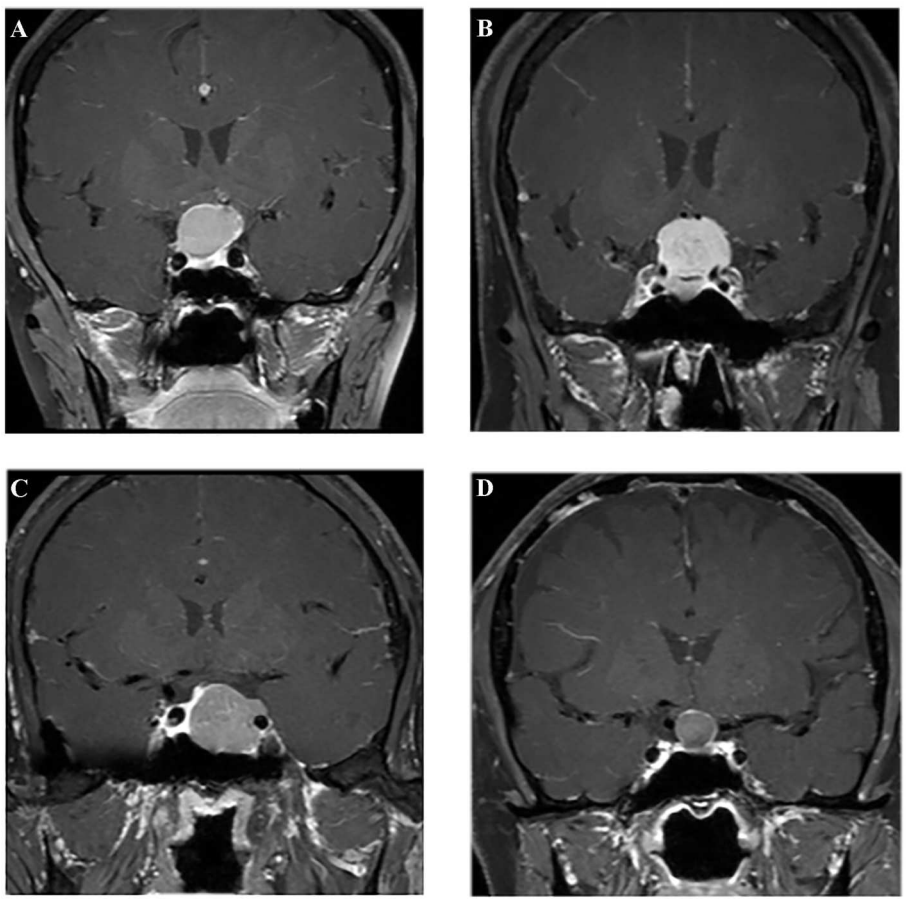


FIGURE 2 | Examples of different lesions on contrast-enhanced T1-weighted image. **(A)** Craniopharyngioma; **(B)** meningioma; **(C)** pituitary adenoma; **(D)** Rathke cleft cyst.

TABLE 1 | Characteristics of patients and lesions.

	Craniopharyngioma	Meningioma	Pituitary adenoma	Rathke cleft cyst	*P*-value
Number	63	64	68	40	
Gender, *n* (%)					0.006
Male	37 (58.7)	18 (28.1)	32 (47.1)	18 (45.0)	
Female	26 (41.3)	46 (71.9)	36 (52.9)	22 (55.0)	
Age, *n* (%)					<0.001
≤18 years	21 (33.3)	2 (3.1)	1 (1.5)	0 (0.0)	
19∼30 years	11 (17.5)	0 (0.0)	7 (10.3)	11 (27.5)	
31∼60 years	27 (42.9)	51 (79.7)	43 (63.2)	26 (65.0)	
>60 years	4 (6.3)	11 (17.2)	17 (25.0)	3 (7.5)	
Mean age (range) (year)	31.62 (2∼73)	49.19 (9∼72)	49.16 (18∼73)	44.23 (21∼68)	
Maximum diameter (mm)	28.86 (12.5∼52.4)	20.41 (8∼40)	23.21 (7∼50.5)	19.87 (8∼38.3)	<0.001
Average time between MR scan and surgery (day)	6.2	7.5	5.3	6.4	0.321

MR, magnetic resonance.

TABLE 2 | Results of the discriminative model of LASSO + LDA in distinguishing lesions in the training group and the testing group.

	Training group				Testing group			
	AUC	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity
Pituitary adenoma vs. craniopharyngioma	0.845	0.851	0.897	0.820	0.804	0.800	0.888	0.734
Meningioma vs. craniopharyngioma	0.882	0.881	0.944	0.832	0.807	0.819	0.863	0.794
Pituitary adenoma vs. Rathke cleft cyst	0.873	0.887	0.861	0.901	0.816	0.836	0.829	0.840

AUC, area under curve; LASSO, least absolute shrinkage and selection operator; LDA, linear discriminant analysis.

selection methods were also evaluated by previous researchers and reported to represent good diagnostic performances (20, 21). To settle the dilemma, five different feature-selection methods were adopted, namely, distance correlation, random forest (RF), LASSO, eXtreme gradient boosting (Xgboost), and gradient boosting decision tree (GBDT). A similar predicament also needs to be solved in regard to the selection of machine-learning classifiers. We employed nine classification algorithms in this study, including linear discriminant analysis (LDA), support vector machine (SVM), RF, Adaboost, *k*-nearest neighbor (KNN), GaussianNB, logistic regression (LR), GBDT, and decision tree (DT). Patients were randomly divided into the training group and the testing group at the ratio of 4:1 on the basis of experiences from previous studies (22–24). The model was first created by the training group and then applied to the independent testing group, and this procedure was repeated over 100 times to conclude the realistic distribution of classification accuracies. Area under receiver operating characteristic (ROC) curve (AUC), accuracy, sensitivity, and specificity were calculated based on the confusion matrix to assess the discriminative ability of different models. Regular statistical analyses of this study were performed using SPSS (Version 22.0, IBM Corp. Armonk, NY, USA), and machine-learning algorithms were programmed with Python Programming Language and scikit-learn package.

RESULTS

Patient Characteristic

A total number of 235 patients who underwent surgical resection of lesions in our neurosurgery department were enrolled in the current study. Among all participants involved, 68 patients were diagnosed with pituitary adenoma, 63 patients were diagnosed with craniopharyngioma, 64 patients were diagnosed with meningioma, and 40 patients were diagnosed with Rathke cleft cyst. The average age of patients was 49.16, 31.62, 49.19, and 44.23 years, respectively. The mean value of the maximum diameter of lesions was 23.21, 28.86, 20.41, and 19.87 mm, respectively. The characteristics of patients and lesions are summarized in **Table 1**.

Machine-Learning Model Assessment

In each group, 45 diagnostic models were established through the combinations of five selection methods and nine classifiers. The combination of LASSO as the selection method and LDA as the classifier (LASSO + LDA) seemed to be the optimal model in differentiating common lesions in the anterior skull base with AUC of more than 0.80 in all three groups. It is worth noting that some combinations represented better performance than LASSO + LDA in a single group, but LASSO + LDA showed the best comprehensive discriminative ability.

Group 1 was the differentiation between pituitary adenoma and craniopharyngioma considering these are the most common tumors located in the sellar/suprasellar region. For LASSO + LDA, ROC analysis demonstrated that AUC, accuracy, sensitivity, and specificity in the training group were 0.845, 0.851, 0.897, and 0.820, respectively. In the testing group, this predictive model was proven to be feasible in discrimination with AUC of 0.804, accuracy of 0.800, sensitivity of 0.888, and specificity of 0.734 (**Table 2**). Besides, other models like RF + RF (AUC = 0.811 in the testing group) and GBDT + RF (AUC = 0.837 in the

TABLE 3 | Results of discriminative models in distinguishing pituitary adenoma from craniopharyngioma in the testing group.

	Distance correlation		RF		LASSO		Xgboost		GBDT	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
LDA	0.719	0.727	0.778	0.782	0.800	0.804	0.730	0.734	0.793	0.799
SVM	0.696	0.712	/	/	0.700	0.717	0.700	0.696	/	/
RF	0.727	0.747	0.804	0.811	0.770	0.780	0.781	0.786	0.841	0.837
Adaboost	0.796	0.799	0.833	0.837	0.785	0.784	0.770	0.774	0.833	0.831
KNN	0.800	0.800	0.689	0.690	0.756	0.765	0.689	0.694	0.722	0.727
GaussianNB	0.744	0.750	0.726	0.730	0.670	0.681	0.715	0.724	0.737	0.741
LR	0.752	0.758	0.822	0.819	0.774	0.783	0.693	0.705	0.767	0.771
GBDT	0.796	0.796	0.859	0.857	0.874	0.866	0.811	0.809	0.844	0.840
DT	0.752	0.754	0.800	0.798	0.767	0.766	0.763	0.757	0.785	0.783

RF, random forest; LASSO, least absolute shrinkage and selection operator; Xgboost, eXtreme gradient boosting; GBDT, gradient boosting decision tree; LDA, linear discriminant analysis; SVM, support vector machine; KNN, k-nearest neighbor; LR, logistic regression; DT, decision tree; AUC, area under curve.
/, over-fitting.

TABLE 4 | Results of discriminative models in distinguishing meningioma from craniopharyngioma in the testing group.

	Distance correlation		RF		LASSO		Xgboost		GBDT	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
LDA	0.846	0.843	0.850	0.842	0.819	0.807	0.800	0.792	0.815	0.809
SVM	0.807	0.804	/	/	0.712	0.732	/	/	/	/
RF	0.769	0.777	0.773	0.780	0.735	0.744	0.796	0.798	0.812	0.822
Adaboost	0.753	0.766	0.746	0.758	0.777	0.784	0.784	0.790	0.773	0.781
KNN	0.838	0.846	0.708	0.713	0.742	0.746	0.669	0.663	0.650	0.656
GaussianNB	0.762	0.753	0.777	0.778	0.700	0.681	0.715	0.691	0.804	0.787
LR	0.796	0.800	0.777	0.783	0.765	0.763	0.735	0.725	0.785	0.780
GBDT	0.769	0.773	0.769	0.774	0.773	0.782	0.765	0.769	0.812	0.816
DT	0.742	0.744	0.723	0.722	0.712	0.710	0.719	0.726	0.765	0.767

RF, random forest; LASSO, least absolute shrinkage and selection operator; Xgboost, eXtreme gradient boosting; GBDT, gradient boosting decision tree; LDA, linear discriminant analysis; SVM, support vector machine; KNN, k-nearest neighbor; LR, logistic regression; DT, decision tree; AUC, area under curve.
/, over-fitting.

testing group) also represented feasible ability in distinguishing pituitary adenoma from craniopharyngioma (**Table 3**).

Group 2 was the differentiation between meningioma and craniopharyngioma, given that they are the most common tumors located in the parasellar region. ROC analysis illustrated the differential ability of LASSO + LDA with AUC of 0.882, accuracy of 0.881, sensitivity of 0.944, and specificity of 0.832 in the training group. In the testing group, AUC of LASSO + LDA was 0.807, accuracy was 0.819, sensitivity was 0.863, and specificity was 0.794 (**Table 2**). Besides, distance correlation + LDA (AUC = 0.843 in the testing group), RF + LDA (AUC = 0.842 in the testing group), GBDT + LDA (AUC = 0.809 in the testing group), and distance correlation + KNN (AUC = 0.846 in the testing group) also represented reliable diagnostic performance in discrimination between meningioma and craniopharyngioma (**Table 4**).

Group 3 was the differentiation between pituitary adenoma and Rathke cleft cyst, which are the most common lesions in the intrasellar region. In the training group, ROC analysis demonstrated that AUC of LASSO + LDA was 0.873 with accuracy of 0.887, sensitivity of 0.861, and specificity of 0.901. In the testing group, this model also represented feasible discriminative ability with AUC of 0.816, accuracy of 0.836, sensitivity of 0.829, and specificity of 0.840 (**Table 2**). In addition, distance correlation + RF also represented good performance in differentiating pituitary adenoma from Rathke cleft cyst with AUC of 0.825 in the testing group (**Table 5**).

The features selected into LASSO + LDA model are listed in **Table 6**. The association between discriminant functions for LASSO + LDA model is represented in **Figure 3**, in which minimal overlap between two clusters was observed in each group. **Figure 4** represents examples of distributions of the direct LDA function for lesions for one of the 100 independent cycles. In group 1, a shift of the LDA function values for craniopharyngioma toward positive values was shown while predominantly negative values for pituitary adenoma. Similar trends could be observed in group 2 and group 3, suggesting that the LASSO

TABLE 5 | Results of discriminative models in distinguishing pituitary adenoma from Rathke cleft cyst in the testing group.

	Distance correlation		RF		LASSO		Xgboost		GBDT	
	Accuracy	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**
LDA	0.841	0.803	0.827	0.804	0.836	0.816	0.754	0.682	0.786	0.767
SVM	0.754	0.678	/	/	0.627	0.500	0.645	0.544	/	/
RF	0.863	0.825	0.768	0.714	0.777	0.710	0.855	0.813	0.823	0.775
Adaboost	0.813	0.794	0.804	0.774	0.818	0.778	0.818	0.778	0.836	0.809
KNN	0.786	0.735	0.745	0.696	0.732	0.666	0.759	0.706	0.768	0.723
GaussianNB	0.841	0.806	0.814	0.786	0.677	0.655	0.800	0.745	0.827	0.792
LR	0.800	0.736	0.823	0.781	0.755	0.683	0.818	0.764	0.805	0.769
GBDT	0.845	0.808	0.832	0.798	0.786	0.743	0.818	0.776	0.850	0.821
DT	0.832	0.798	0.791	0.761	0.736	0.700	0.786	0.757	0.809	0.780

RF, random forest; LASSO, least absolute shrinkage and selection operator; Xgboost, eXtreme gradient boosting; GBDT, gradient boosting decision tree; LDA, linear discriminant analysis; SVM, support vector machine; KNN, k-nearest neighbor; LR, logistic regression; DT, decision tree; AUC, area under curve.
/, over-fitting.

+ LDA model had feasible discriminative ability in the three groups.

DISCUSSION

In the present study, a series of clinical parameters and radiomic features were utilized in differentiating four types of lesions in the anterior skull base. The predictive models were built using five feature-selection methods (distance correlation, RF, LASSO, Xgboost, and GBDT) and nine machine-learning classification algorithms (LDA, SVM, RF, Adaboost, KNN, GaussianNB, LR, GBDT, and DT). The combination of LASSO as the feature-selection method and LDA as the classification algorithm represented the optimal comprehensive performance with AUC of over 0.80 in all of the training groups and the testing groups. Moreover, several models also showed reliable discriminative ability between two types of lesions in a single group. Considering the features we selected could be extracted from routine MR images, the predictive model has the potential to be utilized as a novel, convenient tool in clinical practice.

The most important result of our research was to identify suitable discriminative models for lesions located in the anterior skull base. In previous studies, researchers investigated various combinations and tried to identify the optimal diagnostic or prognostic model. For instance, one study on CT-based survival prediction of non-small cell lung cancer involved models with four selection and classification methods (25). Another study made evaluations on models with 14 selection and 12 classification methods in predicting the overall survival of lung cancer patients (26). Similar studies were performed in bone tumor and head and neck cancer, implicating that the machine-learning model could potentially be a reliable method in differential diagnosis and prognosis prediction (27–30). However, it brought our attention and further investigation that various classifiers were used but that unanimous results on which one could be taken as the universal method were not reached. Considering that the purpose of clinical application of machine learning is to lessen the workload for doctors, simple discriminative models between two types of lesions are relatively meaningless because of complicated and elusive situations in clinical practice. Based on this idea, not only different combinations were tested, but also analyses on four types of lesions were performed simultaneously in three groups in the present study.

TABLE 6 | Parameters selected in the discriminative model of LASSO + LDA.

Group 1	**Group 2**	**Group 3**
Age	Age	minValue
minValue	minValue	meanValue
meanvalue	meanValue	maxValue
maxValue	stdValue	SHAPE_Volume
SHAPE_Volume	maxValue	GLRLM_SRHGE
GLCM_Contrast	SHAPE_Volume	GLRLM_LRHGE
GLRLM_HGRE	GLCM_Contrast	GLRLM_GLNU
GLRLM_SRHGE	GLRLM_HGRE	GLRLM_RLNU
GLRLM_LRHGE	GLRLM_LRHGE	GLZLM_LZE
GLRLM_GLNU	GLRLM_GLNU	GLZLM_HGZE
GLRLM_RLNU	GLRLM_RLNU	GLZLM_SZHGE
GLZLM_LZE	GLZLM_LZE	GLZLM_LZHGE
GLZLM_SZHGE	GLZLM_HGZE	GLZLM_ZLNU
GLZLM_LZLGE	GLZLM_SZHGE	
GLZLM_LZHGE	GLZLM_LZHGE	
GLZLM_GLNU	GLZLM_GLNU	
GLZLM_ZLNU	GLZLM_ZLNU	

GLCM, gray-level co-occurrence matrix; GLZLM, gray-level zone length matrix; NGLDM, neighborhood gray-level dependence matrix; GLRLM, gray-level run length matrix; LASSO, least absolute shrinkage and selection operator; LDA, linear discriminant analysis.

LASSO is a brilliant feature-selection method that tries to retain useful features in both ridge regression and subset selection (31). With the characteristics of avoiding over-fitting, it is suitable for large sets of radiomic features when a relatively small number of samples are involved (28). LDA is a machine-learning classification algorithm that could find a linear model with the best discriminative ability for two classes. The mechanism of LDA is to identify the boundaries around clusters of two classes and to project the statistics into a lower-dimensional space with

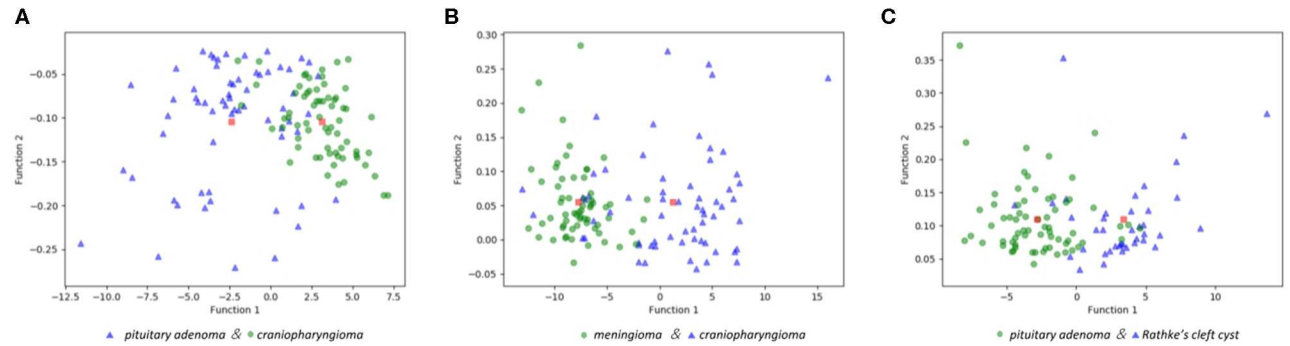


FIGURE 3 | Relationships between the discriminant functions for different lesions in the three groups and for the group centroids. **(A)** Pituitary adenoma vs. craniopharyngioma; **(B)** meningioma vs. craniopharyngioma; **(C)** pituitary adenoma vs. Rathke cleft cyst.

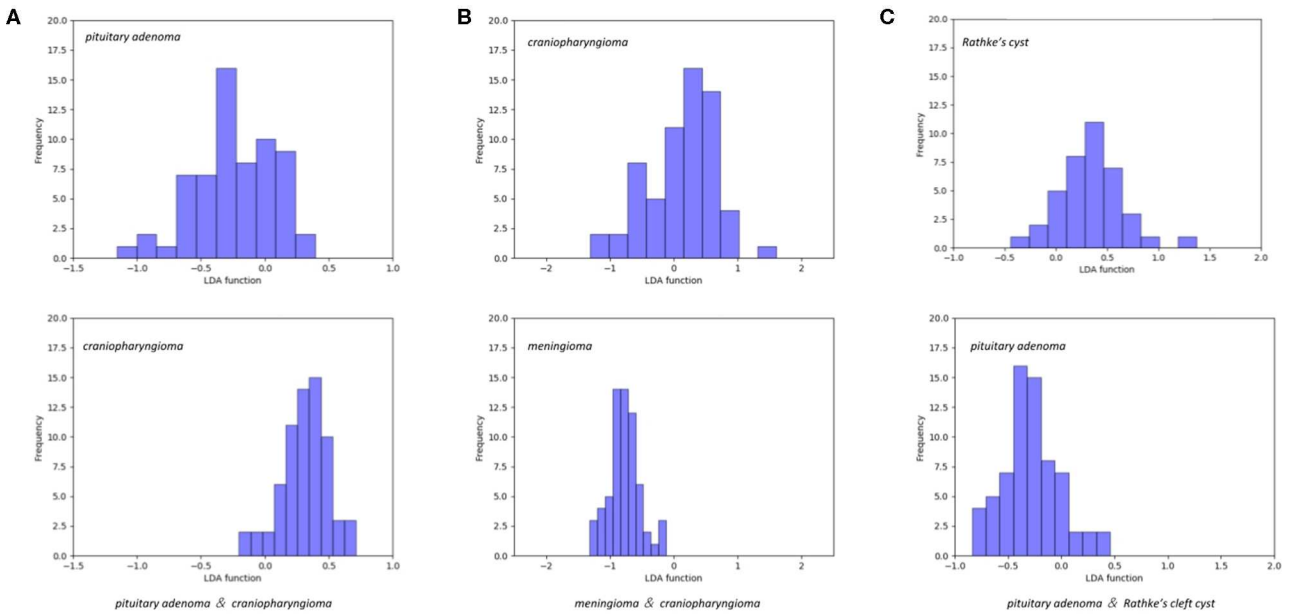


FIGURE 4 | Examples of distributions of the linear discriminant analysis (LDA) function determined for the lesions for one cycle. **(A)** Pituitary adenoma vs. craniopharyngioma; **(B)** meningioma vs. craniopharyngioma; **(C)** pituitary adenoma vs. Rathke cleft cyst.

good discriminative power based on the distance to a centroid of each cluster (32). LDA was reported to be able to reduce the dimensionality and to preserve the class discrimination information as much as possible. The combination of LASSO and LDA showed optimal comprehensive results in all three groups with AUC of more than 0.80 in each group. However, the results of our study were not as good as those of others. One study on the prediction of ATRX mutation in low-grade gliomas represented brilliant results with AUC of 0.925 in the validation group (19). Another study on differentiating sacral chordoma from sacral giant cell tumor represented AUC of 0.984 in the validation group (30). Future researches with more feature-selection methods and machine-learning classifiers are required to verify our results and to explore the optimal model with higher reliability.

There were some limitations in the present study. First, this study was performed in a single center, and only patients with resectable tumors were enrolled. Second, the study cohort, especially the testing cohort, was relatively small, which was a common limitation of other similar studies. Multicenter studies with larger sample sizes are required to validate our results. Third, only the contrast-enhanced T1-weighted imaging was used in radiomic analysis considering that this sequence was most suitable and available for the evaluation of lesions in the anterior skull base. Multi-model imaging statistics need to be integrated into the model to improve its performance in future studies. Fourth, images acquired from different MR scanners may possibly result in the model performance discrepancy. Standardized imaging quality and consistent protocols are required if the predictive models are put into clinical work.

In conclusion, we utilized MRI radiomics and clinical parameters to build predictive models *via* the combinations of selection methods and machine-learning classifiers. Our results indicated that radiomics-based machine learning could preoperatively differentiate common lesions in the anterior skull base with feasible diagnostic performance and facilitate clinical decision making.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Sichuan University. The written informed consent was obtained from participants enrolled in this study.

AUTHOR CONTRIBUTIONS

YZ participated in data interpretation and feature illustration, and revised the manuscript for important intellectual content. LS participated in image evaluation and manuscript revision. CC collected MR images, participated in feature extraction, and drafted the manuscript. XM participated in conceptualization and revised some intellectual content in manuscript. XO collected MR images and participated in feature extraction. JW deployed the machine-learning algorithm and participated in statistical analysis. FX assisted in feature extraction and data collection. JX participated in conceptualization and revised some intellectual content in manuscript. All authors read and approved the submitted version.

FUNDING

This work was supported by 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (ZYJC18007); Key research and development project of science and technology department of Sichuan Province (2019YFS0392).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00752/full#supplementary-material

Supplementary Material 1 | Explanations and formulas of radiomic features.

Supplementary Material 2 | Original data of extracted radiomic features.

Supplementary Material 3 | Examples of ROI delineation in MR images.

REFERENCES

1. Schwartz TH, Fraser JF, Brown S, Tabaee A, Kacker A, Anand VK. Endoscopic cranial base surgery: classification of operative approaches. *Neurosurgery*. (2008) 62:991–1002. doi: 10.1227/01.neu.0000325861.06832.06
2. Famini P, Maya MM, Melmed S. Pituitary magnetic resonance imaging for sellar and parasellar masses: 10-year experience in 2598 patients. *J Clin Endocrinol Metab*. (2011) 96:1633–41. doi: 10.1210/jc.2011-0168
3. Han SJ, Rolston JD, Jahangiri A, Aghi MK. Rathke's cleft cysts: review of natural history and surgical outcomes. *J Neurooncol*. (2014) 117:197–203. doi: 10.1007/s11060-013-1272-6
4. Ivan ME, Han SJ, Aghi MK. Tumors of the anterior skull base. *Exp Rev Neurother*. (2014) 14:425–38. doi: 10.1586/14737175.2014.892830
5. Zamora C, Castillo M. Sellar and parasellar imaging. *Neurosurgery*. (2017) 80:17–38. doi: 10.1093/neuros/nyw013
6. Chin BM, Orlandi RR, Wiggins RH 3rd. Evaluation of the sellar and parasellar regions. *Magn Reson Imaging Clin N Am*. (2012) 20:515-43. doi: 10.1016/j.mric.2012.05.007
7. Nowosielski M, Galldiks N, Iglseder S, Kickingereder P, von Deimling A, Bendszus M, et al. Diagnostic challenges in meningioma. *Neuro Oncol*. (2017) 19:1588–98. doi: 10.1093/neuonc/nox101
8. Kickingereder P, Burth S, Wick A, Gotz M, Eidel O, Schlemmer HP, et al. Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology*. (2016) 280:880–9. doi: 10.1148/radiol.2016160845
9. Prasanna P, Patel J, Partovi S, Madabhushi A, Tiwari P. Radiomic features from the peritumoral brain parenchyma on treatment-naïve multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. *Eur Radiol*. (2017) 27:4188–97. doi: 10.1007/s00330-016-4637-3
10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169
11. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. (2019) 69:127–57. doi: 10.3322/caac.21552
12. Kniep HC, Madesta F, Schneider T, Hanning U, Schonfeld MH, Schon G, et al. Radiomics of brain MRI: utility in prediction of metastatic tumor type. *Radiology*. (2019) 290:479–87. doi: 10.1148/radiol.2018180946
13. Li C, Wang S, Serra A, Torheim T, Yan JL, Boonzaier NR, et al. Multi-parametric and multi-regional histogram analysis of MRI: modality integration reveals imaging phenotypes of glioblastoma. *Eur Radiol*. (2019) 29:4718-29. doi: 10.1007/s00330-018-5984-z
14. Nie D, Lu J, Zhang H, Adeli E, Wang J, Yu Z, et al. Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. *Sci Rep*. (2019) 9:1103. doi: 10.1038/s41598-018-37387-9
15. Wu S, Meng J, Yu Q, Li P, Fu S. Radiomics-based machine learning methods for isocitrate dehydrogenase genotype prediction of diffuse gliomas. *J Cancer Res Clin Oncol*. (2019) 145:543–50. doi: 10.1007/s00432-018-2787-1
16. Rudie JD, Rauschecker AM, Bryan RN, Davatzikos C, Mohan S. Emerging applications of artificial intelligence in neuro-oncology. *Radiology*. (2019) 290:607–18. doi: 10.1148/radiol.2018181928
17. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res*. (2018) 78:4786–9. doi: 10.1158/0008-5472.CAN-18-0125
18. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. (2007) 26:5512–28. doi: 10.1002/sim.3148
19. Li Y, Liu X, Qian Z, Sun Z, Xu K, Wang K, et al. Genotype prediction of ATRX mutation in lower-grade gliomas using an MRI radiomics signature. *Eur Radiol*. (2018) 28:2960–8. doi: 10.1007/s00330-017-5267-0
20. Ditmer A, Zhang B, Shujaat T, Pavlina A, Luibrand N, Gaskill-Shipley M, et al. Diagnostic accuracy of MRI texture analysis for grading gliomas. *J Neurooncol*. (2018) 140:583–9. doi: 10.1007/s11060-018-2984-4

21. Bonora M, Wieckowsk MR, Chinopoulos C, Kepp O, Kroemer G, Galluzzi L, et al. Molecular mechanisms of cell death: central implication of ATP synthase in mitochondrial permeability transition. *Oncogene*. (2015) 34:1608. doi: 10.1038/onc.2014.462
22. Lee J-G, Ko J, Hae H, Sòo-Kang J, Do-Kang Y, Lee JH, et al. Intravascular ultrasound-based machine learning for predicting fractional flow reserve in intermediate coronary artery lesions. *Atherosclerosis*. (2020) 292:171–77. doi: 10.1016/j.atherosclerosis.2019.10.022
23. Yin W-J, Yi YH, Guan XF, Zhou LY, Wang JL, Li Y, et al. Preprocedural prediction model for contrast-induced nephropathy patients. *J Am Heart Assoc*. (2017) 6:004498. doi: 10.1161/JAHA.116.004498
24. Blanc-Durand P, Campedel L, Mule S, Jegou S, Luciani A, Pigneur F, et al. Prognostic value of anthropometric measures extracted from whole-body CT using deep learning in patients with non-small-cell lung cancer. *Eur Radiol*. (2020). doi: 10.1007/s00330-019-06630-w. [Epub ahead of print].
25. Hawkins SH, Korecki JN, Balagurunathan Y, Gu Y, Kumar V, Basu S, et al. Predicting outcomes of nonsmall cell lung cancer using CT image features. *IEEE Access*. (2014) 2:1418–26. doi: 10.1109/ACCESS.2014.2373335
26. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep*. (2015) 5:13087. doi: 10.3389/fonc.2015.00272
27. Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett*. (2017) 403:21–7. doi: 10.1016/j.canlet.2017.06.004
28. Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal Carcinoma. *Clin Cancer Res*. (2017) 23:4259–69. doi: 10.1158/1078-0432.CCR-16-2910
29. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, et al. Exploratory study to identify radiomics classifiers for lung Cancer histology. *Front Oncol*. (2016) 6:71. doi: 10.3389/fonc.2016.00071
30. Yin P, Mao N, Zhao C, Wu J, Sun C, Chen L, et al. Comparison of radiomics machine-learning classifiers and feature selection for differentiation of sacral chordoma and sacral giant cell tumour based on 3D computed tomography features. *Eur Radiol*. (2018) 29:1841–7. doi: 10.1007/s00330-018-5730-6
31. Wu S, Zheng J, Li Y, Yu H, Shi S, Xie W, et al. A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder Cancer. *Clin Cancer Res*. (2017) 23:6904–11. doi: 10.1158/1078-0432.CCR-17-1510
32. Ortega-Martorell S, Olier I, Julia-Sape M, Arus C. SpectraClassifier 1.0: a user friendly, automated MRS-based classifier-development system. *BMC Bioinform*. (2010) 11:106. doi: 10.1186/1471-2105-11-106

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 19 June 2020
doi: 10.3389/fonc.2020.00918

Applying Amide Proton Transfer-Weighted Imaging (APTWI) to Distinguish Papillary Thyroid Carcinomas and Predominantly Solid Adenomatous Nodules: Comparison With Diffusion-Weighted Imaging

Guomin Li[1,2†], Guihua Jiang[2†], Yingjie Mei[3†], Peng Gao[4], Ruijian Liu[4], Min Jiang[4], Yue Zhao[4], Meng Li[2], Yunfan Wu[2], Shishun Fu[1], Mengchen Liu[1], Liming Li[2], Wuming Li[2] and Jianhao Yan[2]*

[1] The Second School of Clinical Medicine, Southern Medical University, Guangzhou, China, [2] The Department of Medical Imaging, Guangdong Second Provincial General Hospital, Guangzhou, China, [3] Philips Healthcare, Hong Kong, China, [4] Department of General Surgery, Guangdong Second Provincial General Hospital, Guangzhou, China

OPEN ACCESS

Edited by:
Xuelei Ma,
Sichuan University, China

Reviewed by:
Ahmed Abdel Razek,
Mansoura University, Egypt
Rita G. Nunes,
Instituto Superior Técnico, Portugal
Yi Zhang,
Zhejiang University, China

***Correspondence:**
Jianhao Yan
yanjianhao@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 23 October 2019
Accepted: 11 May 2020
Published: 19 June 2020

Citation:
Li G, Jiang G, Mei Y, Gao P, Liu R, Jiang M, Zhao Y, Li M, Wu Y, Fu S, Liu M, Li L, Li W and Yan J (2020) Applying Amide Proton Transfer-Weighted Imaging (APTWI) to Distinguish Papillary Thyroid Carcinomas and Predominantly Solid Adenomatous Nodules: Comparison With Diffusion-Weighted Imaging. Front. Oncol. 10:918. doi: 10.3389/fonc.2020.00918

Background: Amide proton transfer-weighted (ATPw) imaging is a novel MRI technique that has been used to identify benign and malignant tumors. The present study evaluated the role of APTw imaging in differentiating papillary thyroid carcinoma from predominantly solid adenomatous nodule.

Methods: This study included 24 cases of solitary papillary thyroid carcinoma, and 20 cases of solid adenomatous nodules. Normal thyroid tissues were examined in 12 healthy subjects. The healthy subjects, eight cases of adenomatous nodule with cystic degeneration, and 12 cases of thyroid goiter, were only considered in the descriptive analysis, not included in our statistical analysis. The mean APTw value and the apparent diffusion coefficients (ADCs) of papillary thyroid carcinoma and solid adenomatous nodule were compared via a Mann-Whitney U test and receiver operating characteristic (ROC)-curve analyses.

Results: The adenomatous nodule (3.3 ± 1.3%) exhibited significantly higher APTw value ($p < 0.05$) than that of the papillary thyroid carcinoma (1.8 ± 0.7%). The optimal cut-off value of the mean APTw value in differentiating papillary thyroid carcinoma from adenomatous nodule was 3.15%, with a sensitivity of 60% and a specificity of 100%. The mean ADC of papillary thyroid carcinoma ($1.2 \pm 0.2 \times 10^{-3}$ mm^2/s) was significantly lower than that of adenomatous nodule ($2.0 \pm 0.4 \times 10^{-3}$ mm^2/s). The optimal cut-off value of the mean ADC was 1.35×10^{-3} mm^2/s, with a sensitivity of 100% and a specificity of 75%. Based on the ROC-curve analysis of APT and ADC, the ADC showed a higher area under the curve (AUC) than that of APT ($AUC_{APT} = 0.84$, $AUC_{ADC} = 0.95$).

Conclusion: APTw imaging may be as useful as DWI for the differentiation of papillary thyroid carcinoma from predominantly solid adenomatous nodule. Although the sensitivity of ADC was greater than that of APT, APT had greater specificity.

Keywords: papillary thyroid carcinoma, predominantly solid thyroid adenomatous nodule, amide proton transfer (APT), diffusion-weighted imaging (DWI), differentiation

INTRODUCTION

Thyroid nodules are becoming increasingly prevalent. Nodular goiters and adenomas are the most common benign thyroid nodules, and papillary thyroid carcinoma is the most common malignant thyroid tumors (1). Nodular goiters and adenomas are usually treated by clinical observation, especially in the elderly. In contrast, the optimal treatment for papillary thyroid carcinoma is surgical excision. Therefore, the precise preoperative differentiation of nodular goiter or adenoma and papillary thyroid carcinoma is of significant practical relevance. Adenomas can occur alone or in combination with nodular goiters. Their morphologies, signals, and enhancements are similar, often resulting in difficult differential diagnoses. In particular, solitary solid nodular goiters are challenging to identify with adenomas. As such, we used adenomatous nodules (2–4) to replace solitary solid nodular goiters or thyroid adenomas in the present study.

Amide proton transfer-weighted (ATPw) imaging is a novel magnetic resonance imaging (MRI) technique that can detect mobile proteins and peptides that contain abundant amide (-CO-NH-) chemical constituents (5, 6). The APTw values can reflect the concentrations of mobile macromolecules, such as proteins and peptides. Early reports of APTw imaging for cancer assessment have focused on the brain. According to the previous literature, high-grade gliomas show higher APTw values than low-grade gliomas (7, 8), and APTw imaging is useful for assessing tumor aggressiveness. Investigators in recent human studies have reported preliminary APT findings in the breast (9), prostate (10), cervix (11), rectum (12), and lung (13). APTw values were higher in cancers than in normal tissues or benign tumors, and APT levels varied between different malignant tumors groups or different histological grades. Furthermore, APT may provide additional information to improve the results of diffusion-weighted imaging (DWI) or other MRI techniques.

The head-neck regions are challenging for molecular MRI techniques because of magnetic field inhomogeneity, and motion and such tissues are prone to artifacts. In a preliminary study on the characterization of head and neck tumors which showed the feasibility of performing APTw imaging in the head and neck by using a technique adapted from the brain, the authors hypothesized that malignant tumors have higher APT levels than healthy tissues and benign tumors and that APT levels differ among malignant tumor groups. They studied the patients with nasopharyngeal undifferentiated carcinoma, squamous cell carcinoma, non-Hodgkin's lymphoma, and benign salivary gland tumors (14, 15).

We previously reported on a study about patients with thyroid tumors that showed the feasibility of performing APTw imaging in the neck. The results showed that the APTw values of malignant nodules of the thyroid are lower than that of benign nodules, which is different from other tumors (16). However, thyroid tumors are prone to cystic change (17), which have a significant influence on the measurement of APTw values. Our previous study samples were simply divided into benign groups and malignant groups. Both the two groups contained different pathological types, and cystic nodules were not excluded. We want to explore the diagnostic performance of APTw imaging in differentiating papillary thyroid carcinoma from predominantly solid adenomatous nodule. Now we need to further group and measure them accurately, calculate the threshold, sensitivity, and specificity of APT and ADC to distinguish solid papillary thyroid carcinoma and solid adenomatous nodule.

MATERIALS AND METHODS

Subjects

The local Institutional Review Board approved this study, and all subjects gave written, informed consent before participation in this study. Between 2018 and 2019, 24 biopsy-proven papillary thyroid carcinomas, 28 cases of adenomatous nodule, and 12 cases of thyroid goiter underwent MRI exam. This study included 12 healthy subjects. The healthy subjects, 8 cases of adenomatous nodules with cystic degeneration, and 12 cases of thyroid goiter were only considered in the descriptive analysis, not included in statistical analysis. Thus, 24 papillary carcinomas (15 females, 9 males; 41.16 ± 13.43 years old; range, 29–68 years old) and 20 adenomatous nodules (13 females, 7 males; 42.80 ± 10.20 years old; range, 22–72 years old) were included in the study.

MRI Protocols

MR imaging was performed with a Philips 3-Tesla (3T) scanner (Ingenia, 3.0 T; Philips Medical Systems, The Netherlands). A 16-channel head-neck coil was used for scanning. The patients underwent T1- [repetition time (TR)/echo time (TE), 570/18 ms] and T2-weighted MR imaging [TR (ms)/TE (ms), 2,500/100] with the section thickness of 4 mm, an intersection gap of 1 mm, field-of-views of 20–25 cm, and an acquisition matrix of 256 × 224. The scan time of T1WI is 85 s and the scan time of T2WI is 150 s. Images were obtained in axial and coronal planes, following scout images in the sagittal plane.

In addition to conventional MR imaging (T1-weighted imaging, T2-weighted imaging, and Gd-enhanced T1-weighted imaging), APTw sequences and reduced field-of-view (r-FOV) diffusion-weighted sequences with different b values (0, 800 mm^2/s) were acquired. Other parameters of DWI were as follows: field-of-views of 116 × 51 mm^2; voxel size of 1.81 × 1.81 mm^2; slice thickness of 4 mm; TR (ms)/TE (ms) of 3,687/62; scan time

of 221 s. APTw imaging was performed using a 3-dimensional (3D) turbo-spin-echo Dixon sequence with these parameters as follows: slice thickness of 4.4 mm, acquisition voxel size of 1.8 × 1.8 mm^2, TR (ms)/TE (ms) of 4,108/5.9, scan time of 259 s, and turbo spin-echo factor of 158. APTw imaging was performed with seven saturation-frequency offsets (offsets = ± 2.7, ± 3.5, ± 4.3 ppm, and 1,540 ppm). The protocol was repeated three times at ±3.5 ppm to increase the signal-to-noise ratio within an appropriate time frame. Saturation radio-frequency pulses for APTw imaging were implemented with an amplitude of 2 μT and a duration of 2 s. B_0 maps were obtained with three acquisitions at 3.5 ppm of different echo times. B0-corrected ATPw images were reconstructed online.

Imaging Analysis of APT and Apparent Diffusion Coefficients (ADCs)

The two radiologists conducting the present study determined by consensus whether the APTw maps and ADC maps were acceptable for statistical analysis. All images were interpreted by two radiologists specializing in head and neck imaging. APTw and ADC imaging were automatically generated via a Philips post-processing workstation. We calculated the mean APTw value and ADC value by drawing a region of interest (ROI). The radiologists drew an ROI around the predominantly solid thyroid nodules or drew a ROI on the central of one leaf of the normal thyroid tissues on the APTw image and ADC map by using the T2WI for reference, and then the mean APTw value and mean ADC value was obtained from the ROI, as shown in **Table 2**. The ROI analysis was repeated by two observers to assess the inter-observer agreement. The two radiologists processed the MR images independently. They were blinded to the histopathologic data.

Statistical Analysis

The APTw values and ADC values of the papillary thyroid carcinoma were compared with that of the thyroid adenomatous nodules using a Mann-Whitney U test. The diagnostic performances of significant APTw parameters for differentiating the papillary thyroid carcinoma from the adenomatous nodules were assessed by using ROC-curve analyses with the AUC. The APTw threshold was acquired by calculating the Youden index, which is the sum of the sensitivity and specificity −1, and the APTw value corresponding to the point where the Yoden index is the largest was considered the APTw threshold. Then the sensitivity, specificity of the optimal thresholds were calculated. Statistical analysis was performed using SPSS software 21.0. All statistical tests were two-sided, and a p-value of <0.05 was considered to indicate a statistically significant difference.

RESULTS

The characteristics of the patients are shown in **Table 1** and the subjects selection flowchart is shown in **Figure 1**. We first assessed the radiographic features of some interesting cases and normal thyroids using several standards sequences (T1-weight images, T2-weight images, Gd- T1-weight images, DWI) and APTw sequences. **Figure 2** shows the normal thyroid tissue and diffuse goiter. They appear homogenously isointense on APTWI, and their APTw values (normal thyroid, 2.15%; diffuse goiter, 2.36%) are similar, and neither is very high. **Figure 3** shows two thyroid nodules with cystic changes. The A cyst rich in serous fluid and appeared hypointense on T1-weight images (T1WI), hyperintense on T2-weight images (T2WI), and hyperintense on APTWI (APTw values = 7.33%). The B cyst is rich in thyroid colloid and appears hyperintense on T1WI, hypointense on T2WI, hypointense on APTWI (APTw values = 1.53%). The solid portion appears isointense on T1WI and hyperintense on T2WI and APTWI (APTw values = 3.56%).

Figure 4 shows two predominantly solid adenomatous nodules. One is an atypical adenomatous nodule and appeared mild enhancement on Gd-T1WI, mild hyperintense on an ADC map, and isointense on APTWI (APTw values = 2.05%). The other is a typical adenomatous nodule and exhibited strong enhancement on Gd-T1WI and hyperintense on the ADC map and APTWI (APTw values = 5.21%).

Figure 5 shows a typical solid adenomatous nodule and a papillary thyroid carcinoma. The adenomatous nodule appeared hyperintense on both the ADC map and T2WI, and the mean APTw value was 6.10%. The papillary thyroid carcinoma appeared hypointense on ADC map, heterogeneous iso-/hypo-intensity on APTWI, and the mean APTw value was 1.93%.

The intraclass correlation coefficients (ICC) showed excellent observer agreement ($ICC_{APT} = 0.92$, $ICC_{ADC} = 0.96$, $p < 0.01$). **Table 2** and **Figure 6** show the APTw values and ADC values of thyroid nodules in this study, and there was a significant difference in the APTw value and ADC of the papillary thyroid carcinoma and adenomatous nodule. The adenomatous nodule (3.3 ± 1.3%) exhibited higher APT-weighted signal intensities than that of papillary carcinoma (1.8 ± 0.7%; $p < 0.01$). The mean ADC of the papillary thyroid carcinoma (1.2 ± 0.2 × 10^{-3} mm^2/s) was significantly lower than that of the adenomatous nodule (2.0 ± 0.4 × 10^{-3} mm^2/s; $p < 0.01$). The optimal cut-off value of the mean APTw value in differentiating papillary thyroid carcinoma from the adenomatous nodule was 3.15%, with a sensitivity of 60% and a specificity of 100%

TABLE 1 | Patient characteristics and pathologies.

Pathology	No.	Female:male	Age (years)
Adenomatous nodule	20	13:7	43 ± 10
Papillary carcinoma	24	15:9	41 ± 13
Total	44	28:16	42 ± 12

TABLE 2 | The APTw values and ADC values of thyroid nodules.

Pathology	APTw value (%)	ADC (mm^2/s)	Diameter (mm)	p-value
Adenomatous nodule	3.3 ± 1.3	2.0 ± 0.4	24 ± 9	<0.001
Papillary carcinoma	1.8 ± 0.7	1.2 ± 0.2	11 ± 5	

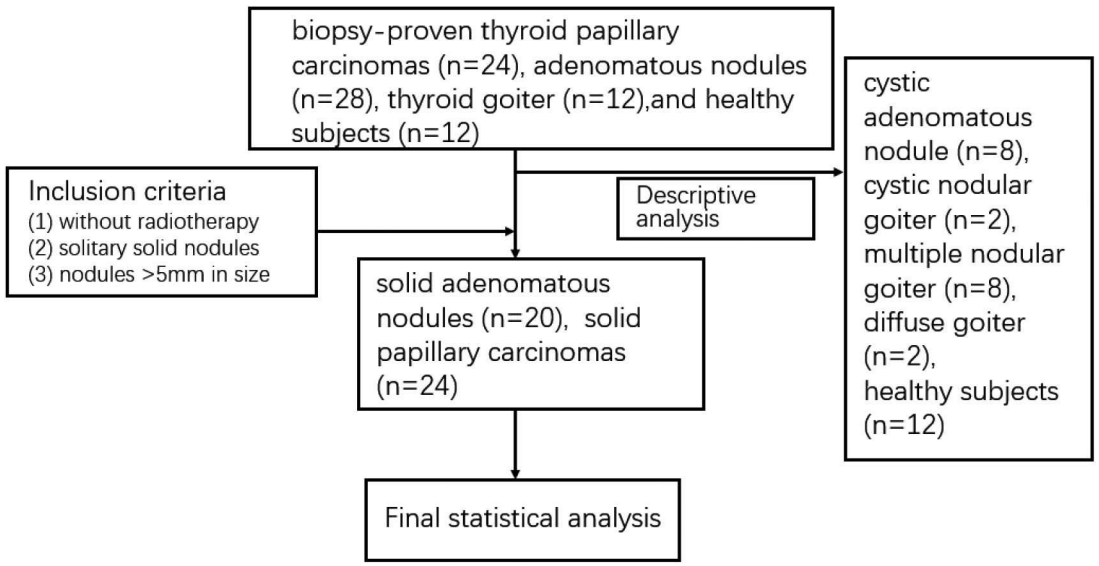


FIGURE 1 | Subject selection flowchart.

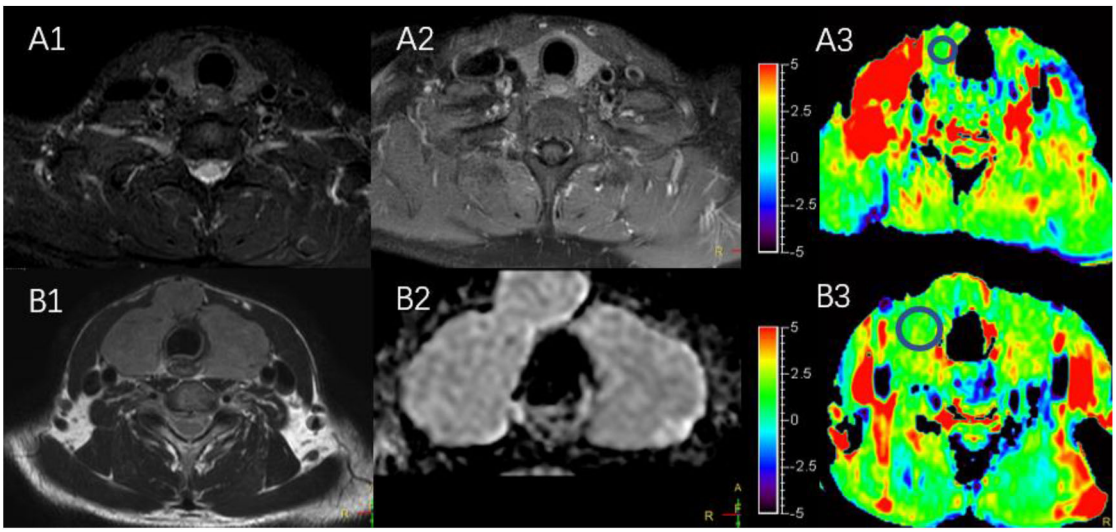


FIGURE 2 | (A1–A3) MR images of a normal thyroid, including T2WI **(A1)**, Gd-T1WI **(A2)**, and APTWI **(A3)**. **(B1–B3)** MR images of diffuse goiter, including T2WI **(B1)**, ADC map **(B2)**, and APTWI **(B3)**.

(**Figure 7**). The mean ADC of the papillary thyroid carcinoma was significantly lower than that of the adenomatous nodule. The optimal cut-off value of the mean ADC in differentiating papillary carcinoma from adenomatous nodule was 1.35×10^{-3} mm^2/s, with a sensitivity of 100% and specificity of 75%. The ROC curve analysis revealed that ADC exhibited a higher AUC value compared to that of APT ($AUC_{APT} = 0.84$, $AUC_{ADC} = 0.95$). The r-FOV DWI showed a better diagnostic performance than that of APTw imaging. Although the sensitivity of DWI (100%) was significantly higher than that of APT (60%), the specificity of APT (100%) was substantially higher than that of ADC (75%).

DISCUSSION

In this study, we explored the diagnostic performance of using APTw imaging to differentiate papillary thyroid carcinoma from the solid adenomatous nodule. The aim was to differentiate papillary thyroid carcinoma from adenomatous nodule so that the patients with papillary thyroid carcinoma would be able to receive appropriate treatment at an earlier stage while avoiding unnecessary surgery in the patients with adenomatous nodules. The present study showed a significant difference between the ADC and APTw value of papillary thyroid carcinoma and adenomatous nodule, in which the most adenomatous nodules

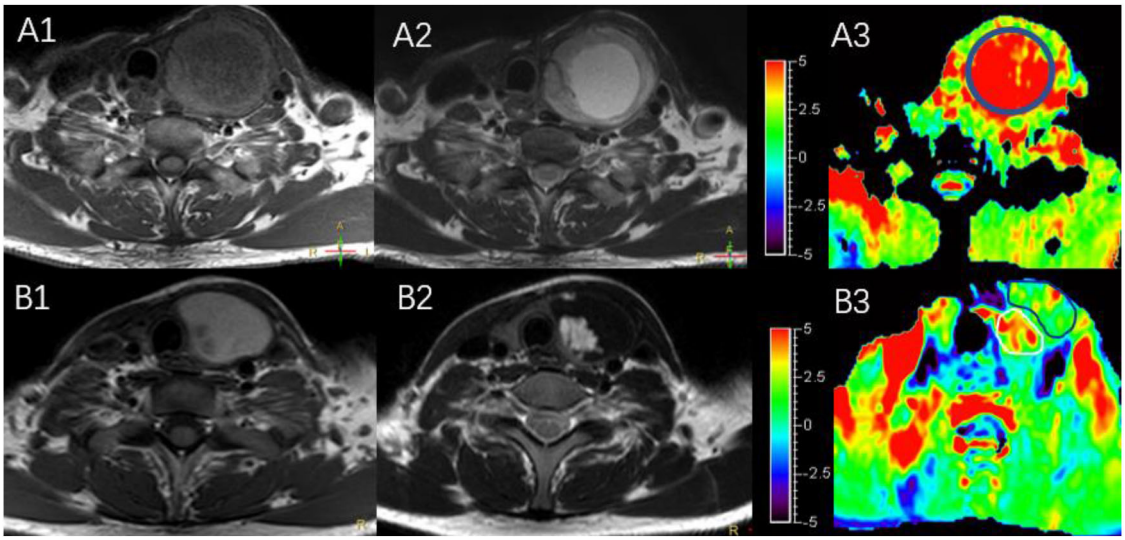


FIGURE 3 | MR images of two predominantly cystic thyroid nodules, including T1WI **(A1,B1)**, T2WI **(A2,B2)**, and APTWI **(A3,B3)**.

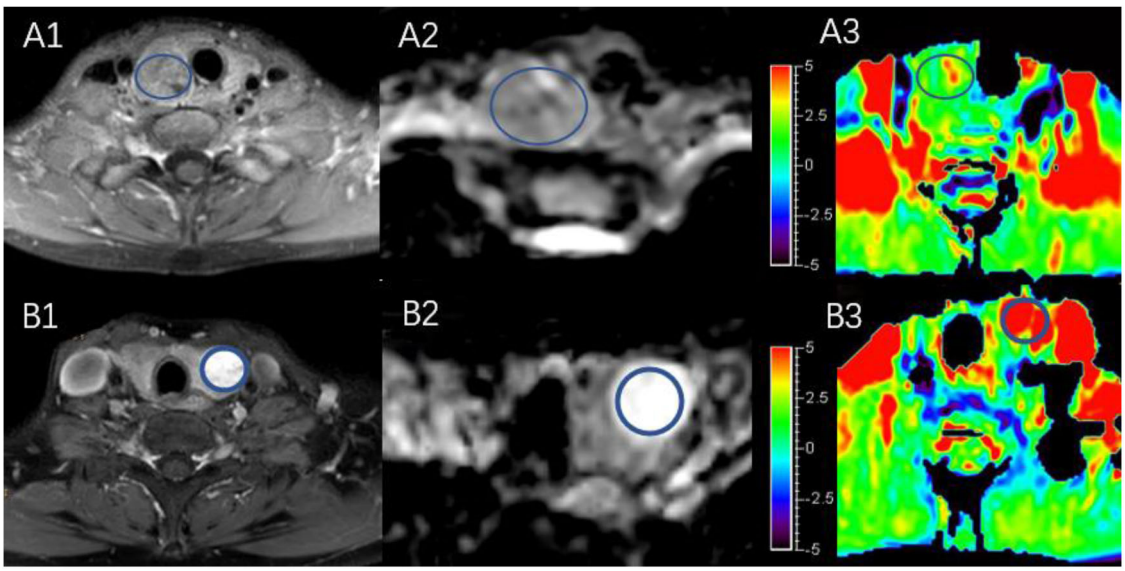


FIGURE 4 | Both **(A,B)** are predominantly solid adenomatous nodules on Gd-T1WI **(A1,B1)**, ADC map **(A2,B2)**, and APTWI **(A3,B3)**.

had higher mean ADC and APTw value than papillary thyroid carcinoma. It is not clear why adenomatous nodules have higher APTw value than papillary thyroid carcinoma, as opposed to other tumors.

Diffusion-weighted imaging provides a better characterization of tissues because it can reflect the random motion of water molecules, which is disturbed by intracellular macromolecules. Previous studies have evaluated the role of diffusion-weighted imaging in differentiating benign from malignant thyroid nodules (18–24). The APTw values can reflect the concentrations of mobile macromolecules, such as proteins and peptides. Our previous studies on the thyroid established a positive correlation between APTw values with ADC. Part of the reason for this may be because the APTWI detects free protein rather than solid proteins.

As shows in **Figure 2**, the APTw value of normal thyroid tissue and diffuse goiters are similar, and neither is very high despite relatively abundant colloid components in the diffuse goiter. The components of the cystic thyroid zone consist mainly of serous fluid, thyroid colloid (thyroglobulin), and blood from different periods, and they exhibit characteristic MR signals (25, 26). Serous fluid often appears hypointense on T1WI and hyperintense on T2WI, similar to that of water. Thyroid colloid contains macromolecular thyroglobulin, which

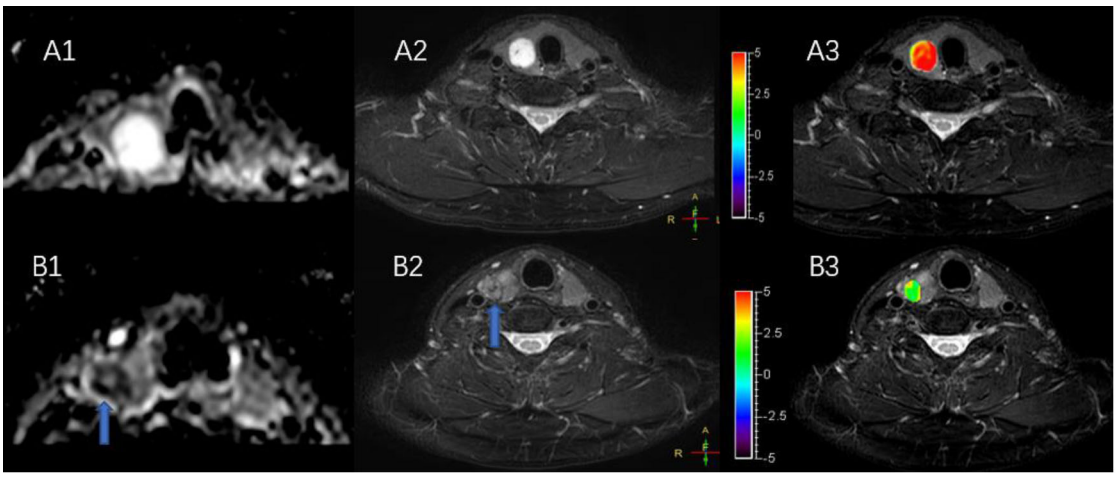


FIGURE 5 | (A,B) Show a solid adenomatous nodule and a papillary thyroid carcinoma on the ADC map **(A1,B1)**, T2WI **(A2,B2)**, and the combination of T2WI and APTWI **(A3,B3)**.

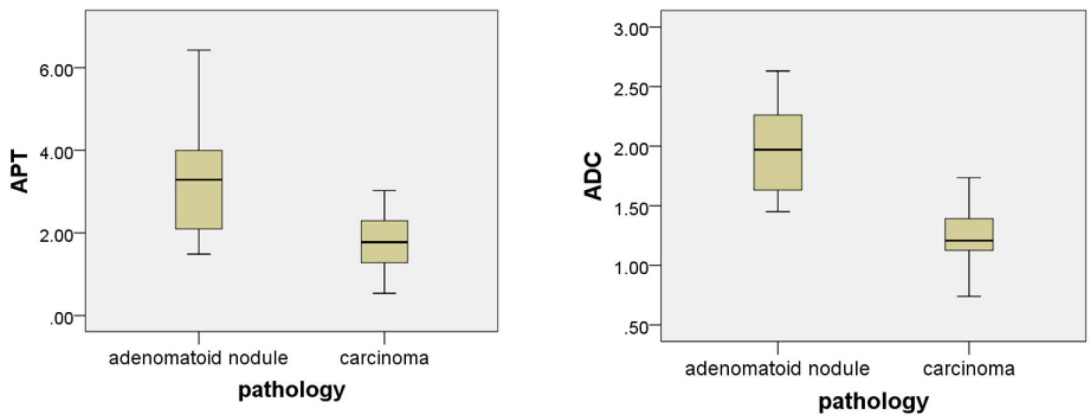


FIGURE 6 | Box plot of the APT (%) and the ADC (mm^2/s) of adenomatous nodule and papillary thyroid carcinoma.

shortens the T1 relaxation time and shows a homogenous high signal on T1WI. Blood fluid from different periods can display various heterogeneous signal intensities (27, 28). **Figure 3** exhibits a thyroid nodule with the cystic change, and the components of the cystic thyroid zone consist mainly of thyroid colloid, but the APTw value is low. It is speculated that the thyroid colloid does not show a high signal intensity on APTWI, and the reasons why the APTw value of adenomatous nodule was higher than that of papillary thyroid carcinoma is not that adenomatous nodule contains abundant thyroid colloid.

In the present study, most solid adenomatous nodules showed significantly high APTw value, but some were similar to normal thyroid tissue. The typical adenomatous nodule that showed high signal on the APTw image exhibited isointense on T1WI, hyperintensity on T2WI, and strong enhancement on Gd-T1WI, indicating that there is abundant microvessel on the typical adenomatous nodule. The typical adenomatous nodule had a high ADC value, indicating active water-molecule movement. On the contrary, the atypical adenomatous nodule exhibited isointense on the ADC map and slight enhancement on Gd-T1WI, indicating the restricted water-molecule movement and the less microvessel compared with typical adenomatous nodule (29). The blood supply of papillary thyroid carcinoma is not as abundant as in typical adenomatous nodule, and papillary thyroid carcinoma has a high density of tumor cells, small extracellular space, and high cytoplasmic viscosity (21, 30–34). In conclusion, abundant blood supply may underlie why adenomatous nodule has higher APTw value than papillary thyroid carcinoma.

The present study had some limitations. First, the sample size was small. Second, the head and neck are challenging regions in which to perform functional MRIs because of field inhomogeneity, relatively low signal-to-noise ratio, movement artifacts, and difficulties with imaging fat suppression. Third, some thyroid microcarcinomas may occur in adenomatous

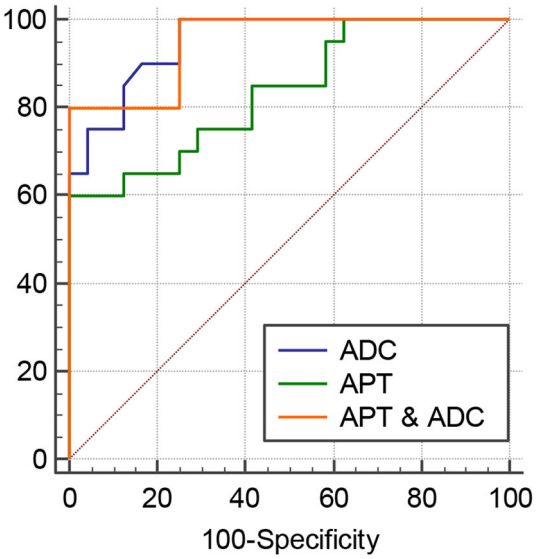


FIGURE 7 | ROC curves of APT, ADC, and the combination of APT and ADC for differentiation between papillary thyroid carcinoma and adenomatous nodule: the AUC was 0.95 (ADC), 0.95 (APT and ADC), and 0.835 (APT).

nodules. In addition, there is some biases because the ROIs were drawn manually on the APTWI and ADC maps by using the anatomic images for reference. The APT value of adenomatous nodules is not absolutely higher than that of papillary thyroid carcinoma, but it is because most papillary carcinomas are relatively small when they are found. At this stage, the papillary carcinoma has incompletely developed blood vessels and relatively less blood vessels. If the supply of blood vessels to the papillary cancer in the late stage becomes rich, then like other malignant tumors, the APT value of papillary thyroid carcinoma will increase and close to the adenomatous nodule.

CONCLUSIONS

APTw imaging may be useful for the differentiation of papillary thyroid carcinoma from predominantly solid adenomatous nodule. DWI had higher accuracy and sensitivity but lower specificity than APTw imaging. From our present results, we hypothesize that plentiful blood supply may be the main reason why the APTw value of the typical adenomatous nodule is higher than that of papillary thyroid carcinoma.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of Guangdong Second Provincial General Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GJ designed the experiment. GL and RL carried out the experiment. LL, WL, PG, YZ, MJ, and MLi collected and sorted out the data. YW, SF, and MLiu helped on data management and processing. JY and GL wrote the manuscript. YM revised the manuscript.

FUNDING

This study was funded by the National Natural Science Foundation of China (Grant number: 81701111), the Science and Technology Planning Project of Guangdong Province, China (2017045), and Medical Science and Technology Foundation of Guangdong Province (A2016411).

REFERENCES

1. Schob S, Voigt P, Bure L, Meyer HJ, Wickenhauser C, Behrmann C, et al. Diffusion-weighted imaging using a readout-segmented, multishot EPI sequence at 3 T distinguishes between morphologically differentiated and undifferentiated subtypes of thyroid carcinoma–a preliminary study. *Transl Oncol.* (2016) 9:403–10. doi: 10.1016/j.tranon.2016.09.001
2. Derwahl M, Studer H. Hyperplasia versus adenoma in endocrine tissues: are they different? *Trends Endocrinol Metab.* (2002) 13:23–8. doi: 10.1016/S1043-2760(01)00519-7
3. Qi L, Xue K, Li C, He W, Mao D, Xiao L, et al. Analysis of CT morphologic features and attenuation for differentiating among transient lesions, atypical adenomatous hyperplasia, adenocarcinoma *in situ*, minimally invasive and invasive adenocarcinoma presenting as pure ground-glass nodules. *Sci Rep.* (2019) 9:14586. doi: 10.1038/s41598-019-50989-1
4. Schreiner AM, Yang GC. Adenomatoid nodules are the main cause for discrepant histology in 234 thyroid fine-needle aspirates reported as follicular neoplasm. *Diagn Cytopathol.* (2012) 40:375–9. doi: 10.1002/dc.21499
5. Zhou J, Lal B, Wilson DA, Laterra J, van Zijl PC. Amide proton transfer (APT) contrast for imaging of brain tumors. *Magn Reson Med.* (2003) 50:1120–6. doi: 10.1002/mrm.10651
6. Zhou J, Blakeley JO, Hua J, Kim M, Laterra J, Pomper MG, et al. Practical data acquisition method for human brain tumor amide proton transfer (APT) imaging. *Magn Reson Med.* (2008) 60:842–9. doi: 10.1002/mrm.21712
7. Zhou J, Zhu H, Lim M, Blair L, Quinones-Hinojosa A, Messina SA, et al. Three-dimensional amide proton transfer MR imaging of gliomas: Initial experience and comparison with gadolinium enhancement. *J Magn Reson Imaging.* (2013) 38:1119–28. doi: 10.1002/jmri.24067
8. Bai Y, Lin Y, Zhang W, Kong L, Wang L, Zuo P, et al. Noninvasive amide proton transfer magnetic resonance imaging in evaluating the grading and cellularity of gliomas. *Oncotarget.* (2017) 8:5834–42. doi: 10.18632/oncotarget.13970
9. Dula AN, Arlinghaus LR, Dortch RD, Dewey BE, Whisenant JG, Ayers GD, et al. Amide proton transfer imaging of the breast at 3 T: establishing reproducibility and possible feasibility assessing chemotherapy response. *Magn Reson Med.* (2013) 70:216–24. doi: 10.1002/mrm.24450
10. Takayama Y, Nishie A, Sugimoto M, Togao O, Asayama Y, Ishigami K, et al. Amide proton transfer (APT) magnetic resonance imaging of prostate cancer: comparison with Gleason scores. *MAGMA.* (2016) 29:671–9. doi: 10.1007/s10334-016-0537-4
11. Li B, Sun H, Zhang S, Wang X, Guo Q. Amide proton transfer imaging to evaluate the grading of squamous cell carcinoma of the cervix: a comparative

study using (18) F FDG PET. *J Magn Reson Imaging.* (2019) 50:261–8. doi: 10.1002/jmri.26572
12. Nishie A, Takayama Y, Asayama Y, Ishigami K, Ushijima Y, Okamoto D, et al. Amide proton transfer imaging can predict tumor grade in rectal cancer. *Magn Reson Imaging.* (2018) 51:96–103. doi: 10.1016/j.mri.2018.04.017
13. Ohno Y, Kishida Y, Seki S, Yui M, Miyazaki M, Koyama H, et al. Amide proton transfer-weighted imaging to differentiate malignant from benign pulmonary lesions: comparison with diffusion-weighted imaging and FDG-PET/CT. *J Magn Reson Imaging.* (2018) 47:1013–21. doi: 10.1002/jmri.25832
14. Law B, King AD, Ai QY, Poon D, Chen W, Bhatia KS, et al. Head and neck tumors: amide proton transfer MRI. *Radiology.* (2018) 288:782–90. doi: 10.1148/radiol.2018171528
15. Yuan J, Chen S, King AD, Zhou J, Bhatia KS, Zhang Q, et al. Amide proton transfer-weighted imaging of the head and neck at 3 T: a feasibility study on healthy human subjects and patients with head and neck cancer. *NMR Biomed.* (2014) 27:1239–47. doi: 10.1002/nbm.3184
16. Liu R, Jiang G, Gao P, Li G, Nie L, Yan J, et al. Non-invasive amide proton transfer imaging and ZOOM diffusion-weighted imaging in differentiating benign and malignant thyroid micronodules. *Front Endocrinol (Lausanne).* (2018) 9:747. doi: 10.3389/fendo.2018.00747
17. Kim YJ, Baek JH, Ha EJ, Lim HK, Lee JH, Sung JY, et al. Cystic versus predominantly cystic thyroid nodules: efficacy of ethanol ablation and analysis of related factors. *Eur Radiol.* (2012) 22:1573–8. doi: 10.1007/s00330-012-2406-5
18. Wang Q, Guo Y, Zhang J, Shi L, Ning H, Zhang X, et al. Utility of high b-value (2000 sec/mm^2) DWI with RESOLVE in differentiating papillary thyroid carcinomas and papillary thyroid microcarcinomas from benign thyroid nodules. *PLoS ONE.* (2018) 13:e0200270. doi: 10.1371/journal.pone.0200270
19. Razek AA, Sadek AG, Kombar OR, Elmahdy TE, Nada N. Role of apparent diffusion coefficient values in differentiation between malignant and benign solitary thyroid nodules. *AJNR Am J Neuroradiol.* (2008) 29:563–8. doi: 10.3174/ajnr.A0849
20. Noda Y, Kanematsu M, Goshima S, Kondo H, Watanabe H, Kawada H, et al. MRI of the thyroid for differential diagnosis of benign thyroid nodules and papillary carcinomas. *AJR Am J Roentgenol.* (2015) 204:W332–5. doi: 10.2214/AJR.14.13344
21. Ekinci O, Boluk SE, Eren T, Ozemir IA, Boluk S, Salmaslioglu A, et al. Diffusion-weighted magnetic resonance imaging for the detection of thyroid cancer. *Cir Esp.* (2018) 96:620–6. doi: 10.1016/j.cireng.2018.10.002
22. Abdel RA, Sadek AG, Gaballa G. Diffusion-weighed MR of the thyroid gland in Graves' disease: assessment of disease activity and prediction of outcome. *Acad Radiol.* (2010) 17:779–83. doi: 10.1016/j.acra.2010.01.014
23. Abdel RA, Abd AS, El-Said A. Role of diffusion-weighted magnetic resonance (MR) imaging in differentiation between Graves' disease and painless thyroiditis. *Pol J Radiol.* (2017) 82:536–41. doi: 10.12659/PJR.902416
24. Abdel RA. Routine and advanced diffusion imaging modules of the salivary glands. *Neuroimaging Clin N Am.* (2018) 28:245–54. doi: 10.1016/j.nic.2018.01.010
25. Tennvall J, Olsson M, Moller T, Akerman M, Ranstam J, Biorklund A, et al. Thyroid tissue characterization by proton magnetic resonance relaxation time determination. *Acta Oncol.* (1987) 26:27–32. doi: 10.3109/02841868709092973
26. Noma S, Kanaoka M, Minami S, Sagoh T, Yamashita K, Nishimura K, et al. Thyroid masses: MR imaging and pathologic correlation. *Radiology.* (1988) 168:759–64. doi: 10.1148/radiology.168.3.3406406
27. Beomonte ZB, Cardone G, Tella S, Innacoli M, Cisternino S, Calvisi G, et al. Magnetic resonance in the diagnosis of thyroid diseases. *Radiol Med.* (1992) 84:36–42.
28. Kusunoki T, Murata K, Nishida S, Tomura T, Inoue M. Histopathological findings of human thyroid tumors and signal intensities of magnetic resonance imaging (MRI). *Nihon Jibiinkoka Gakkai Kaiho.* (1994) 97:1406–11. doi: 10.3950/jibiinkoka.97.1406
29. Nachiappan AC, Metwalli ZA, Hailey BS, Patel RA, Ostrowski ML, Wynne DM. The thyroid: review of imaging features and biopsy techniques with radiologic-pathologic correlation. *Radiographics.* (2014) 34:276–93. doi: 10.1148/rg.342135067
30. Vermoolen MA, Kwee TC, Nievelstein RA. Apparent diffusion coefficient measurements in the differentiation between benign and malignant lesions: a systematic review. *Insights Imaging.* (2012) 3:395–409. doi: 10.1007/s13244-012-0175-y
31. Hao Y, Pan C, Chen W, Li T, Zhu W, Qi J. Differentiation between malignant and benign thyroid nodules and stratification of papillary thyroid cancer with aggressive histological features: whole-lesion diffusion-weighted imaging histogram analysis. *J Magn Reson Imaging.* (2016) 44:1546–55. doi: 10.1002/jmri.25290
32. Schmidt H, Schwenzer NF, Gatidis S, Kustner T, Nikolaou K, Schick F, et al. Systematic evaluation of amide proton chemical exchange saturation transfer at 3 T: effects of protein concentration, pH, and acquisition parameters. *Invest Radiol.* (2016) 51:635–46. doi: 10.1097/RLI.0000000000000292
33. Gupta N, Norbu C, Goswami B, Chowdhury V, Ravishankar L, Gulati P, et al. Role of dynamic MRI in differentiating benign from malignant follicular thyroid nodule. *Auris Nasus Larynx.* (2011) 38:718–23. doi: 10.1016/j.anl.2011.02.002
34. Aghaghazvini L, Sharifian H, Yazdani N, Hosseiny M, Kooraki S, Pirouzi P, et al. Differentiation between benign and malignant thyroid nodules using diffusion-weighted imaging, a 3-T MRI study. *Indian J Radiol Imaging.* (2018) 28:460–4. doi: 10.4103/ijri.IJRI_488_17

Conflict of Interest: YM was employed by the company Philips Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 27 July 2020
doi: 10.3389/fonc.2020.01121

Radiomics Based on CECT in Differentiating Kimura Disease From Lymph Node Metastases in Head and Neck: A Non-Invasive and Reliable Method

Ying Zhang[1,2], *Shujing Yu*[2], *Li Zhang*[2] *and Liqing Kang*[3]*

[1] Graduate School, Tianjin Medical University, Tianjin, China, [2] Department of CT Diagnosis, Cangzhou Central Hospital, Cangzhou, China, [3] Department of Magnetic Resonance Imaging, Cangzhou Central Hospital, Cangzhou, China

Background: Kimura disease may be easily misdiagnosed as malignant tumors such as lymph node metastases based on imaging and clinical symptoms. The aim of this article is to investigate whether the radiomic features and the model based on the features on venous-phase contrast-enhanced CT (CECT) images can distinguish Kimura disease from lymph node metastases in the head and neck.

Methods: A retrospective analysis of 14 patients of head and neck Kimura disease (a total of 38 enlarged lymph nodes) and 39 patients with head and neck lymph node metastases (a total of 39 enlarged lymph nodes), confirmed by biopsy or surgery resection, was conducted. All patients accepted CECT within 10 days before biopsy or surgery resection. Radiomic features based on venous-phase CECT were generated automatically from Artificial-Intelligence Kit (AK) software. All lymph nodes were randomly divided into the training set ($n = 54$) and testing set ($n = 23$) in a ratio of 7:3. ANOVA + Mann–Whitney, Spearman correlation, least absolute shrinkage and selection operator, and Gradient Descent were introduced for the reduction of the highly redundant features. Binary logistic regression model was constructed based on the selected features. Receiver operating characteristic was used to evaluate the diagnostic performance of the features and the model. Finally, a nomogram was established for model application.

Results : Seven features were screened out at the end. Significant difference was found between the two groups for all the features with area under the curves (AUCs) ranging from 0.759 to 0.915. The AUC of the model's identification performance was 0.970 in the training group and 0.977 in the testing group. The disease discrimination efficiency of the model was better than that of any single feature.

Conclusions : The radiomic features and the model based on these features on venous-phase CECT images had very good performance for the discrimination between Kimura disease and lymph node metastases in the head and neck.

Keywords: Kimura disease, lymph node, metastases, radiomics, nomogram, texture analysis, differential diagnosis, CT

OPEN ACCESS

Edited by:
Xuelei Ma,
Sichuan University, China

Reviewed by:
Ahmad Chaddad,
Guilin University of Electronic Technology, China
Oliver Riesterer,
Aarau Cantonal Hospital, Switzerland

***Correspondence:**
Liqing Kang
13333367921@163.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *02 October 2019*
Accepted: *04 June 2020*
Published: *27 July 2020*

Citation:
Zhang Y, Yu S, Zhang L and Kang L (2020) Radiomics Based on CECT in Differentiating Kimura Disease From Lymph Node Metastases in Head and Neck: A Non-Invasive and Reliable Method. Front. Oncol. 10:1121. doi: 10.3389/fonc.2020.01121

INTRODUCTION

Kimura disease, also known as eosinophilic lymphogranuloma, is a rare lymphoproliferative disease, which occurs in the head and neck with unknown origin and expounded systemically in 1948 by Kimura (1). The clinical symptoms of Kimura disease include a painless soft tissue mass, with peripheral lymphadenopathy or lymphadenectasis in the neck and submandibular region (2). The imaging findings of Kimura disease are non-specific; even though the lesions may have some features such as well-defined boundaries, lack of liquefaction necrosis, and calcification or fusion trend, it is still difficult to be distinguished from other lymphadenectasis for some cancer patients with lymph node metastases without symptoms of primary tumors (3–5). The main treatment of Kimura disease is radiotherapy instead of radical surgery, which is preferred for some kind of lymph node metastases in the head and neck. So, it is essential to make an accurate differential diagnosis for clinical intervention. Currently, the diagnosis of Kimura disease is mainly based on the judgment of clinical features and the histopathological examination. However, the clinical judgment is inaccurate, and it is uneconomical and uncomfortable for patients to undergo non-comprehensive sampling and time-consuming and invasive surgical resection or biopsy. An accurate, non-invasive, and efficient method of disease identification is urgently needed.

Radiomics is a newly emerging form of imaging analysis using a series of data mining algorithms or statistical analysis tools on high-throughput imaging features extracted from radiographic data to obtain diagnostic or prognostic information (6–8). By building appropriate models with refined features, successful assessment, and prediction abilities in various challenging clinical tasks can be achieved (9–13). Recent studies of radiomics have provided insights in precision medicine in oncologic practice related to tumor detection, subtype classification, lymph node metastases, survival, and therapeutic response evaluation (14). A review walking through several steps necessary for radiomic analysis in brain tumor in detail showed how it is able to use radiomics in diseases (15). As far as we have known, the application of radiomics for differential lymph node lesions of Kimura disease from lymph node metastases in the head and neck has not been reported in the literature yet.

The purpose of this study was to investigate whether radiomic features extracted from contrast-enhanced CT (CECT) images and the model build on the features could be used for differentiating Kimura disease from metastases.

MATERIALS AND METHODS

General Information

This retrospective study was approved by the ethics committee of Cangzhou Central Hospital, and the informed consent requirement was waived. The research method was in accordance with the standard guidelines and regulations. The clinical histopathologic and radiological data were collected from July 2011 to August 2018. The cohort inclusion criteria were as follows: (a) lymph nodes with histopathologically confirmed Kimura disease and lymph node metastases in head and neck by means of biopsy or surgery resection, (b) patients with CECT performed within 10 days before the pathological examination, and (c) lymph nodes without liquefaction necrosis or calcification with the minimum diameter not <1.0 cm (4, 16). The exclusion criteria were (a) poor image quality with artifacts and (b) patients who had previously received related therapy. Eventually, a total of 77 lymph nodes were included in our study—among them were 14 patients (12 males and two females; mean age, 36.5 years old; range, 16–51 years) with a total of 38 lymph nodes diagnosed as Kimura disease in head and neck. Eight lymph nodes were located in level I, 10 in level II, one in level III, two in level V, and 17 in level VIII. There were 39 patients (20 males and 19 females; mean age, 59.2 years old; range, 30–77 years) with 39 lesions diagnosed as lymph node metastases. Ten lymph nodes were located in level II, eight in level III, 14 in level IV, five in level V, and two in level VI. The level of lymph nodes is defined according to the method described by Gregoire et al. (17). The lymph node metastases were derived from variously sourced cancerous foci (see **Table S1** for detailed information).

CT Image Acquisition

All enrolled patients underwent CECT examination (Light Speed 64, Waukesha, WI, USA). All patients took the supine position. The range of the scan was from the skull base to the sternal notch. The scan parameters were as follows: tube of voltage of 120 KV, tube current of auto Am, slice thickness 2.5 mm, interval 2.5 mm, and pitch 1.375. Ultravist (350 mg I/Ml, 1.5 ml/kg) was injected with a rate of 3.5 ml/s through the elbow vein by a high-pressure injection. Axial arterial-phase and venous-phase CT images were obtained at 25–30 and 60–70 s after injection and were exported in DICOM format.

Radiomic Features

VOI Segmentation and Radiomic Feature Acquisition

The venous-phase images were used for image feature extraction as the distribution of the contrast agent in the lesions was more homogeneous, and the image quality was better for distinguishing the lesions from the adjacent tissue (18–20). The technical process of the entire study is shown in **Figure 1**. The lesions were delineated on the venous-phase CECT images using the ITK-SNAP software (available at www.itksnap.org) in soft-tissue window (window width, 35; window level, 400). Two experienced radiologists (ZY, reader #1, radiology resident; ZL, reader #2; both doctors have 10 years of experience in imaging) blinded to the clinical outcomes were involved in ROI segmentation. The whole-tumor volume was determined by manually drawing a region of interest along the border of the tumor on each consecutive slice covering the whole lesion. Therefore, a three-dimensional (3D) volume of interest (VOI) was finally obtained. The radiomic features were automatically calculated by AK software (Artificial Intelligence Kit, GE Healthcare, Shanghai, China). The features extracted by the AK software comply with the standards set by the Image Biomarker Standardization Initiative. In total, 396 imaging features were extracted in each lesion, including (1) histogram features, such as mean, uniformity, skewness, kurtosis, energy, and entropy; (2) form factor features, such as volume CC, surface, surface

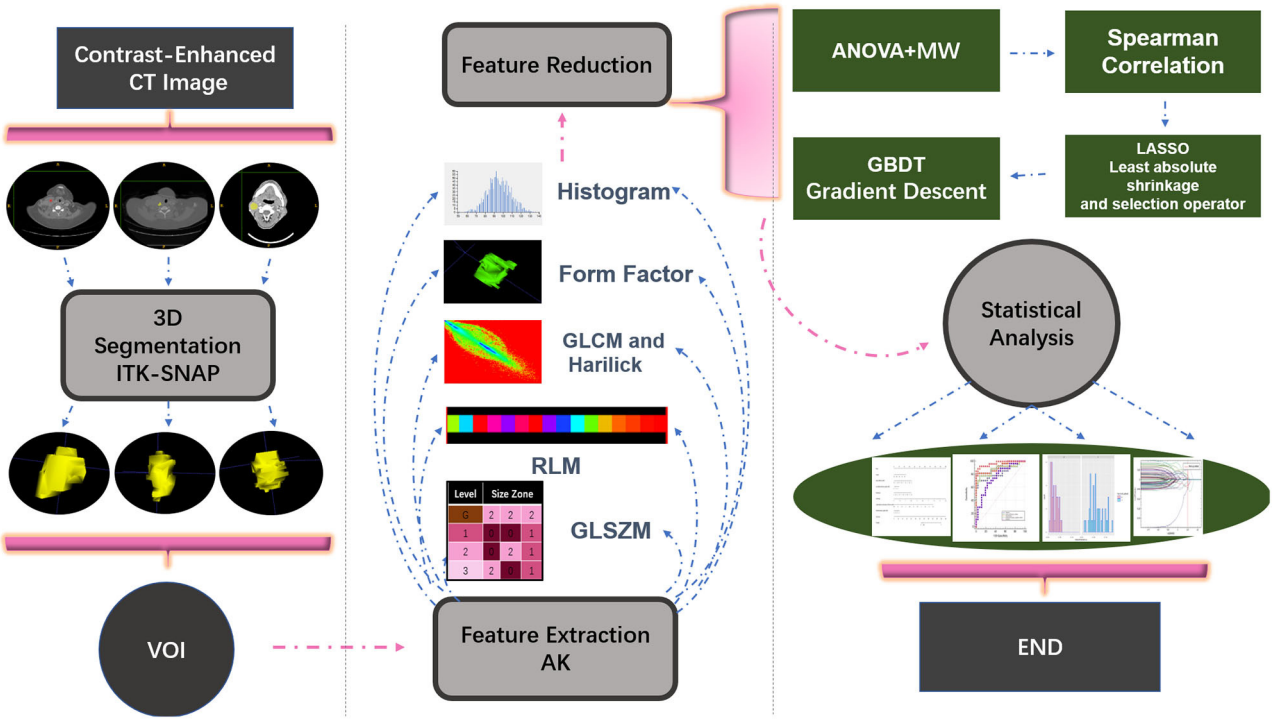


FIGURE 1 | Illustration of the process of data analysis. First, each layer of the lesion was manually segmented and automatically merged into a three-dimensional volume of interest (VOI) in the software. Then, upon extraction of the VOI radiomic features, relevant statistical methods were carried out for feature dimensionality reduction, and finally a statistical analysis on the selected features was performed and a model classifier was established.

volume ratio, compactness, and maximum 3D diameter; and (3) texture features including Gray level co-occurrence matrix (GLCM), Gray level run length matrix (GLRLM), Gray level size zone matrix, and Haralick parameters. The offset of GLCM and GLRLM were 1, 4, and 7. Features pre-processing was conducted in two steps: step 1—outliers and null values were replaced by mean values, and step 2—values standardization was carried out to eliminate the influence of the dimension. Feature dimension reduction was performed as follows: First, analysis of variance (ANOVA) and Mann–Whitney *U*-test were performed. Second, Spearman correlation test was conducted to remove the highly correlated variables. Third, in the LASSO model, the value of the minimum error rate among the 10-fold cross-validation was selected to construct the penalty function to compress the unimportant variable coefficients to zero (**Figure 2**). Gradient Descent algorithm for further feature screening was performed when the features were still redundant. In the study, the morphological features of the lesions were excluded. An analysis was made only about the texture features of the lesions (21).

Radiomic Modeling and Validation

All lymph nodes were randomly divided into training set (n = 54) and testing set (n = 23) with a ratio of 7:3. A total of 27 of the 38 lymph nodes of Kimura disease and 27 of 39 lymph nodes of metastases were included in the training set and 11 other lymph nodes of Kimura disease and 12 lymph nodes of metastases were in the testing set. Then, the binary logistic regression model was constructed based on the training set data to validate the model with the testing set data. The features and model identification performance were quantified by the area under the receiver operating characteristic (ROC) curve (AUC) in the training and the testing sets. Radiomic nomogram was then constructed on the basis of the binary logistic regression model. Radscore was calculated for each lesion and then converted into the risking probability of lymph node metastasis. A decision curve analysis was performed to evaluate the clinical benefit of the nomogram model developed in the testing dataset. The x axis of the curve is the threshold of the predicted probability outcome by the nomogram model. The y axis is the clinical decision net benefit for patients based on the discrimination result under this threshold.

Statistical Analysis

Statistical analysis was performed by R studio (1.1.463, packages: "verification," "pROC," "rms," "glmnet," "caret," and "rmda") and IBM SPSS Statistics 22. With regard to the reproducibility of volumetric and texture analysis, inter-observer reliability was assessed by intra-class correlation coefficient (ICC) test. Delong test was used for significant difference test among AUCs (22). Hosmer–Lemeshow test was used for evaluating model fit-goodness. The normal distribution test was performed using Shapiro–Wilk on continuous quantitative variables. Levene's test was used for equality of variances. $P > 0.05$ was considered to be normal distribution and variance is equal. Independent-sample

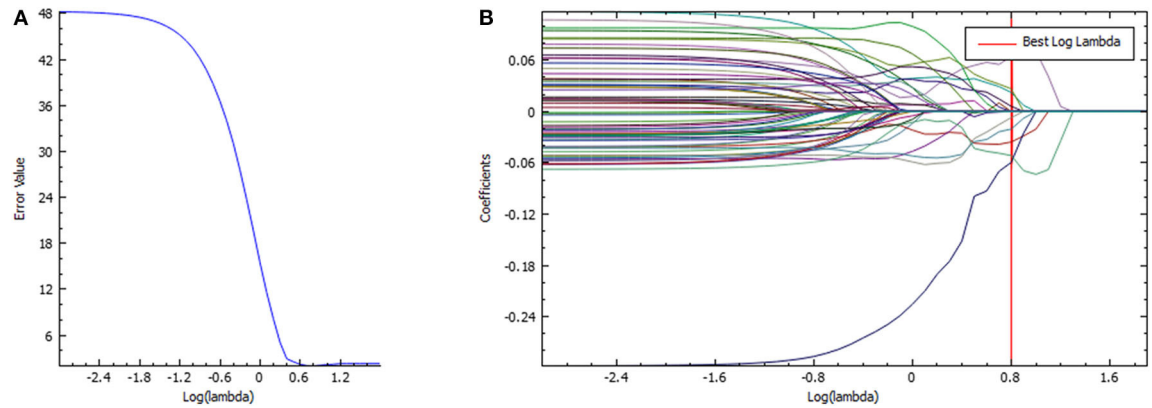


FIGURE 2 | Feature selection in the LASSO model. **(A)** LASSO error graph tuning parameter (λ) selection in the LASSO model used 10-fold cross-validation *via* minimum criteria. The error value was plotted vs. log(λ). Seven features with the smallest error value were selected. **(B)** LASSO coefficient profiles of the texture features. The red vertical line is drawn at the value selected using 10-fold cross-validation in log(λ) sequence and coefficients with non-zero values are indicated.

t-test was used for significant difference in variable distribution if normally distributed; otherwise, the non-parametric Mann–Whitney *U*-test was used. The qualitative variables were compared with chi-square or Fisher's exact test. $P < 0.05$ was considered as statistically significant.

RESULTS

The ICC values of the inter-observer of our research were 0.76–0.97, which suggest great accordance between two readers and the reliability of VOI sketching (22, 23). Three hundred ninety-six radiomic features were extracted automatically by AK software. The morphological features of the lesions were excluded and seven features were left after the redundancy reduction step, including one histogram feature, four GLCM features, and two GLRLM features. The seven features were significantly different between the two groups (all $P < 0.05$) (**Figure 3**, **Figure S1**). In the histogram feature of variance, the first quantile of the lymph nodes of the metastases group was significantly higher than the maximum value of the lymph nodes of the Kimura disease group. The variance value of metastases is generally greater than that of Kimura disease, which suggests that the image-brightness-changing gradient of metastases was steeper than that of Kimura disease. In the GLCM feature cluster, the first quantile of the Inertia_AllDirection_offset1 and HaraVariance of the metastases group is slightly larger than the fourth quantile of the Kimura group. The greater the value, the greater is the difference in the lesions. While the first quantile of the Kimura group of InverseDifferenceMoment_angle90_offset7 and sumAverage was greater than the fourth quantile of the metastases group, the larger the value of these two features, the smaller is the lesion difference. In the GLRLM cluster, the first quantile of the LongRunHighGreyLevelEmphasis_AllDirection_offset4 feature of the Kimura group was significantly higher than the fourth quantile of the metastases group, while in the ShortRunEmphasis_angle90_offset7 feature, the median of the metastases group was greater than the fourth quantile of the Kimura group. The first quantile of the metastases group is slightly lower than that of the Kimura group. The larger the value, the greater is the difference in gray value between adjacent pixels in the lesion (**Figure 3**). Two sets of mapped images of CECT and radiomic features of patients with Kimura disease and lymph node metastases are shown, respectively, in **Figure 4**, wherein the histogram is the gray scale distribution of the entire lesion. The gray distribution of the Kimura disease patient is more concentrated than that of the metastases patient. The variation of the run length matrix of metastases patients is greater than that of the Kimura patients, and the GLCM shows that the lesion complexity of the Kimura disease patient is less than that of the metastases patient.

Radiomic Model Building and Validation

A binary logistic regression model was established using the seven distinctive features. The radscore value of each lesion (Formula 1) was obtained, and the predicting risking probability of lymph nodes of metastases was obtained (Formula 2).

$$Radscore = 4.290 + 8.476A + 2.587B - 2.232C + 7.690D + 1.142E + 0.092F - 10.934G \quad (1)$$

$$Probability\ positive\ prediction\ probability = \frac{1}{1 + e^{(-radscore)}} \quad (2)$$

(A: Variance, B: Inertia_AllDirection_offset1, C: HaraVariance, D: LongRunHighGreyLevelEmphasis_AllDirection_offset4, E: ShortRunEmphasis_angle90_offset7, F: InverseDifferenceMoment_angle90_offset7, and G: sumAverage).

If the coefficient of the variable is negative, the smaller the value, the greater the risk probability will be. If the coefficient of the variable is positive, the smaller the value, the smaller the risk of developing lymph nodes of metastases will be.

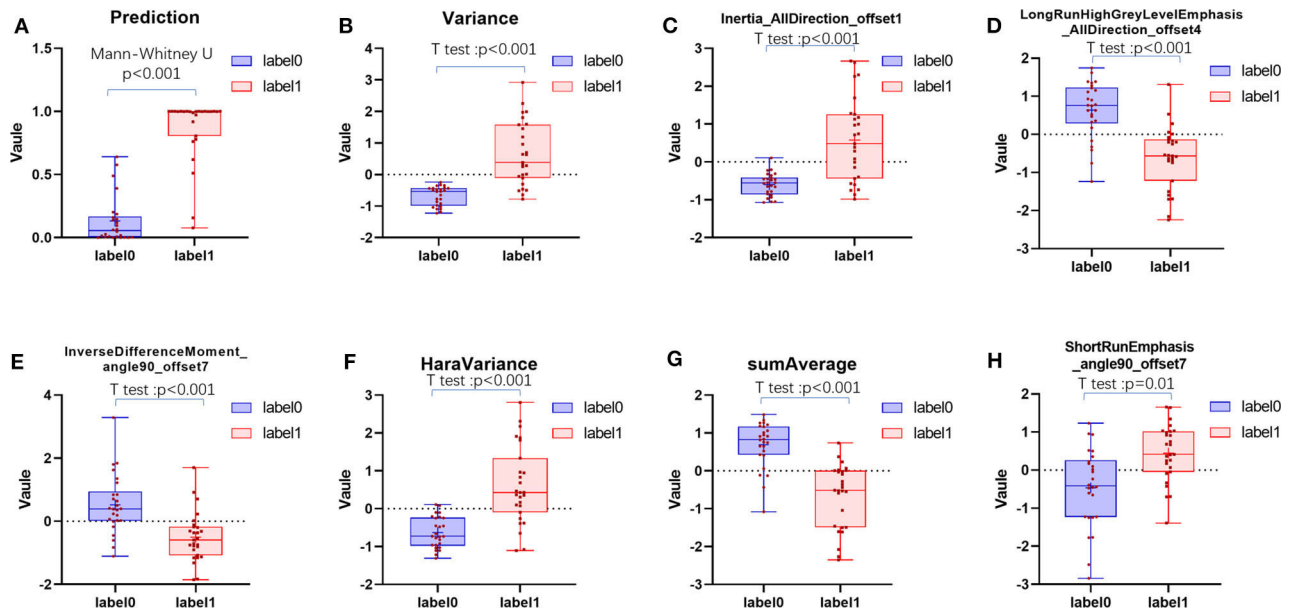


FIGURE 3 | The distribution of positive prediction probability and the selected feature values between the two groups **(A–H)**. Label 0 represents the Kimura disease group and the label 1 represents the lymph node metastases group. There was a significant difference in the distribution of values between the two groups, and all *P*-values were <0.05.

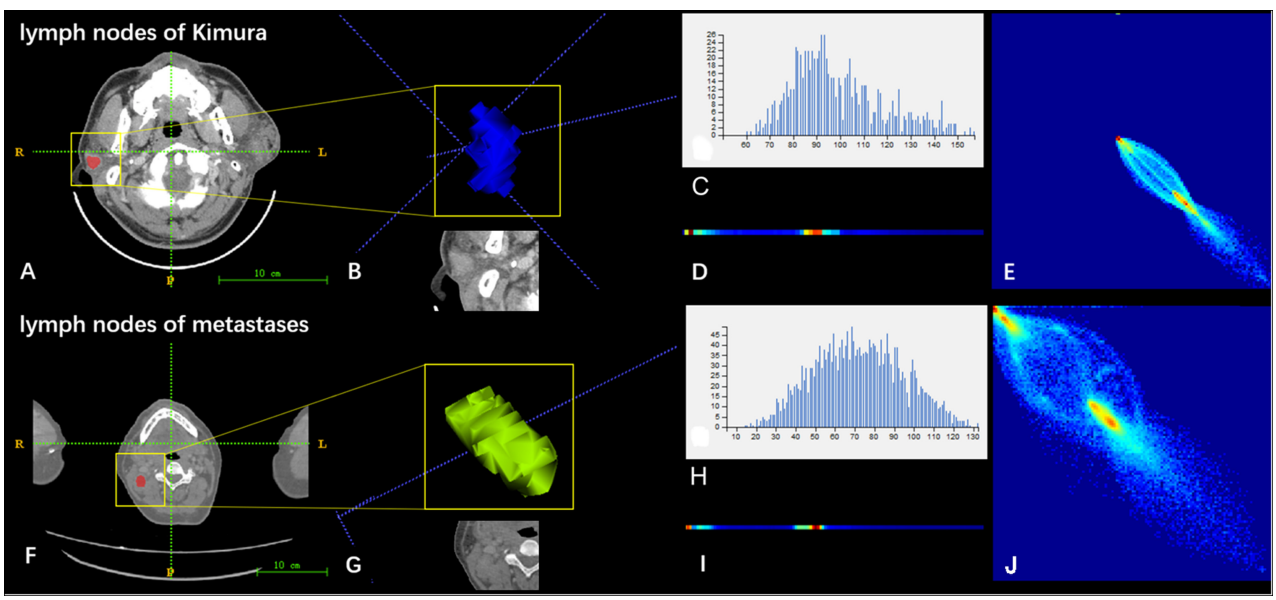


FIGURE 4 | Two sets of mapped images of radiomic features of patients with Kimura disease **(A–E)** and lymph node metastases **(F–J)**. **(A,F)** CECT images. **(B,G)** Volumes of interest. **(C,H)** Grayscale distribution histogram of the lesions. The distribution of **(H)** was more dispersed than the distribution of **(C)**. **(D,I)** Run length matrix features. The frequency of changing of the gray scale of **(D)** was smaller than that of **(I)**. **(E,J)** Gray level co-occurrence matrix features. The heterogeneity of lesions in **(J)** was greater than that of **(E)**.

The radiomic signature showed favorable predictive efficacy. The risking probability according to radscore shows a significant difference between the two groups ($P < 0.001$) (**Figure 3**). As can be seen from the figure, the prediction probability of the lymph nodes of the Kimura disease group is much lower than that of the lymph nodes of the metastases group [cutoff value: 0.490—this cutoff value is taking into account disease prevalence (50.9%); the value larger than 0.490 is considered to be the metastasis group, while the value smaller than 0.490 is thought to be the Kimura disease group]. The positive and negative predictive

TABLE 1 | The parameters of the model and the seven features of disease identification performance.

Features	AUC (95% CI)	Sensitivity	Specificity	Optimal criterion	Significance level *P* (area = 0.5)	OR	+LR (95% CI)	-LR (95% CI)	+PV (95% CI)	-PV (95% CI)	Cost
Variance	0.910 (0.830–0.991)	81.48	96.15	>-0.329	<0.0001	4800.048	21.19 (3.1–146.0)	0.19 (0.09–0.4)	95.7 (76.1–99.3)	83.3 (69.3–91.7)	0.113
Inertia_AllDirection_offset1	0.828 (0.711–0.945)	70.37	96.15	>-0.205	<0.0001	13.296	18.30 (2.6-127.0)	0.31 (0.2-0.6)	95.0 (73.2-99.2)	75.8 (63.5-84.9)	0.170
InverseDifference Moment_angle90_offset7	0.823 (0.705–0.942)	77.78	84.62	≤-0.179	<0.0001	1.096	5.06 (2.0-12.7)	0.26 (0.1–0.5)	84 (67.6–93.0)	78.6 (64.0–88.3)	0.189
HaraVariance	0.879 (0.779–0.978)	77.78	92.31	>-0.098	<0.0001	0.015	10.11 (2.6–38.9)	0.24 (0.1–0.5)	91.3 (73.2–97.6)	80 (66.2–89.1)	0.151
sumAverage	0.915 (0.837–0.99)	96.30	80.77	≤0.373	<0.0001	0	5.01 (2.3–11.0)	0.046 (0.007–0.3)	83.9 (70.2–92.0)	95.5 (75.3–99.3)	0.113
LongRunHighGreyLevel Emphasis_AllDirection_offset4	0.869 (0.766–0.972)	92.59	76.92	≤0.281	<0.0001	2,187.088	4.01 (2.0–8.2)	0.096 (0.02–0.4)	80.6 (67.2–89.4)	90.9 (72.2–97.5)	0.151
ShortRunEmphasis_angle90_offset7	0.759 (0.629–0.890)	66.67	76.92	>0.228	<0.0001	3.133	2.89 (1.4–6.1)	0.43 (0.2–0.8)	75 (58.6–86.4)	69 (55.6–79.8)	0.283
Prediction	0.970 (0.931–1)	92.59	92.31	>0.490	<0.0001	-	12.04 (3.2–45.8)	0.080 (0.02–0.3)	92.6 (76.7–97.9)	92.3 (75.9–97.9)	0.076

+LR, positive likelihood ratio; -LR, negative likelihood ratio; +PV, positive predictive value; -PV, negative predictive value; 95% CI, 95% confidence interval (binomial exact).

FIGURE 5 | Receiver operating characteristic plots of the selected seven features and the model's prediction probability based on them. According to the area under the curve index, the identification effectiveness of the binary logistic regression model based on the seven features is greater than the discrimination performance of any single feature.

values of prediction were higher than the performance values of the other seven features. According to the OR values of the seven features, the importance of the features can be ranked, where extreme values <1 or >1 indicate that the feature is more important, and the values of sumAverage and variance are more extreme, so the contribution of these two features to the model was greater, which is consistent with their AUC value trend (**Table 1**). The ROC curves of seven features and the predicted probabilities are established to evaluate the performance of each feature and model (**Figure 5**). From the figure, we can see that the model prediction probability (AUC: 0.970) is optimal for disease detection, followed by sumAverage (AUC: 0.915) and variance (AUC: 0.910). The AUC value of each variable was significantly different from the AUC of 0.5 (all $P < 0.0001$), indicating that each variable is reliable for the prediction and the identification of the disease. According to the AUC DeLong test between prediction and the other seven features, except for sumAverage and variance, the predictive power of prediction was significantly different from the other five features (all $P < 0.05$). All the information above can be seen in **Table 2**. In general, the eight variables in the figure have a good distinguishing effect on the disease. Sensitivity and specificity are obtained according to the most approximate Youden index. The optimal criterion value indicates the cutoff value which was assigned to the metastatic tumor group.

We further tested the fit-goodness of the established binary logistic regression model. The results show that the model is

TABLE 2 | Significant difference test between area under the curve values of the model and the seven features.

Comparison of variables	Difference between areas	Standard error	95% confidence interval	*z* statistic	Significance level
Prediction ~ HaraVariance	0.0912	0.0417	0.00945 to 0.173	2.187	$P = 0.0288$
Prediction ~ LongRunHighGreyLevelEmphasis_AllDirection_offset4	0.101	0.05	0.00306 to 0.199	2.021	$P = 0.0433$
Prediction ~ Inertia_AllDirection_offset1	0.142	0.0521	0.0403 to 0.245	2.733	$P = 0.0063$
Prediction ~ ShortRunEmphasis_angle90_offset7	0.211	0.0623	0.0887 to 0.333	3.384	$P = 0.0007$
Prediction ~ InverseDifferenceMoment_angle90_offset7	0.147	0.0563	0.0363 to 0.257	2.604	$P = 0.0092$
Prediction ~ sumAverage	0.0556	0.0361	−0.0152 to 0.126	1.54	$P = 0.1236$
Prediction ~ Variance	0.0598	0.0313	−0.00152 to 0.121	1.911	$P = 0.0560$

FIGURE 6 | Receiver operating characteristic curve **(A)** and classification effect diagram **(B)** for verifying the model with testing data. **(A)** The area under the curve is 0.977, the sensitivity is 1.0, and the specificity is 0.909 in testing data. **(B)** Pink and blue represent label 0 and label 1 according to the gold standard, respectively. The model uses 0 as the classification threshold, with blues <0 and pinks >0 being cases of model misclassification. It can be seen that the classification effect of the model in the testing group is generally good.

in good agreement with the actual model ($\chi^2 = 2.127$, $P = 0.977$, Hosmer–Lemeshow test). The model was validated in the testing group and found to have good generalization ability. The AUC of the testing set was 0.977, the sensitivity was 1, and the specificity was 0.909 (**Figure 6A**). The model performed to be a good classifier on the testing set data (**Figure 6B**). The decision curve of the model shows that the net benefit of making decisions based on the established model is much greater than treating all patients or not treating all patients between the probability threshold of 0.08–1.0 (**Figure 7**).

We have constructed a nomogram of the predictive model for model application. After we get the patient's image feature data, normalize the feature, then get the corresponding points according to the values of these seven features, and finally add these seven points to get the total point, the total point is vertically corresponding to the probability scale line. The probability of having a metastatic tumor in this patient is available (**Figure 8**).

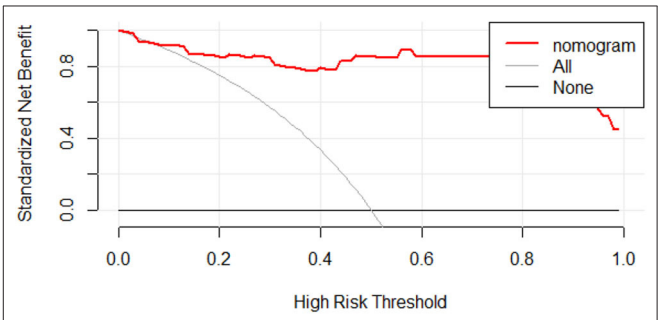


FIGURE 7 | Decision curve analysis for radiomic discrimination model. The Y-axis represents the standardized net benefit. All: assuming that all patients will be treated. None: assuming that no patient will be treated. Red line: the nomogram prediction performance based on model. When making a decision based on a nomogram, the standard net benefit obtained is greater than the treatment of all patients or none in the range of threshold probability 0.08–1.0.

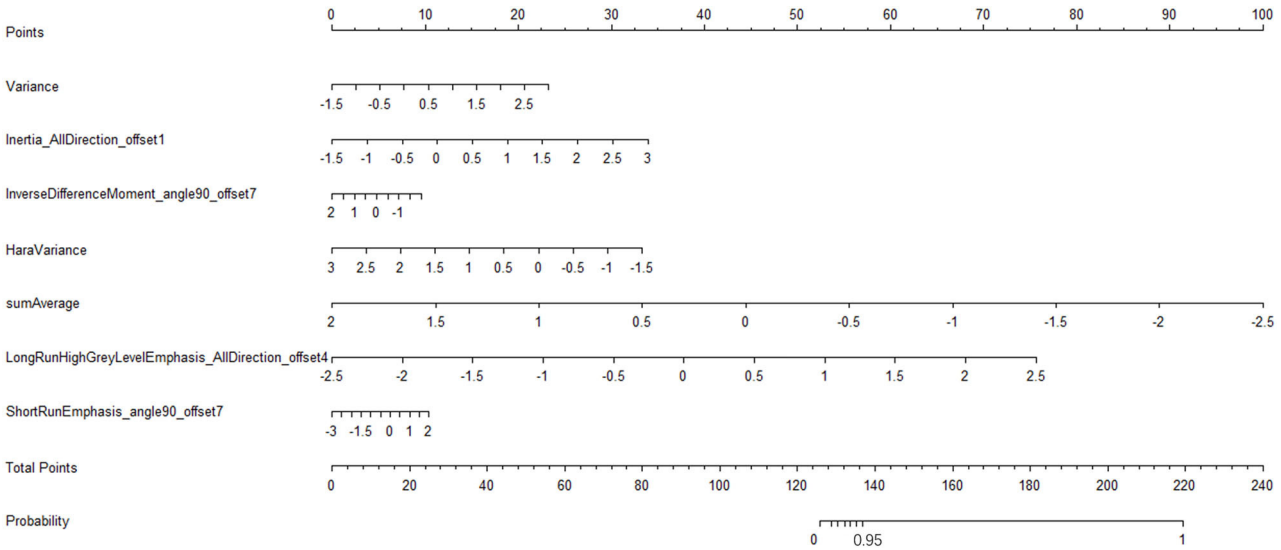


FIGURE 8 | Radiomic nomogram to detect lymph node lesion. The radiomic nomogram was developed in the training set with the seven features. In the nomogram, first, make a vertical line according to the value of the selected seven features to determine the corresponding value of points. Then, the total points are the sum of the seven points above. Finally, make a vertical line according to the value of the total points to determine the probability of lymph node metastases.

DISCUSSION

The present study is the first to distinguish Kimura disease from lymph node metastases in the head and neck using radiomic features and the model based on the features on venous-phase CECT images. In our study, seven features and predicted probabilities have a good distinguishing effect on the disease. The predicted probabilities are optimal for disease detection.

The morbidity of Kimura disease is low, and the majority of reports about Kimura disease focus on clinical findings, with a few of the imaging findings reported (24–26). There are three types of CT manifestations of Kimura disease in the head and neck: (1) multiple nodular type, showing multiple nodules, with clear borders and uniform enhancement, (2) diffuse mass type, characterized by ill-defined diffuse subcutaneous soft tissue mass, with thickening of the adjacent skin and mild to moderate heterogeneous enhancement, mainly located in the subcutaneous fat space of the maxillofacial region, and (3) mixed type, with characteristics of both of the above types.

Most patients having subcutaneous tumor-like nodules with large parotid gland and local lymph node involvement are easy to be misdiagnosed as malignant tumors, which make a differential diagnosis difficult even using CT and MRI. Lymph node metastasis is a high-risk factor for the prognosis (27); more than one metastatic lymph node will increase the risk of recurrence (28). Metastasis of tumors to sentinel lymph nodes can predict disease progression and often guides a treatment scheme (29). In clinical practice, many patients are required for further CT or MRI to detect whether primary tumors exist. There are different treatment modalities for Kimura disease, and the postoperative recurrence rate is high (2, 26). The main treatment of Kimura disease is different from malignant tumors. There is no need to do radical surgery. Therefore, it is necessary to make a clear diagnosis before surgery. Although some scholars have summarized some imaging features of Kimura disease in the head and neck, it is necessary to combine clinical and laboratory examinations to improve the diagnostic accuracy for lacking image characteristics and the pretty low diagnostic accuracy.

In the recent years, radiomics increasingly draws attention and has demonstrated that it may be a tool that can obtain high-fidelity target information to comprehensively evaluate lesions, especially the texture features in the image that are not recognized by the naked eye and reveal the inherent heterogeneity of the tissue, reflecting the subtle differences between different tissues. Radiomics can be combined with the imaging appearance to further improve the differential diagnosis ability of the lesion (30, 31). The AK, an imaging analytic software used in this study, has been used in many research reports (32, 33). A previous study showed that radiomic feature-based CT imaging signatures allow the prediction of lymph node metastasis in cancer and could facilitate the preoperative individualized prediction of lymph node status (20).

Kimura disease is a rare disease; the lymph nodes involved in the case are often multiple. Therefore, the diseased lymph nodes were selected as the research object, and the AK software was used for feature extraction and dimensionality reduction. A total of 396 features were extracted and seven texture features were selected to identify Kimura disease from lymph node metastases, and a logistic regression model was established. In order to avoid model over-fitting, we adopted 10-fold cross-validation using training set data and testing set data for the established model. The mean AUCs of models in the two sets were 0.7812 and 0.7628. The AUC of the testing set was 0.977, which is a strong validation of the good performance index of the logistic regression model established in the study. In this

study, the seven screened out features showed a significantly different distribution between the two groups, and from the point of view of the features themselves, the results showed that the heterogeneity of metastatic tumors was greater than that of Kimura disease on CECT. This conclusion may be explained by a previous basic study. Lee et al. (29), by using comparative transcriptomics and metabolomics analyses of primary and lymph node metastasis tumors in mice, found that lymph node metastasis requires that tumor cells undergo a metabolic shift toward fatty acid oxidation (FAO). Transcriptional coactivator yes-associated protein (YAP) is selectively activated in lymph node metastatic tumors, leading to the up-regulation of genes in the FAO signaling pathway. Several bioactive bile acids accumulated to high levels in metastatic lymph node metastasis, and these bile acids activated YAP in tumor cells, likely through the nuclear vitamin D receptor. The study showed that lymph node metastases are complex.

It is also obvious that both the CECT image and the radiomic features image of the lymph nodes in a metastatic tumor patient have a greater changing rate and more complexity than that of the Kimura disease (**Figure 8**). The discriminated efficiency of the model is better than any single feature for the two diseases, and the disease identification ability of the model, in addition to variance and sumAverage, is significantly different from the other five features. The results show that the model has a higher identification accuracy; the decision curve of the model shows a greater standard net benefit within a wide threshold probability (0.08–1.0) than treating all patients or treating no patient. So, we prefer using this model as a basis for decision making to identify these two kinds of diseases. The nomogram is one of the important applications of the model. Through the nomogram, the risk of each patient can be predicted (20, 34, 35). Using the model to classify the data of the testing set, it is found that the correctness of the classification is good, which may be due to the small amount of sample data in the testing set. This study demonstrates that radiomics can help identify Kimura disease in the head and neck and lymph node metastases, and the established nomogram can predict the risk of lymph node metastases in patients. Radiomics can be used as an intelligent-aided tech to diagnose diseases.

There are some limitations in our study. First, this is a single-institution retrospective analysis. The sample size is rather small because of the low morbidity of Kimura disease. Second, because of lack of data, we did not integrate clinical features and genetic and immunohistochemical data into a statistically predictive model. Third, this study lacks an external validation. Therefore, the sample should be expanded and multi-center independent samples are needed to further improve the accuracy of the model. In the future, some clinical data will be integrated into a statistically predictive model.

In summary, our results showed that CECT images contain much useful information which could be used to differentiate Kimura disease from lymph node metastases, but which could not be seen through naked eyes. Radiomic technology can deeply explore the image heterogeneity information, which may be an effective and non-invasive way for differential diagnosis between Kimura disease and lymph node metastases.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

ETHICS STATEMENT

This retrospective study was approved by the ethics committee of Cangzhou Central Hospital, and the informed consent requirement was waived.

AUTHOR CONTRIBUTIONS

LK is the guarantor of the article. LK, YZ, SY, and LZ contributed to the conception and design. YZ and LZ collected and assembled the data. YZ contributed to data analysis and interpretation. All the authors contributed to manuscript writing and gave final approval of the manuscript.

ACKNOWLEDGMENTS

We would like to thank Dr. Jialiang Ren from GE Healthcare for his statistical analysis of the study and article revision.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.01121/full#supplementary-material

Figure S1 | Histogram distribution of the screened out seven radiomic features.

Table S1 | Pathology of 39 Lymph Node Metastases.

REFERENCES

1. Kimura T, Yoshimura S, Ishikaura E. On the unusual granulation combined with hyperplastic changes of lymphatic tissue. *Trans Soc Pathol Jpn.* (1948) 37:179–80.
2. Jiang Y, Hua Q, Ren J, Zeng F, Sheng J, Liao H, et al. Eosinophilic hyperplastic lymphogranuloma: clinical diagnosis and treatment experience of 41 cases. *Am J Otolaryngol.* (2017) 38:626–9. doi: 10.1016/j.amjoto.2017.07.007
3. Matsuo T, Tanaka T, Kinomura M. Nephrotic syndrome during the tapering of oral steroids after pathological diagnosis of Kimura disease from a lacrimal gland mass: case report and review80 of 10 Japanese patients. *J Clin Exp Hematop.* (2017) 57:147–52. doi: 10.3960/jslrt.17028
4. Zhang L, Yao L, Zhou WW, Ma JN, Zhang CQ. Computerized tomography features and clinicopathological analysis of Kimura disease in head and neck. *Exp Ther Med.* (2018) 16:2087–93. doi: 10.3892/etm.2018.6431
5. Malhotra M, Varshney S, Singh N. Kimura disease. a rare case report from himalayan region. *Indian J Otolaryngol Head Neck Surg.* (2017) 69:571–4. doi: 10.1007/s12070-016-1045-6
6. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuze S, et al. Promises and challenges for the implementation of computational

medical imaging (radiomics) in oncology. *Ann Oncol.* (2017) 28:1191–206. doi: 10.1093/annonc/mdx034
7. Lambin P, Leijenaar RTH, Deis TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicin. *Nat Rev Clin Oncol.* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
8. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
9. Aerts HJWL, Rios Valazquez E, Leijanaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2014) 5:4006. doi: 10.1038/ncomms5006
10. Kickingereder P, Gotz M, Muschelli J, Wick A, Neuberger U, Shinohara RT, et al. Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. *Clin Cancer Res*. (2016) 22:5765–71. doi: 10.1158/1078-0432.CCR-16-0702
11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
12. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology.* (2016) 281:947–57. doi: 10.1148/radiol.2016152234
13. Xu X, Zhang HL, Liu QP, Sun SW, Zhang J, Zhu FP, et al. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. *J Hepatol.* (2019) 70:1133–44. doi: 10.1016/j.jhep.2019.02.023
14. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol.* (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
15. Chaddad A, Kucharczyk MJ, Daniel P, Sabri S, Jean-Claude, BJ, Niazi T. et al. Radiomics in glioblastoma: current status and challenges facing clinical implementation. *Front Oncol.* (2019) 9:374. doi: 10.3389/fonc.2019.00374
16. Park SW, Kim HJ, Sung KJ, Lee JH, Park IS. Kimura disease: CT and MR imaging findings. *AJNR Am J Neuroradiol.* (2012) 33:784–8. doi: 10.3174/ajnr.A2854
17. Gregoire V, Ang K, Budach W, Grau C, Hamoir M, Langendijk JA. et al. Delineation of the neck node levels for head and neck tumors: a 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines. *Radiother Oncol.* (2014) 110:172–81. doi: 10.1016/j.radonc.2013.10.010
18. Feng C, Lu F, Shen Y, Li A, Yu H, Tang H, et al. Tumor heterogeneity in gastrointestinal stromal tumors of the small bowel: volumetric CT texture analysis as a potential biomarker for risk stratification. *Cancer Imaging.* (2018) 18:46. doi: 10.1186/s40644-018-0182-4
19. Lu W, Zhong L, Dong D, Fang M, Dai Q, Leng S, et al. Radiomic analysis for preoperative prediction of cervical lymph node metastasis in patients with papillary thyroid carcinoma. *Eur J Radiol.* (2019) 118:231–8. doi: 10.1016/j.ejrad.2019.07.018
20. Jiang Y, Wang W, Chen C, Zhang X, Zha X, Lv W, et al. Radiomics signature on computed tomography imaging: association with lymph node metastasis in patients with gastric cancer. *Front Oncol.* (2019) 9:340. doi: 10.3389/fonc.2019.00340
21. Seidler M, Forghani B, Reinhold C, Pérez-Lara A, Romero-Sanchez G, Muthukrishnan N, et al. Dual-energy CT texture analysis with machine learning for the evaluation and characterization of cervical lymphadenopathy. *Comput Struct Biotechnol J.* (2019) 17:1009–15. doi: 10.1016/j.csbj.2019.07.004
22. Yuan M, Zhong Y, Zhang YD, Yu TF, Li H, Wu JF. Volumetric analysis of intravoxel incoherent motion imaging for assessment of solitary pulmonary lesions. *Acta Radiol.* (2017) 58:1448–56. doi: 10.1177/0284185117698863
23. Kang Y, Lee JW, Koh YH, Hur S, Kim SJ, Chai JW, et al. New MRI grading system for the cervical canal stenosis. *AJR Am J Roentgenol.* (2011) 197:134–40. doi: 10.2214/AJR.10.5560
24. Gopinathan A, Tan TY. Kimura's disease: imaging patterns on computed tomography. *Clin Radiol.* (2009) 64:994–9. doi: 10.1016/j.crad.2009.07.003
25. Zhang R, Ban XH, Mo YX, Lv MM, Duan XH, Shen J, et al. Kimura's disease: the CT and MRI characteristics in fifteen cases. *Eur J Radiol.* (2011) 80:489–97. doi: 10.1016/j.ejrad.2010.09.016
26. Li X, Wang J, Li H, Zhang M. Misdiagnosed recurrent multiple Kimura's disease: a case report and review of the literature. *Mol Clin Oncol.* (2019) 10:352–6. doi: 10.3892/mco.2018.1793
27. Carlson RW, Scavone JL, Koh WJ, McClure JS, Greer BE, Kumar R. et al. NCCN framework for resource stratification: a framework for providing and improving global quality oncology care. *J Natl Compr Canc Netw.* (2016) 14:961–9. doi: 10.6004/jnccn.2016.0103
28. Tsai CS, Lai CH, Wang CC, Chang JT, Chang TC, Tseng CJ, et al. The prognostic factors for patients with early cervical cancer treated by radical hysterectomy and postoperative radiotherapy. *Gynecol Oncol.* (1999) 75:328–33. doi: 10.1006/gyno.1999.5527
29. Lee CK, Jeong SH, Jang C, Bae H, Kim YH, Park I, et al. Tumor metastasis to lymph nodes requires YAP-dependent metabolic adaptation. *Science.* (2009) 363:644–9. doi: 10.1126/science.aav0173
30. Chaddad A, Tanougast C. Extracted magnetic resonance texture features discriminate between phenotypes and are associated with overall survival in glioblastoma multiforme patients. *Med Biol Eng Comput.* (2016) 54:1707–18. doi: 10.1007/s11517-016-1461-5
31. Cui WJ, Wang C, Jia L, Ren S, Duan SF, Cui C, et al. Differentiation between G1 and G2/G3 phyllodes tumors of breast using mammography and mammographic texture analysis. *Front Oncol.* (2019) 9:433. doi: 10.3389/fonc.2019.00433
32. Hu B, Xu K, Zhang Z, Chai R, Li S, Zhang L. A radiomic nomogram based on an apparent diffusion coefficient map for differential diagnosis of suspicious breast findings. *Chin J Cancer Res*. (2018) 30:432–8. doi: 10.21147/j.issn.1000-9604.2018.04.06
33. Shu Z, Fang S, Ding Z, Mao D, Cai R, Chen Y, et al. MRI-based radiomics nomogram to detect primary rectal cancer with synchronous liver metastases. *Sci Rep.* (2019) 9:3374. doi: 10.1038/s41598-019-39651-y
34. Zhao C, Lou Y, Wang Y, Wang D, Tang L, Gao X, et al. A gene expression signature-based nomogram model in prediction of breast cancer bone metastases. *Cancer Med.* (2019) 8:200–8. doi: 10.1002/cam4.1932
35. Liang W, Yang P, Huang R, Xu L, Wang J, Liu W, et al. A combined nomogram model to preoperatively predict histologic grade in pancreatic neuroendocrine tumors. *Clin Cancer Res.* (2019) 25:584–94. doi: 10.1158/1078-0432.CCR-18-1305

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 20 August 2020
doi: 10.3389/fonc.2020.01416

Radiomics Nomogram for Prediction of Peritoneal Metastasis in Patients With Gastric Cancer

Weicai Huang[1†], Kangneng Zhou[2†], Yuming Jiang[1†], Chuanli Chen[3], Qingyu Yuan[3], Zhen Han[1], Jingjing Xie[4], Shitong Yu[1], Zepang Sun[1], Yanfeng Hu[1], Jiang Yu[1], Hao Liu[1], Ruoxiu Xiao[2], Yikai Xu[3], Zhiwei Zhou[5,6]* and Guoxin Li[1]**

[1] Department of General Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, China, [2] School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China, [3] Department of Medical Imaging Center, Nanfang Hospital, Southern Medical University, Guangzhou, China, [4] Center for Drug and Clinical Research, Nanfang Hospital, Southern Medical University, Guangzhou, China, [5] Department of Gastric Surgery, Sun Yat-sen University Cancer Center, Guangzhou, China, [6] State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

OPEN ACCESS

Edited by:
Xuelei Ma,
Sichuan University, China

Reviewed by:
Ziyu Li,
Peking University Cancer Hospital, China
Di Dong,
Chinese Academy of Sciences, China

***Correspondence:**
Yikai Xu
Yikaivip@163.com
Zhiwei Zhou
zhouzhw@sysucc.org.cn
Guoxin Li
gzliguoxin@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 22 October 2019
Accepted: 06 July 2020
Published: 20 August 2020

Citation:
Huang W, Zhou K, Jiang Y, Chen C, Yuan Q, Han Z, Xie J, Yu S, Sun Z, Hu Y, Yu J, Liu H, Xiao R, Xu Y, Zhou Z and Li G (2020) Radiomics Nomogram for Prediction of Peritoneal Metastasis in Patients With Gastric Cancer. Front. Oncol. 10:1416. doi: 10.3389/fonc.2020.01416

Objective: The aim of this study is to evaluate whether radiomics imaging signatures based on computed tomography (CT) could predict peritoneal metastasis (PM) in gastric cancer (GC) and to develop a nomogram for preoperative prediction of PM status.

Methods: We collected CT images of pathological T4 gastric cancer in 955 consecutive patients of two cancer centers to analyze the radiomics features retrospectively and then developed and validated the prediction model built from 292 quantitative image features in the training cohort and two validation cohorts. Lasso regression model was applied for selecting feature and constructing radiomics signature. Predicting model was developed by multivariable logistic regression analysis. Radiomics nomogram was developed by the incorporation of radiomics signature and clinical *T* and *N* stage. Calibration, discrimination, and clinical usefulness were used to evaluate the performance of the nomogram.

Results: In training and validation cohorts, PM status was associated with the radiomics signature significantly. It was found that the radiomics signature was an independent predictor for peritoneal metastasis in multivariable logistic analysis. For training and internal and external validation cohorts, the area under the receiver operating characteristic curves (AUCs) of radiomics signature for predicting PM were 0.751 (95%CI, 0.703–0.799), 0.802 (95%CI, 0.691–0.912), and 0.745 (95%CI, 0.683–0.806), respectively. Furthermore, for training and internal and external validation cohorts, the AUCs of radiomics nomogram for predicting PM were 0.792 (95%CI, 0.748–0.836), 0.870 (95%CI, 0.795–0.946), and 0.815 (95%CI, 0.763–0.867), respectively.

Conclusions: CT-based radiomics signature could predict peritoneal metastasis, and the radiomics nomogram can make a meaningful contribution for predicting PM status in GC patient preoperatively.

Keywords: gastric cancer, peritoneum, metastasis, radiomics, nomogram

INTRODUCTION

Gastric cancer (GC) is one of the most common human malignancies and the third leading cause of cancer-related deaths worldwide (1–3). Surgical resection is the major treatment for GC patient (4); however, patients in advanced gastric cancer with peritoneal metastasis, a non-curable factor, showed poor prognosis (5). Peritoneal metastasis (PM) primary occurs in T4 stage (6, 7). Accurate evaluation of PM status in GC patients is essential for treatment decision and prognosis. It was found that some biomarkers and histopathological factors (e.g., *T*, *N* staging, Greater Omental Milky Spot, and Troponin I2) could predict PM status in GC (7, 8), but they just provided low prediction or were merely available postoperatively. Preoperative assessment of PM can provide useful information for performing adjuvant treatment and avoid unnecessary surgical resection, thus contributing to pretreatment decision. Computed tomography (CT), which could detect obvious parietal peritoneum thickening and ascites, is regarded as a popular non-invasive method to diagnose PM (9). However, the signs of obvious parietal peritoneum thickening or large amount of ascites did not always exist in every GC patient with PM. There still exist GC patients who were CT-diagnosed PM-negative but confirmed PM-positive during subsequent laparoscopies (9). Therefore, CT-diagnosed PM has low sensitivity. Laparoscopy, a golden criterion for detecting PM status, is strongly recommended by the European Society for Medical Oncology (ESMO) and National Comprehensive Cancer Network (NCCN) guidelines to perform for its diagnosis but is controversial due to the different clinical *T* stage and health status; for example, some patients could not suffer the intraperitoneal high pressure during the performance of laparoscopy. Laparoscopy is an invasive diagnostic procedure; hence, it is not suitable for each patient. Therefore, accurate preoperative prediction of PM status is very important for GC patients, especially at the late stage.

Radiomics, an arising field that involves converting digital medical images into mineable data, analyzing data, and improving medical decision, has attracted increasing attention in recent years (10, 11). With radiomics, the accuracy of diagnosis, prognosis, and prediction could be improved, especially in oncology (12). Radiomics enables the non-invasive profiling of tumor heterogeneity (13, 14) through integrating complex imaging features. The applications of radiomics mainly focus on individualized therapy associated with cancer such as tumor detection, lymph node metastasis (LNM), subtype classification, survival, and treatment reaction assessment (13, 15–17). It was reported that CT texture was associated with prognosis in patients with GC (18); however, an ideal method that can change complex imaging features into a signature for predicting PM is still urgent to be developed. For GC patients especially those in clinical T3–T4, who are more likely to have higher risk of PM, developing a predictive model with radiomics signature to predict PM status is quite necessary.

The purpose of this study was to establish a radiomics signature for predicting the PM status in GC patients on the basis of preoperative CT information and to further develop a radiomics model that incorporates the clinicopathological findings and radiomics signature for the personal prediction of PM status in patients with GC preoperatively.

MATERIALS AND METHODS

Patients

This study was approved by the ethics committee of every participating center, and the informed consent requirement was waved. We retrospectively selected three independent cohorts of patients with GC in pathological T4 stage. The training and internal validation cohorts comprising 562 and 106 consecutive patients, respectively, with total or partial radical gastrectomy were obtained from Sun Yat-sen University Cancer Center between January 2008 and December 2012 and January 2013 and December 2013, respectively. The external validation cohort that comprised 287 consecutive patients was obtained from Nanfang Hospital of Southern Medical University (Guangzhou, China) between January 2007 and December 2013. Clinicopathological data of each patient were collected retrospectively. **Table 1** shows the characteristics of the 955 GC patients. Patients were included if they met the following criteria: performed standard unenhanced and contrast-enhanced abdominal CT <4 weeks before surgery, with fundamental clinicopathological data, without combined malignant tumor, without preoperative chemotherapy, and being confirmed T4 GC histologically. Exclusion criteria were as follows: CT could not distinguish the lesions of the neoplasm, and anticancer therapy was performed previously. Fundamental clinicopathological data, such as gender, age, size, location, cancer antigen 19-9 (CA19-9), carcinoembryonic antigen (CEA), status of preoperative differentiation, were obtained from medical records. We also collected the dates of baseline CT imaging and the clinical *T* stage (cT) and *N* stage (cN) of patients. The details are shown in **Figure S1**.

Peritoneal metastasis status was divided into two outcome categories: PM-negative status [PM(−)] and PM-positive status [PM(+)]. All the diagnoses of PM status were based on the laparoscopy surgery and pathological examination.

Image Acquisition

Contrast-enhanced abdominal CT of patients was performed by the multidetector row CT (MDCT) systems (256-MDCT scanner Brilliance iCT, Philips Healthcare, Cleveland, OH, USA; 64-section LightSpeed VCT, GE Medical Systems, Milwaukee, WI, USA; or GE Lightspeed 16, GE Healthcare Milwaukee, WI, USA). To standardize the image acquisition, portal venous phase contrast-enhanced CT images were retrieved from the picture archiving and communication system (PACS) (Carestream, Canada) in Digital Imaging and Communications in Medicine (DICOM) format for feature extraction (19). The details are shown in **Supplementary Materials**.

Imaging Texture Analysis

All the CT images were reviewed by two experienced radiologists who both had clinical experience in abdominal CT study interpretation for more than 10 years. Based on the consensus of these two radiologists, the tumor manual segmentation

TABLE 1 | Characteristics of patients with gastric cancer (GC) in each cohort.

Variables	Training Cohort (*N* = 562)		Internal Validation Cohort (*N* = 106)		External Validation Cohort (*N* = 287)	
	N	%	*N*	%	*N*	%
AGE (YEARS)						
≥60	238	42.3	46	43.4	125	43.6
<60	324	57.7	60	56.6	162	56.4
GENDER						
Male	385	68.5	76	71.7	207	72.1
Female	177	31.5	30	28.3	80	27.9
SIZE						
≥4 cm	409	72.8	79	74.5	196	68.3
<4 cm	153	27.2	27	25.5	91	31.7
DIFFERENTIATION						
Well or moderate	75	13.3	21	19.8	60	20.9
Poor or undifferentiation	487	86.7	85	80.2	227	79.1
LAUREN TYPE						
Intestinal	157	27.9	36	34	100	34.8
Mixed and diffuse	405	72.1	70	66	187	65.2
LOCATION						
Cardia	218	38.8	32	30.2	59	20.6
Body	113	20.1	31	29.2	48	16.7
Antrum	183	32.6	36	34	140	48.8
Whole	48	8.5	7	6.6	40	13.9
CEA						
Elevated	139	24.7	30	28.3	50	17.4
Normal	423	75.3	76	71.7	237	82.6
CA19-9						
Elevated	130	23.1	19	17.9	62	21.6
Normal	432	76.9	87	82.1	225	78.4
cT STAGE						
T3	99	17.6	27	25.5	34	11.8
T4a	337	60	48	45.3	232	80.8
T4b	126	22.4	31	29.2	21	7.4
cN STAGE						
N0	159	28.3	23	21.7	69	24
N1	131	23.3	18	17	58	20.2
N2	125	22.2	22	20.8	69	24
N3	147	26.2	43	40.6	91	31.8
PM						
PM(–)	472	84	89	84	225	78.4
PM(+)	90	16	17	16	62	21.6

was performed and checked with *ITK-SNAP* software (www.itksnap.org) (20). They constructed manually a single region of interest (ROI) that covered the entire area of the lesion on the transverse image section, which depicted the maximum lesion diameter for each lesion. Based on the above procedure, we used an available radiomics analysis package (https://github.com/mvallieres/radiomics/) in Matlab R2016a (The MathWorks Inc.) to extract and calculated the radiomics features of these images. The inter- and intraobserver variability of radiomics feature extraction of the two radiologists was initially analyzed with 100 randomly chosen images for ROI-based texture feature extraction; details are shown in **Supplementary Methods 2**. The final feature pool included first-order intensity features, shape features, and second- and higher-order textural features. The detailed mathematical definitions of all imaging features are listed in **Supplementary Materials**.

Radiomics Feature Selection and Signature Development

In order to predict the PM status of patients with GC, we used the least absolute shrinkage and selection operator (LASSO) logistic regression model to select the optimal radiomics features

from the primary texture features, and then, the development of the radiomics score (Rad-score) was constructed in the training cohort (21). For further detecting and addressing the collinearity among features, scatterplot correlation matrix with Person correlation coefficient was applied to investigate the interrelationship among the primary selected features and PM status, and if features had a correlation coefficient that was higher than 0.80 between each other, then the one with the highest collinearity was excluded from the analysis (22–24). In this study, we used the R software (version 3.5.3) with the "glmnet" package to perform the LASSO regression (25, 26). A detailed information is provided in **Supplementary Materials**.

Development of an Individualized Prediction Model

Estimation of univariate relationships between PM status and potential predictors were developed with logistic regression analysis. The basis of the PM status prediction model was established by multivariate logistic regression analysis. Different variables were analyzed by using univariate logistic regression analysis in the training cohort, and statistically significant variable values with $P < 0.05$ in each cohort were then entered into the multivariate analyses. We used the likelihood ratio test with Akaike's information criterion as the stopping rule to apply the backward step-wise selection (27). Based on the prediction model, a nomogram was then constructed.

Validation of the Prediction Model

The accuracy of the prediction model was assessed by measuring both discrimination and calibration. We used 1,000 bootstrapping resamples for evaluating both discrimination and calibration. Discriminative ability was measured by the area under the receiver operating characteristic curve (AUC). The consistency between the predicted and actual probability of PM status was graphically represented by calibration plots.

Clinical Use

The potential net benefit of the predictive models was assessed by the decision curve analysis (DCA), which is popular as a new method for evaluating predictive model recently (28). The decision strategy based on every threshold probability would show a potential net benefit by using this method. In this study, we quantified the net benefits at different threshold probabilities with the use of the DCA to identify the clinical usefulness of the Rad-score, cT stage, cN stage, and radiomics nomogram.

Statistical Analysis

For continuous variable values, the two-tailed *t*-test, unpaired, one-way ANOVA, or Mann–Whitney test were used for comparison. For categorical variables, the χ^2 test or Fisher's exact test was used to compare. Standardization of the image features was applied by transforming the data of each feature into new scores with a mean of zero and a standard deviation of 1 (z-score transformation) (29, 30). Most of the statistical tests were examined by using SPSS version 21.0 (IBM) and R version 3.5.3 (http://www.r-project.org), and the nomograms and calibration plots were conducted by using the R version 3.5.3 with rms package. For all tests, $P < 0.05$ was thought to be statistically significant.

RESULTS

Clinical Characteristics

Table 1, **Tables S1–S3** are the clinicopathological characteristics of the training cohort ($n = 562$), internal ($n = 106$), and external validation cohort ($n = 287$). As are shown in the tables, the clinical characteristics of patients among the three cohorts have no statistical difference. PM(+) status occurred in 90 (16.0%) patients, 17 (16.0%) patients, and 62 (21.6%) patients in the training, internal, and external validation cohort, respectively (**Table 1**).

The inter- and intraclass correlation coefficients (ICCs) for the two radiologists' extracted features were both higher than 0.75, indicating that the inter- and intraobserver agreements of radiomics feature extraction of the two radiologists were good, so all outcomes were calculated on the basis of the measurements of the first radiologist. Details are summarized in **Supplementary Materials**.

Feature Selection and Radiomics Signature Building

On the basis of 562 patients in the primary cohort, 292 features were extracted from the CT image by using Matlab R2016a. Then, the 292 features were reduced to 11 potential predictors (**Figure S2**) by being featured with non-zero coefficients in the LASSO logistic regression model. Features that showed high collinearity with each other were excluded from analysis (**Figure S3**). Finally, four PM status-related features were selected for constructing a meaningful radiomics signature, presented with a Rad-score calculation formula: **Rad-score** = −1.429359e−03 × Eccentricity + 1.232216e−02 × Exten − 9.887834e−02 × GLCM_IMC1-0.5 + 8.977322e-02 × GLCM_MaximumProbability-0.5. Details are shown in **Supplementary Materials**. The associations among the clinicopathological characteristics, Rad-score, and PM status in different cohorts are showed in **Tables S1–S3**.

The relationships among PM status, Rad-score, and clinical characteristics in each cohort were determined by three heatmaps (**Figures S4–S6**). In the training and internal and external validation cohorts, significant positive relationship was found between Rad-score and PM status.

Diagnostic Validation of Radiomics Signature

In the training cohort, the Rad-score between PM(−) and PM(+) patients was significantly different ($P < 0.01$, **Figure 1**). Similarly, in the two validation cohorts, the Rad-score was also confirmed to be significantly different between PM(−) and PM(+) patients ($P < 0.01$, **Figure 1**). Higher Rad-scores were found in PM(+) patients in the training and internal and external validation cohorts consistently. The radiomics signature displayed an AUC for predicting PM status of 0.751 (95% CI, 0.703–0.799) in the training cohort and 0.802 (95% CI, 0.691–0.912) and 0.745 (95% CI, 0.683–0.806) in internal and external

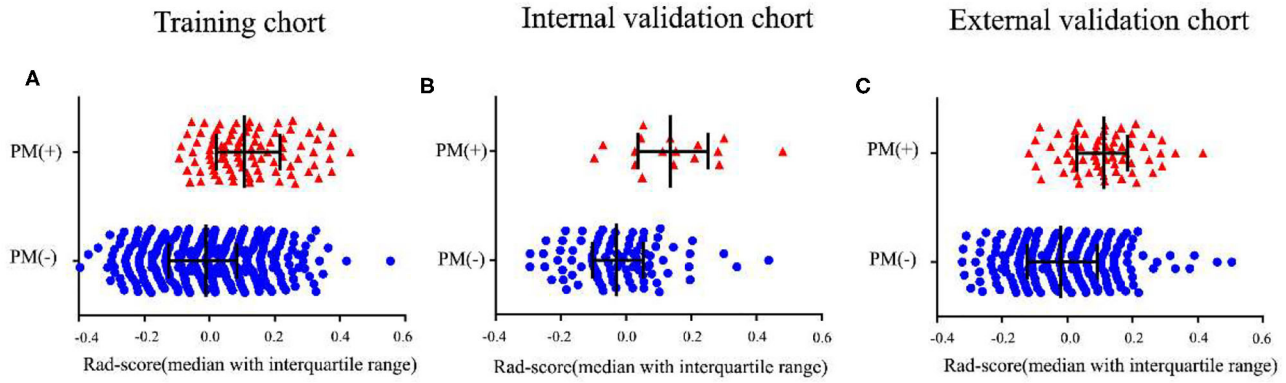


FIGURE 1 | Distribution of radiomics scores according to the PM status. **(A–C)** The values of the radiomics scores (Rad-scores) of each patient and the median values with interquartile range of Rad-scores. PM, peritoneal metastasis status.

TABLE 2 | Multivariate relationship between Rad-score and preoperative clinicopathological characteristics with peritoneal metastasis in each cohort.

Variables	PM(+) vs. PM(–)					
	OR (95%CI)	*P*	OR (95%CI)	*P*	OR (95%CI)	*P*
	Training cohort		Internal validation cohort		External validation cohort	
Rad-score	**6.364 (3.387–11.959)**	**<0.0001**	**13.151 (2.175–79.512)**	**0.005**	**6.544 (2.924–14.642)**	**<0.0001**
Location						
Cardia	0.362 (0.156–0.838)	0.018	0.925 (0.052–16.590)	0.958	0.797 (0.271–2.339)	0.679
Body	0.803 (0.335–1.923)	0.623	2.581 (0.306–21.744)	0.383	0.667 (0.206–2.156)	0.499
Antrum	0.877 (0.394–1.951)	0.747	1.846 (0.224–15.206)	0.569	0.746 (0.295–1.890)	0.537
Whole	**Reference**		**Reference**		**Reference**	
cT Stage						
T3	0.300 (0.120–0.754)	0.01	0.150 (0.020–1.103)	0.062	/	0.997
T4a	0.684 (0.395–1.184)	0.175	0.183 (0.040–0.842)	0.029	0.221 (0.076–0.640)	0.005
T4b	**Reference**		**Reference**		**Reference**	
cN Stage						
N0	0.268 (0.126–0.571)	0.001	/	0.998	0.177 (0.047–0.665)	0.01
N1	0.617 (0.315–1.208)	0.159	0.085 (0.007–1.071)	0.057	0.466 (0.169–1.284)	0.14
N2	0.935 (0.499–1.753)	0.835	1.150 (0.287–4.612)	0.843	1.750 (0.799–3.833)	0.162
N3	**Reference**		**Reference**		**Reference**	

PM(+), PM-positive status; PM(–), PM-negative status; CI, confidence interval; OR, odds ratio.

validation cohorts, respectively. Furthermore, when performing the stratified analysis according to clinicopathological risk factors, the Rad-score was still significantly associated with PM status in the training and internal and external validation cohorts (**Tables S4**–**S6**).

Development of Radiomics Nomogram

Using univariate analysis, we found that the radiomics signature was correlated with PM status significantly (**Table S7**). Variable values that demonstrated significance were used for multivariable analysis. In the training and internal and external validation cohorts, radiomics signature still remained an independent and powerful predictor for PM status in the multivariate logistic regression analysis (**Table 2**). According to the multivariate analysis, we developed a nomogram that integrated the radiomics signature and cT and cN stages in the training cohort (**Figure 2**). Using the nomogram, first draw a vertical line to the top point row to assign points for each variable; next, add the points from each variable together and drop a vertical line from the total points row to obtain the probability of PM status. For example, for a patient with a Rad-score of 0.4 and CT reported T4aN2 gastric cancer, the radiomics nomogram would predict a total score of more than 100, which indicates that the probability to suffer from peritoneal metastasis would be higher than 50%. The relationship between nomogram score and PM status is shown in **Figure S7**. Higher Nom-scores were found in PM(+) patients in the training cohorts and internal and external validation cohorts consistently (**Figure S7**).

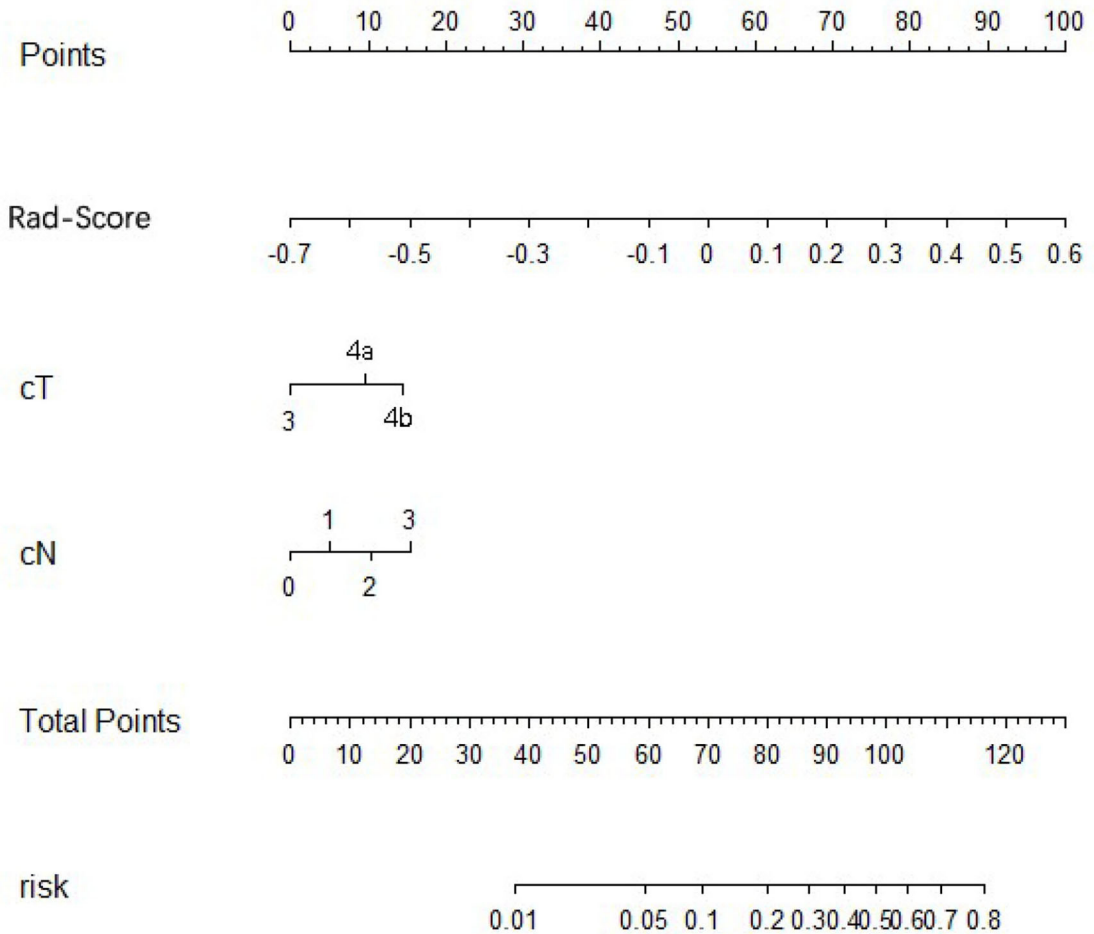


FIGURE 2 | Development of radiomics nomogram in training cohort. The radiomics nomogram incorporating the radiomics signature and cT stage and cN stages was developed in the training cohort.

Validation of the Nomogram

In the training cohort, an appropriate agreement between prediction and observation was yielded by the calibration curve of the radiomics nomogram (**Figure 3A**). ROC analysis showed good diagnostic performance of the Rad-score and nomogram in predicting gastric cancer peritoneal metastasis in each cohort (**Table S8**). The prediction performance of the model was moderate, with an AUC of 0.792 (95% CI, 0.748–0.836) in the training cohort (**Figure 4A**, **Table 3**, **Table S8**). In the validation cohort, it also displayed excellent prediction efficacy (**Figures 3B,C**), with AUCs of 0.870 (95% CI, 0.795–0.946) and 0.815 (95% CI, 0.763–0.867) in the internal and external validation cohorts, respectively (**Figures 4B,C**, **Table 3**, **Table S8**). The decision curve analysis for the nomogram in different cohorts are shown in **Figure 5**.

DISCUSSION

In this study, we first established a radiomics signature based on four texture features. This PM-related radiomics signature was obviously correlated with PM and was an independent predictor of PM status in GC. Second, we constructed the preoperative individualized prediction of PM status by developing and validating a radiomics nomogram that incorporated the radiomics signature and cT and cN stages. Both of the nomogram and the radiomics signature can be used to assist clinicians to predict peritoneal metastasis non-invasively.

We excluded patients with pathological T1–T3 tumors because, compared to pathological T4, these parts of patients are less expected to have PM (6, 7). If we had included patients with pT1–pT3 tumors, the total incidence rate of PM would decrease, reducing the specificity and sensitivity of the model for predicting PM. Actually, those CT-reported T3 or T4 patients are more likely to have the risk of PM (9), and most of them were eventually conformed T3 or T4 stage but less even none T1 or T2. All of the cancer centers in this study were in the same situation.

The LASSO method is a powerful method for the regression of high-dimensional predictors (31, 32). In this study, we shrank the regression coefficients with the LASSO method to examine

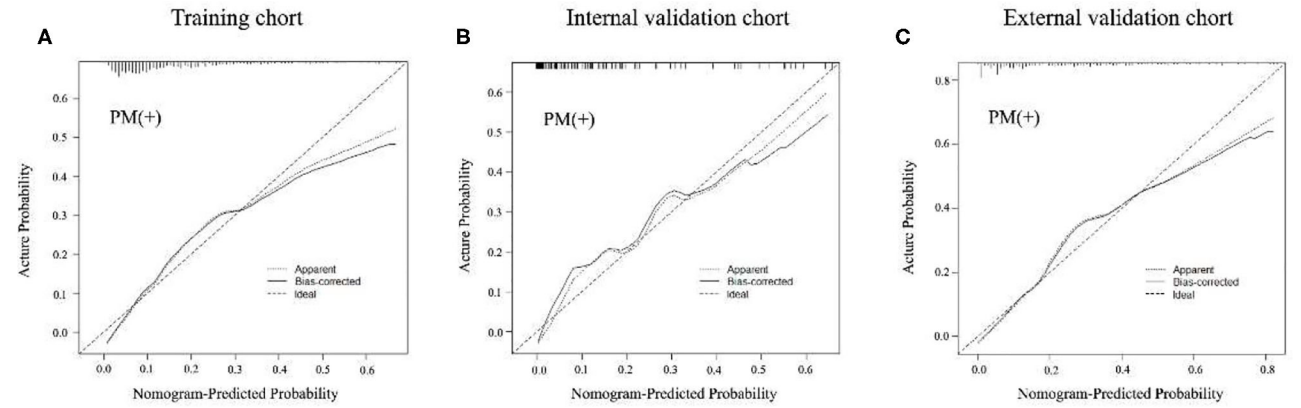


FIGURE 3 | Calibration curves of radiomics nomogram of peritoneal metastasis (PM) status prediction in training and internal and external validation cohorts. The calibration curves described the calibration in agreement between predicted and observed outcome. The 45-degree reference line means a perfect calibration with the outcome by ideal model. The solid line is the performance of the nomogram, without correction for overfit. The dotted line is the bootstrap-corrected performance of the nomogram, with a scatter estimate for future accuracy. **(A–C)** Calibration curves in the training cohort, internal and external validation cohorts.

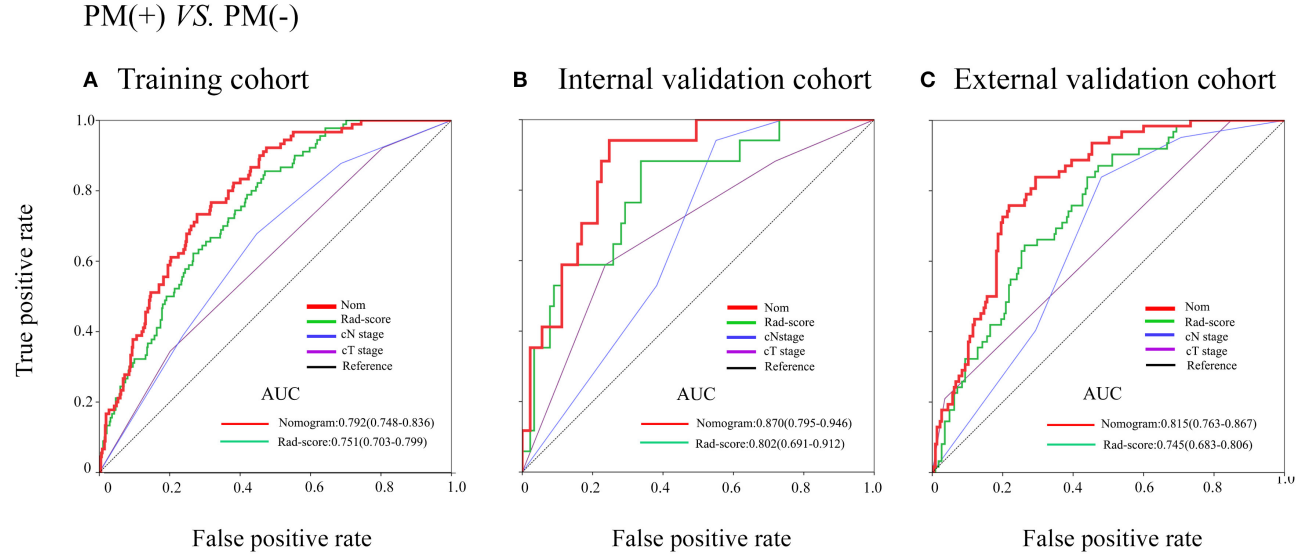


FIGURE 4 | Receiver operating characteristic (ROC) curves of the models in training and internal and external validation cohorts. **(A–C)** ROC curves in the training cohort, internal and external cohorts.

the predictor-outcome association, and as a result, 292 candidate radiomics features that were extracted from the primary CT image were reduced to 11 potential predictors. After addressing the collinearity, four PM-related features were selected for the construction of the radiomics signature.

Although CT is very popular and important in preoperative diagnosis, the accuracy of CT for preoperatively identifying PM status was very limited in patients with GC (9). PET-CT was a good method for predicting the LNM status preoperatively and had value on distant organ metastasis (33), and Findlay et al. also pointed out in their study that when staging patients with gastric cancer, 18F-fluorodeoxyglucose (18F-FDG) PET-CT could show useful information in identifying unsuspected metastasis (34); however, its accuracy for PM did not demonstrate an advantage over CT (33). Several studies have demonstrated that some clinicopathological factors, like CEA or CA19-9 level, size of the tumor, invasion depth, Borrmann type, and differentiation type, showed relationships with LNM (35). Some nomograms were developed for predicting LNM in patient with GC by using above clinicopathological factors, but outer validations still require to be applied on these nomograms. What is more, no particular nomogram has been used widely in clinical settings (17). Dong et al. found that PM was associated with the texture of the nearby peritoneum of tumor (36), but actually nearby peritoneum contains lots of positive-metastasis lymph nodes, which may increase the false positive rate of PM, as

the PM is also strongly associated with lymph node stage (9, 17). This had also been proven in our study. Hence, CT texture of the nearby peritoneum of the tumor still needs further examination.

CT images contain many medical information, and nowadays, radiologists could easily acquire complementary anatomical information of human tissues and definite tumor visually from CT images. However, there remain large amount of digital information for precise analysis, and different analytical tools could dig out different kinds of information. Recently, the rise in deep learning in medical research arouse hot topic among researchers, especially in disease detection and diagnosis (37, 38). Our previous study found that deep-learning-based CT image signature could help in predicting survival for patients with GC and in identifying which patients could benefit from adjuvant chemotherapy (39). However, deep learning requires large amount of sample for training. Radiomics, which can convert digital medical images to mineable data and analyze these data to improve detection, diagnosis, stage, and prediction power, may help us to improve the accuracy of detecting PM status preoperatively (11, 12, 15). Giganti et al. reported that texture features from DW-MRI and CE-MDCT could be promising non-invasive biomarker in evaluating the prognosis of gastric cancer (40), suggesting that texture analysis from medical images could facilitate clinical decision. Recently, Liu et al. reported that venous CT radiomics analysis could provide interesting information for predicting occult PM in gastric cancer (41), and before this study, a radiomics signature that could predict LNM in colorectal cancer was developed (17); therefore, we aimed to present a predictive model for preoperative prediction of PM status by connecting the preoperative clinicopathological factors and the radiomics features.

In East Asian countries, radical gastrectomy with chemotherapy is the standard treatment for advanced GC, and the usefulness of neoadjuvant chemotherapy is recently being measured (42, 43). In some clinical trials, preoperative chemotherapy was performed in GC patients with extensive metastasis. Recently, study found that patients with peritoneal metastases of gastric cancer may benefit from cytoreductive surgery (CRS) and hyperthermic intraperitoneal chemotherapy (HIPEC) (44). Therefore, the radiomics nomogram for preoperatively predicting PM status may contribute to make an adequate preoperative medical decision and select patients who could benefit from above treatment.

Four-feature radiomics signature and two preoperative clinical factors (cT and cN stages, which are easily obtained from CT) are integrated in our radiomics nomogram. According to the nomogram, the status of the disease could be comprehensively reflected, and the accuracy of prediction could be obviously improved. Calibration plots and ROC analysis were used to validate the nomogram. The nomogram showed excellent

TABLE 3 | Predictive accuracy of newly developed nomogram, Rad-score, and clinicopathological characteristics.

Variables	AUC (95% CI)
	PM(+) vs. PM(–)
TRAINING COHORT	
Nomogram	0.792 (0.748–0.836)
Rad-score	0.751 (0.703–0.799)
cN stage	0.639 (0.580–0.698)
cT stage	0.604 (0.542–0.667)
INTERNAL VALIDATION COHORT	
Nomogram	0.870 (0.795–0.946)
Rad-score	0.802 (0.691–0.912)
cN stage	0.669 (0.557–0.780)
cT stage	0.689 (0.548–0.830)
DATION COHORT	
Nomogram	0.815 (0.763–0.867)
Rad-score	0.745 (0.683–0.806)
cN stage	0.664 (0.597–0.731)
cT stage	0.647 (0.570–0.723)

AUC, area under the receiver operating characteristic curve; CI, confidence interval; PM, peritoneal metastasis; Rad-score, radiomics score.

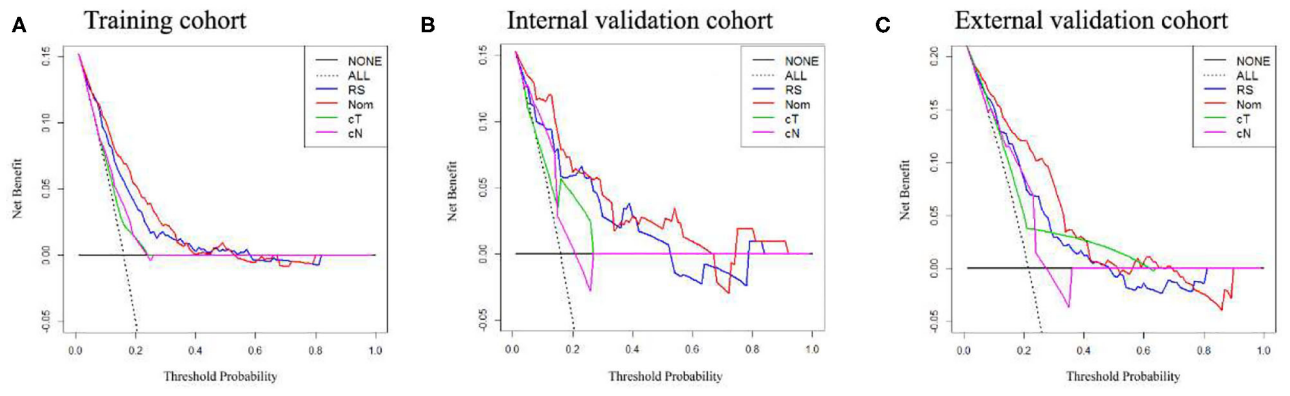


FIGURE 5 | Decision curve analysis of the radiomics nomogram, Rad-score, and clinical *T* and *N* stage in each cohort. **(A–C)** Calibration curves in the training cohort, internal and external validation cohorts.

prediction with a good calibration. What is more, high AUCs were demonstrated in our radiomics nomogram both in internal and external cohorts when predicting the PM status; this could provide more valuable information for determining the need for adjuvant therapy and the adequacy of surgical resection, thus aiding in pretreatment decision making.

In this study, our radiomics signature and nomogram could provide meaningful message for preoperatively predicting PM status. In the future, we will put a preoperative prediction model into effect to help provide proper surgical procedures or select candidates with high risk for laparoscopy exploration and treatment based on the comprehensive consideration of the information of the radiomics features.

However, there are still limitations in our study. Although the radiomics signature and nomogram could provide meaningful message for predicting PM, the nomograms were developed and externally validated in three retrospective data sets from two Chinese institutions; thus, these results need to be validated in a larger population with a multicenter and prospective study in the future, which could develop high-level evidence needed for clinical use. In addition, this predictive model is suitable for those who were preoperatively diagnosed cT3 or cT4, as these patients are more likely to have the risk of PM; however, clinical and pathological stages are sometimes inconsistent, especially in gastric cancer (45). For example, those who were diagnosed clinical T3 or T4 sometimes were found to be T1 or T2, and some patients who were diagnosed clinical T1 or T2 were confirmed T3 or T4 after surgery. For those who were pT3 or pT4 but were diagnosed cT1 or cT2, they would miss the opportunity to use this model for prediction. Furthermore, some serological matter like CA125 and HER-2 were not included in this study, as this study is retrospective and the above data were not available from each patient in that period.

In conclusion, this study demonstrated that the radiomics signature based on CT can be used as a predictor for predicting peritoneal metastasis in GC patients. Besides, this study revealed that the radiomics nomogram, which combined the clinicopathological risk factors with the radiomics signature, can be effectively used to promote the preoperative individualized prediction of PM status in patients with GC.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Nanfang Hospital, Nanfang Hospital, Southern Medical University; the Ethics committee of Sun Yat-sen University Cancer Center, Sun Yat-sen University Cancer Center. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

GL, YX, and ZZ: guarantor of the article. GL, YX, ZZ, YJ, WH, and KZ: conception and design. YJ, WH, KZ, CC, QY, ZH, SY, YH, JX, JY, and HL: collection and assembly of data. WH, YJ, RX, ZH, JX, ZS, and JY: data analysis and interpretation. All authors: manuscript writing and final approval of manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from National Natural Science Foundation of China (81872013, 81672446, 81600510), The National Key Research and Development Program of China (2017YFC0108300), Natural Science Foundation of Guangdong Province (2019A1515011445), Outstanding Youth Development Scheme of Nanfang Hospital, Southern Medical University (2018J007), and Director's Foundation of Nanfang Hospital (2016B010).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.01416/full#supplementary-material

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012). *CA A Cancer J Clin.* (2015) 65:87–108. doi: 10.3322/caac.21262
2. Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015. *JAMA Oncol.* (2017) 3:524–8. doi: 10.1001/jamaoncol.2016.5688
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
4. Menges M, Hoehler T. Current strategies in systemic treatment of gastric cancer and cancer of the gastroesophageal junction. *J Cancer Res Clin.* (2009) 135:29–38. doi: 10.1007/s00432-008-0425-z
5. Fujitani K, Yang H, Mizusawa J, Kim Y, Terashima M Han S, et al. Gastrectomy plus chemotherapy versus chemotherapy alone for advanced gastric cancer with a single non-curable factor (REGATTA): a phase 3, randomised controlled trial. *Lancet Oncology.* (2016) 17:309–18. doi: 10.1016/S1470-2045(15)00553-7
6. Gretschel S, Siegel R, Estevez-Schwarz L, Hunerbein M, Schneider U, Schlag PM. Surgical strategies for gastric cancer with synchronous peritoneal carcinomatosis. *Br J Surg.* (2006) 93:1530–5. doi: 10.1002/bjs.5513
7. Zhang X, Liu X, Sun F, Li S, Gao W, Wang Y. Greater omental milky spot examination for diagnosis of peritoneal metastasis in gastric cancer

patients. *J Laparoendosc Adv Surg Tech A*. (2017) 27:106–9. doi: 10.1089/lap.2016.0295
8. Sawaki K, Kanda M, Miwa T, Umeda S, Tanaka H Tanaka C, et al. Troponin I2 as a specific biomarker for prediction of peritoneal metastasis in gastric cancer. *Ann Surg Oncol.* (2018) 25:2083–90. doi: 10.1245/s10434-018-6480-z
9. Kim SJ, Kim HH, Kim YH, Hwang SH, Lee HS Park DJ, et al. Peritoneal metastasis: detection with 16- or 64-detector row CT in patients undergoing surgery for gastric cancer. *Radiology.* (2009) 253:407–15. doi: 10.1148/radiol.2532082272
10. Yip SS, Aerts HJ. Applications and limitations of radiomics. *Phys Med Biol.* (2016) 61:R150–66. doi: 10.1088/0031-9155/61/13/R150
11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
12. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
13. Banerjee S, Wang DS, Kim HJ, Sirlin CB, Chan MG, Korn RL, et al. A computed tomography radiogenomic biomarker predicts microvascular invasion and clinical outcomes in hepatocellular carcinoma. *Hepatology.* (2015) 62:792–800. doi: 10.1002/hep.27877
14. O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol.* (2017) 14:169–86. doi: 10.1038/nrclinonc.2016.162
15. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RTH, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol.* (2015) 114:345–50. doi: 10.1016/j.radonc.2015.02.015
16. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2014) 5:5644. doi: 10.1038/ncomms5644
17. Huang Y, Liang C, He L, Tian J, Liang C, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol.* (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
18. Giganti F, Antunes S, Salerno A, Ambrosi A, Marra P, Nicoletti R, et al. Gastric cancer: texture analysis from multidetector computed tomography as a potential preoperative prognostic biomarker. *Eur Radiol.* (2017) 27:1831–9. doi: 10.1007/s00330-016-4540-y
19. Jiang Y, Chen C, Xie J, Wang W, Zha X, Lv W, et al. Radiomics signature of computed tomography imaging for prediction of survival and chemotherapeutic benefits in gastric cancer. *Ebiomedicine.* (2018) 36:171–82. doi: 10.1016/j.ebiom.2018.09.007
20. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage.* (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015
21. Daghir-Wojtkowiak E, Wiczling P, Bocian S, Kubik Ł, Koślinski P, Buszewski B, et al. Least absolute shrinkage and selection operator and dimensionality reduction techniques in quantitative structure retention relationship modeling of retention in hydrophilic interaction liquid chromatography. *J Chromatogr A*. (2015) 1403:54–62. doi: 10.1016/j.chroma.2015.05.025
22. Kocak B, Durmaz ES, Ates E, Kilickesmez O. Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol.* (2019) 25:485–95. doi: 10.5152/dir.2019.19321
23. Sheikh K, Lee SH, Cheng Z, Lakshminarayanan P, Peng L, Han P, et al. Predicting acute radiation induced xerostomia in head and neck cancer using MR and CT radiomics of parotid and submandibular glands. *Radiat Oncol.* (2019) 14:131. doi: 10.1186/s13014-019-1339-4
24. Kocak B, Durmaz ES, Ates E, Kaya OK, Kilickesmez O. Unenhanced CT texture analysis of clear cell renal cell carcinomas: a machine learning–based study for predicting histopathologic nuclear grade. *Am J Roentgenol.* (2019) 212:W132–9. doi: 10.2214/AJR.18.20742
25. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* (2010) 33:1–22. doi: 10.18637/jss.v033.i01
26. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Softw.* (2011) 39:1–13. doi: 10.18637/jss.v039.i05
27. Schuit E, Kwee A, Westerhuis M, van Dessel H, Graziosi G, van Lith J, et al. A clinical prediction model to assess the risk of operative delivery. *BJOG.* (2012) 119:915–23. doi: 10.1111/j.1471-0528.2012.03334.x
28. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* (2016) 26:565–74. doi: 10.1177/0272989X06295361
29. Jiang Y, Yuan Q, Lv W, Xi S, Huang W, Sun Z, et al. Radiomic signature of 18 F fluorodeoxyglucose PET/CT for prediction of gastric cancer survival and chemotherapeutic benefits. *Theranostics.* (2018) 8:5915–28. doi: 10.7150/thno.28018
30. Haga A, Takahashi W, Aoki S, Nawa K, Yamashita H, Abe O, et al. Standardization of imaging features for radiomics analysis. *J Med Invest.* (2019) 66:35–7. doi: 10.2152/jmi.66.35
31. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* (1997) 16:385–95. doi: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3
32. Zhang JM, Song WP, Chen ZM, Wei JM, Liao YP, Lei JP, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol.* (2013) 14:1295–306. doi: 10.1016/S1470-2045(13)70491-1
33. Lee DH, Kim SH, Joo I, Hur BY, Han JK. Comparison between 18F-FDG PET/MRI and MDCT for the assessment of preoperative staging and resectability of gastric cancer. *Eur J Radiol.* (2016) 85:1085–91. doi: 10.1016/j.ejrad.2016.03.015
34. Findlay JM, Antonowicz S, Segaran A, El Kafsi J, Zhang A, Bradley KM, et al. Routinely staging gastric cancer with 18F-FDG PET-CT detects additional metastases and predicts early recurrence and death after surgery. *Eur Radiol.* (2019) 29:2490–8. doi: 10.1007/s00330-018-5904-2
35. Yamaguchi H, Satoh Y, Ishigami H, Kurihara M, Yatomi Y, Kitayama J. Peritoneal lavage CEA mRNA levels predict conversion gastrectomy outcomes after induction chemotherapy with intraperitoneal paclitaxel in gastric cancer patients with peritoneal metastasis. *Ann Surg Oncol.* (2017) 24:3345–52. doi: 10.1245/s10434-017-5997-x
36. Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol.* (2019) 30:431–8. doi: 10.1093/annonc/mdz001
37. Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, et al. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLos Med.* (2018) 15:e1002711. doi: 10.1371/journal.pmed.1002711
38. Peng H, Dong D, Fang M, Li L, Tang L, Chen L, et al. Prognostic value of deep learning PET/CT-based radiomics: potential role for future individual induction chemotherapy in advanced nasopharyngeal carcinoma. *Clin Cancer Res.* (2019) 25:4271–9. doi: 10.1158/1078-0432.CCR-18-3065
39. Jiang Y, Jin C, Yu H, Wu J, Chen C, Yuan Q, et al. Development and validation of a deep learning CT signature to predict survival and chemotherapy benefit in gastric cancer: a multicenter, retrospective study. *Ann Surg.* (2020). doi: 10.1097/SLA.0000000000003778. [Epub ahead of print].
40. Giganti F, Tang L, Baba H. Gastric cancer and imaging biomarkers: part 1-a critical review of DW-MRI and CE-MDCT findings. *Eur Radiol.* (2019) 29:1743–53. doi: 10.1007/s00330-018-5732-4
41. Liu S, He J, Liu S, Ji C, Guan W, Chen L, et al. Radiomics analysis using contrast-enhanced CT for preoperative prediction of occult peritoneal metastasis in advanced gastric cancer. *Eur Radiol.* (2019) 30:239–46. doi: 10.2139/ssrn.3297887
42. Inoue T, Yachida S, Usuki H, Kimura T, Hagiike M, Okano K, et al. Pilot feasibility study of neoadjuvant chemoradiotherapy with S-1 in patients with locally advanced gastric cancer featuring adjacent tissue invasion or

JGCA bulky N2 lymph node metastases. *Ann Surg Oncol.* (2012) 19:2937–45. doi: 10.1245/s10434-012-2332-4

43. Japanese Gastric Cancer Association. Japanese gastric cancer treatment guidelines 2014 (ver. 4). *Gastric Cancer.* (2017) 20:1–19. doi: 10.1007/s10120-016-0622-4
44. Rau B, Brandl A, Thuss-Patience P, Bergner F, Raue W, Arnold A, et al. The efficacy of treatment options for patients with gastric cancer and peritoneal metastasis. *Gastric Cancer.* (2019) 22:1226–37. doi: 10.1007/s10120-019-00969-1
45. Burbidge S, Mahady K, Naik K. The role of CT and staging laparoscopy in the staging of gastric cancer. *Clin Radiol.* (2013) 68:251–5. doi: 10.1016/j.crad.2012.07.015

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 15 September 2020
doi: 10.3389/fonc.2020.01629

Clinical Trials for Artificial Intelligence in Cancer Diagnosis: A Cross-Sectional Study of Registered Trials in ClinicalTrials.gov

*Jingsi Dong[1], Yingcai Geng[1], Dan Lu[2], Bingjie Li[1], Long Tian[1], Dan Lin[1] and Yonggang Zhang[3,4]**

[1] Department of Lung Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [2] Department of Otorhinolaryngology, Head and Neck Surgery, West China Hospital, Sichuan University, Chengdu, China, [3] Department of Periodical Press and National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, Chengdu, China, [4] Chinese Evidence-Based Medicine Center, West China Hospital, Sichuan University, Chengdu, China

OPEN ACCESS

Edited by:
Chunxiao Guo,
University of Minnesota, United States

Reviewed by:
Song Xu,
Tianjin Medical University General Hospital, China
Yifeng Zhou,
Soochow University, China
Xiaoxing Li,
Sun Yat-sen University, China
Louis Joseph Vaickus,
Dartmouth College, United States
David Valle-Cruz,
Universidad Autónoma del Estado de México, Mexico

***Correspondence:**
Yonggang Zhang
jebm_zhang@yahoo.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 19 March 2020
Accepted: 27 July 2020
Published: 15 September 2020

Citation:
Dong J, Geng Y, Lu D, Li B, Tian L, Lin D and Zhang Y (2020) Clinical Trials for Artificial Intelligence in Cancer Diagnosis: A Cross-Sectional Study of Registered Trials in ClinicalTrials.gov. Front. Oncol. 10:1629. doi: 10.3389/fonc.2020.01629

Objective: Clinical trials are the most effective way to judge the merits of diagnosis and treatment strategies. The in-depth mining of clinical trial data enables us to grasp the application trend of artificial intelligence (AI) for cancer diagnosis. The aim of this study was to analyze the characteristics of registered trials on AI for cancer diagnosis.

Methods: Clinical trials on AI for cancer diagnosis registered on the ClinicalTrials.gov database were searched and downloaded. Statistical analysis was performed by using SPSS 20.0 software.

Results: A total of 97 registered trials were included. Of them, only 27 (27.8%) were interventional trials and 70 (72.1%) were observational trials. Fifteen (15.4%) trials had been completed. Fifty trials were in recruitment, and another 18 remained unrecruited. The number of cases included in the clinical trials tended to be large, 31 (32.0%) trials including samples ranging from 100 to 499 cases and 17 (17.5%) trials including samples ranging from 500 to 999 cases. Of the 27 interventional trials, only two trials reported trials' phase. Most (85.2%) interventional trials were for diagnosis, and a few (3.7%) were for the purpose of both the diagnosis and therapy of cancers. For the observational clinical trials, 46 (65.7%) were cohort studies, and 11 (15.7%) were case-only studies. Among the observational trials, 46 (65.7%) were prospective studies and 13 (18.6%) were retrospective studies. Among 97 trials, 37 (38.1%) involved colorectal cancer, 11 (11.3%) involved breast cancer, 43 (44.3%) were for imaging diagnosis, 33 (34.0%) were for endoscopic diagnosis, and 11 (11.3%) were for pathological diagnosis. For the interventional trials, 11 trials were parallel assignment (40.7%), and 14 were single group assignment (51.9%). Among the 27 interventional trials, 18 (66.7%) trials were performed without masking, 6 (22.2%) trials were performed with single masking, only 1 (3.7%) was performed with double masking, and 2 (7.4%) was performed with triple masking.

Conclusion: It appears that most registered trials on AI for cancer diagnosis are observational design, and more trials are needed in this field.

Keywords: clinical trial, diagnosis, artificial intelligence, ClinicalTrials.gov, cancer

INTRODUCTION

With the development of new computing methods combined with the availability of training data, the application of powerful mathematical algorithms in the field of artificial intelligence (AI) has been promoted. It is hard to define AI precisely. It has been suggested that a machine is intelligent if its working behavior is indistinguishable from that of a human being (1). In modern concepts, AI refers to the ability of machines to communicate, reason, and operate independently at work in a manner similar to that of humans. AI programs have been developed and applied in many medical areas, including diagnosis, treatment, drug development, and patient cares; in addition, there are increased researches regarding AI in various specialties, especially in cancer diagnosis (2–5).

Because cancer is still the leading cause of death worldwide (6), accurate diagnosis of cancers is essentially important. In terms of the imaging diagnosis of tumors (7) [i.e., pathological diagnosis (8) and endoscopic diagnosis (9)], the performance of AI is as good as that of human experts. In the era of big data, medical activities are accelerating to produce a vast amount of health-disease data (10). With the help of AI, doctors can provide medical services to patients more efficiently and accurately (11, 12). At present, the most critical problem for AI application in imaging is that there is no gold standard of AI for cancer diagnosis (12); thus, many researchers performed trials to assess AI for cancer diagnosis, because trials are the most effective way to judge the merits of diagnosis and treatment strategies (13). Most trials were registered in ClinicalTrials.gov, which is a public clinical trial registry platform jointly launched by the US Food and Drug Administration (FDA) and the US National Library of Medicine (NLM). Studying characteristics of registered trials will help to know the development of trials in specific filed. Up to now, there is no such study on AI for cancer diagnosis; thus, we performed the current study. The aim of this study was to analyze and summarize the characteristics of AI for cancer diagnosis. The in-depth mining of clinical trial data from ClinicalTrials.gov enables us to grasp the application trend of AI in cancer diagnosis earlier.

MATERIALS AND METHODS

Inclusion Criteria and Exclusion Criteria

The inclusion criteria were registered trials on AI for cancer diagnosis; in each trial, cancer can be a single cancer or all kinds of cancer. The exclusion criteria were AI in purely therapeutic applications and incomplete registration information.

Data Search

According to our previous studies (14, 15), we search the ClinicalTrials.gov on February 20, 2020, for trials on AI for cancer diagnosis. In case of missing any trials, we used the following words: artificial intelligence, deep learning, machine learning, etc. All searched results were downloaded. The data were updated on June 18, 2020.

Trial Screening and Data Extraction

Two authors independently screened the trials according to the inclusion and exclusion criteria. In case of any disagreement, discussion was performed. And then, two authors independently extracted the data of the included trials. The following information was extracted: registered number, study type, start date, status of the trial, study result, study sample, participant age, primary sponsor, location, primary purpose, phases of the trial, allocation, intervention model, masking and intervention, and types of cancer.

Data Analysis

The methodology of the study is similar to our previous study (14). This is a cross-sectional study, so a descriptive analysis was used to analyze the characteristics of registered trials. The outcomes included year, status of trials, study results, age, enrolment, sponsor, location, funding source, characteristics of study designs, type of cancer, and application method. Statistical analysis was performed by using SPSS 20.0 software. P-value <0.05 was considered statistically significant.

RESULTS

The Characteristics of the Included Trials

On June 18, 2020, 884 results were searched from the ClinicalTrials.gov website. After a careful review of the clinical trial information, 787 results were excluded. Finally, a total of 97 trials were included in this study.

The characteristics of the included trials are shown in **Table 1**. Of the 97 trials, only 27 (27.8%) were interventional trials and 70 (72.1%) were observational trials. Fifteen (15.4%) trials were completed, the largest proportion of trials (50 trials) were in recruitment, and another 18 trials remained unrecruited. None of the trials had available results. Eighty-seven (89.7%) trials included subjects over the age of 18, and 10 trials (10.3%) included patients of all ages. The number of cases included in the trials tended to be large, with 31 (32.0%) trials including samples ranging from 100 to 499 cases and 17 (17.5%) trials including samples ranging from 500 to 999 cases. Universities were listed as the primary sponsor for 57 (58.8%) trials, hospitals for 30 (30.9%) trials, and industries for 10 (10.3%) trials. Of all the trials, 48 (49.5%) trials were performed in Asia, 15 in Europe, 33 in North America, and 1 in Australia.

Characteristics of the Study Design

The characteristics of interventional trials are displayed in **Table 2**. Of the 27 interventional trials, only two trials reported phase (1 in phase 1 and 1 in phase 3), and other trials did not report phases. Among all of the interventional trials, most (85.2%) of the interventional trials were for diagnosis, a few (3.7%) were for the purpose of both the diagnosis and therapy of cancers, 2 (7.4%) trials were for the primary purpose of screening, and 1 (3.7%) trial was for the primary purpose of device feasibility. Among all of the interventional trials, there were 10 randomized trials (37%) and 4 (14.8%) non-randomized trials, and 13 (48.1%) trials did not mention the allocation value. There were 11 parallel assignment (40.7%) and 14 single

TABLE 1 | Characteristics of included trials.

Characteristics	Number	Percentage (%)
Study type		
Interventional	27	27.8
Observational	70	72.1
Year		
2007–2016	11	11.3
2017–2018	29	29.9
2019–2020	57	58.8
Status		
Completed	15	15.4
Recruiting	50	51.5
Active, not recruiting	6	6.1
Not yet recruiting	18	18.5
Unknown status	7	7.2
Withdrawn	1	1.0
Study results		
Has available results	0	0
No available results	97	100
Participant age (y)		
1 to older	10	10.3
Older than 18	87	89.7
Enrollment		
<99	17	17.5
100–499	31	32.0
500–999	17	17.5
>999	32	33.0
Sponsor		
University	57	58.8
Hospital	30	30.9
Industry	10	10.3
Location		
Europe	15	15.5
North America	33	34.0
Asia	48	49.5
Australia	1	1.0
Funded by		
Other		
Industry	8	8.2
Other and industry	6	6.2
Other	83	85.6

TABLE 2 | Study design elements of interventional trials.

Characteristics	Number	Percentage (%)
Primary purpose		
Diagnostic and treatment	1	3.7
Diagnostic	23	85.2
Screening	2	7.4
Device feasibility	1	3.7
Phase		
Phase 1	1	3.7
Phase 2	None	–
Phase 3	1	3.7
Phase 4	None	–
Not applicable	25	92.6
Allocation		
Randomized	10	37.0
Non-randomized	4	14.8
Missing value	13	48.1
Intervention model		
Parallel assignment	11	40.7
Sequential assignment	1	3.7
Crossover assignment	1	3.7
Single group assignment	14	51.9
Masking		
Single	6	22.2
Double	1	3.7
Triple	2	7.4
Without	18	66.7

TABLE 3 | Study design elements of observational trials.

Characteristics	Number	Percentage (%)
Observational model		
Case-only	11	15.7
Cohort	46	65.7
Defined Population	1	1.4
Other	12	17.1
Time perspective		
Prospective	46	65.7
Retrospective	13	18.6
Cross- Sectional	3	4.3
Other	8	11.4

group assignment (51.9%). Among the 27 interventional trials, 18 (66.7%) trials were performed without masking, six (22.2%) trials were performed with single masking, only one (3.7%) was performed with double masking, and two (7.4%) was performed with triple masking.

Table 3 shows the characteristics of the observational trials ($n = 70$). For the observational trials, 46 (65.7%) were cohort studies, 11 (15.7%) were case-only studies, 1 (1.4%) was defined population, and 12 (17.1%) were other. Among the observational trials, 46 (65.7%) were prospective studies and 13 (18.6%) were retrospective studies.

Overview of Clinical Trials for Diagnosis

Table 4 shows the overview of trials for diagnosis. All 97 trials were designed for a variety of cancers, 37 trials (38.1%) were for colorectal cancer, and 11 (11.3%) were for breast cancer. Of the 97 trials, 43 (44.3%) were for imaging diagnosis, 33 (34%) for endoscopic diagnosis, and 11 (11.3%) for pathological diagnosis. To verify whether colonoscopy would be much more effective with the assistance of an automatic quality control system (AQCS), a prospective interventional trial was performed (NCT03622281). The enrolled patients were randomly assigned into the AQCS group and the control group who received

TABLE 4 | Overview of clinical trials in diagnosis.

Characteristics	Number	Percentage (%)
Tumor types		
Breast Cancer	11	11.3
Colorectal Cancer	37	38.1
Esophageal Cancer	3	3.1
Gastrointestinal Cancer	3	3.1
Glioma	6	6.2
Ovarian	2	2.1
Prostate	3	3.1
Pituitary	2	2.1
Lung Cancer	10	10.3
Skin Cancer	6	6.2
Other Cancer	14	14.4
Application method		
Endoscopy	33	34.0
Imaging	43	44.3
Pathology	11	11.3
Biomarker	6	6.2
Biopsy	4	4.1

colonoscopy without AQCS. An increased adenoma detection rate was seen in the AQCS group, which indicated that AQCS could practically improve the accuracy of colonoscopy (16).

Another trial with published results aimed to confirm whether a designed chatbot was not inferior to physicians regarding the satisfaction of breast cancer patients with the information provided (NCT03556813). Two groups of randomly assigned patients asked 12 predefined questions that were previously answered by a chatbot or a medical committee and received the response from either a chatbot or a physician. The chatbot group had higher success rates (69 vs. 64%) than the physician group, which showed noninferiority ($P < 0.001$) (17).

DISCUSSION

This study provided an assessment of the clinical trials registered on ClinicalTrials.gov about AI diagnosis in cancer. Most trials were observational, and only a few were interventional. Most trials were registered after 2017, indicating that the application of AI for cancer diagnosis was a new technology. More than half of the trials were in recruitment, and only one trial published available results. Most trials tended to be large sample size studies, and only a few studies had an expected sample size of <100. Most trials were sponsored by universities and hospitals. Notably, the vast majority ($n = 48$) of trials were conducted in Asia, with only 15 trials initiated in Europe and 33 trials initiated in North America. Most interventional trials used randomization, but most did not use blinding methods. Most observational trials were prospective design.

Ten trials were designed to diagnose lung cancer. In terms of the imaging diagnosis of lung cancer, it was reported that machine learning could predict the histological type of lung cancer through the imaging characteristics of PET/CT (18). In 2017, Song et al. (19) reported that CT imaging features could be used to predict the pathological type of micropapillary adenocarcinoma, but whether it could be recognized by AI had not been reported (20). In 2016–2017, Luo et al. (21) and Yu et al. (22) reported that the automatic analysis method of AI could perform pathological image analysis to predict the prognosis of patients with lung cancer. The CT diagnosis and pathological diagnosis of lung cancer are important prerequisites for the treatment of lung cancer (23–25). It was expected that AI could provide more functions for accurate diagnosis in the future.

In reviewing all included trials, the highest proportion trials were for colorectal cancer (37, 38.1%). This suggests that application of AI in diagnosis of colorectal cancer is a hot topic. Because the overall rate of missed polyps is as high as 22%, the associated colorectal cancer after colonoscopy is of concern (26). Computer-aided diagnosis (CAD) offers a promising solution for reducing the rate of missed diagnoses with colonoscopy (27). AI technology must address a number of important issues before it can be incorporated into routine clinical practice. The key stages for the implementation of CAD technology in routine colonoscopy have been detailed elsewhere, particularly by Mori et al. (28), who described the following steps: product development and feasibility studies, clinical trials, regulatory approval, and insurance reimbursement (29).

Eleven (11.3%) trials were for the pathological diagnosis of cancer. In 2019, Chen et al. (30) reported that thanks to ARM technology, AI can be integrated into the microscopic workflow to improve the efficiency and consistency of the microscopic inspection of biological specimens. The technology would be used to diagnose cancers. Among all the studies of AI in the field of cancer pathological diagnosis, the implementation in breast cancer was earlier and more mature (8), and AI diagnosis had an excellent application in the diagnosis of primary breast cancer and metastatic breast cancer (31). In terms of cancer imaging diagnosis, AI could also improve the specificity of diagnosis by integrating patient information and image analysis (32). In this study, we found that the diagnostic modes of lung cancer were mainly imaging diagnosis and pathological diagnosis. The diagnostic mode of colorectal cancer was mainly endoscopic diagnosis. The diagnostic modes of breast cancer were mainly imaging diagnosis and pathological diagnosis. In addition, both published studies used the randomization method, suggesting that the comparison of AI and traditional methods in tumor diagnosis was more operable in trials.

With the aid of AI, the detection rate of polyps and adenomas under endoscopy will be greatly improved (33). This apparent advantage, however, remains to be demonstrated in multicenter studies. The deficiency of AI in cancer diagnosis is also obvious. Chatbots designed to aid diagnosis could communicate with breast cancer patients like doctors, showing the potential to help doctors. But chatbot's questions are too routine to fully help doctors' diagnoses and decisions. The main purpose of this study is to understand the current situation of the application of AI in the field of medicine, which has a good hint to the scholars in related fields. Therefore, our study did not focus on the specific results of each trial.

The study had several limitations. First, not all studies were registered on ClinicalTrials.gov, which limits the application of the results in the future. However, ClinicalTrials.gov contained more than 80% of the trials on the World Health Organization's International Clinical Trials Registry Platform (34). Therefore, even if our study cannot cover all trials, it still reflected the mainstream of such trials. Second, the study analyzed data from clinical trials in which AI was used to diagnose cancer, but due to the short time span of the clinical trials, most clinical trials have not published results, so the analysis of results was insufficient. Third, this study included the application of AI in the diagnosis of all cancer types, but as only 97 trials were included, specific cancer types were not targeted for detailed analysis.

In conclusion, the current study presents the characteristics of registered trials on AI for cancer diagnosis. It suggested that more trials are needed to provide more evidence.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article.

AUTHOR CONTRIBUTIONS

YZ designed the study. JD and YG provided the source for the study and edited the manuscript. JD, LT, and DLi searched, extracted, assessed, and analyzed the data and drafted the manuscript. JD, BL, and DLu analyzed the data. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partly supported by the Health Commission of Sichuan Province (18PJ432).

REFERENCES

1. Turing AM. Computing machinery and intelligence. *Mind LIX*. (1950) 433–60. doi: 10.1093/mind/LIX.236.433
2. Adamson AS, Welch HG. Machine learning and the cancer-diagnosis problem - no gold standard. *N Engl J Med*. (2019) 381:2285–7. doi: 10.1056/NEJMp1907407
3. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. (2019) 69:127–57. doi: 10.3322/caac.21552
4. Mori Y, Berzin TM, Kudo SE. Artificial intelligence for early gastric cancer: early promise and the path ahead. *Gastrointest Endosc*. (2019) 89:816–7. doi: 10.1016/j.gie.2018.12.019
5. Strom P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol*. (2020) 21:222–32. doi: 10.1016/S1470-2045(19)30738-7
6. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. (2015) 65:87–108. doi: 10.3322/caac.21262
7. Mak RH, Endres MG, Paik JH, Sergeev RA, Aerts H, Williams CL, et al. Use of crowd innovation to develop an artificial intelligence-based solution for radiation therapy targeting. *JAMA Oncol*. (2019) 5:654–61. doi: 10.1001/jamaoncol.2019.0159
8. Ehteshami Bejnordi B, Veta M, Johannes Van Diest P, Van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*. (2017) 318:2199–210. doi: 10.1001/jama.2017.14585
9. Rees CJ, Koo S. Artificial intelligence - upping the game in gastrointestinal endoscopy? *Nat Rev Gastroenterol Hepatol*. (2019) 16:584–5. doi: 10.1038/s41575-019-0178-y
10. Zeng T, Huang T, Lu C. Editorial: machine learning advanced dynamic omics data analysis for precision medicine. *Front Genet*. (2019) 10:1343. doi: 10.3389/fgene.2019.01343
11. Obermeyer Z, Lee TH. Lost in thought - the limits of the human mind and the future of medicine. *N Engl J Med*. (2017) 377:1209–11. doi: 10.1056/NEJMp1705348
12. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. (2019) 380:1347–58. doi: 10.1056/NEJMra1814259
13. Feizabadi M, Fahimnia F, Mosavi Jarrahi A, Naghshineh N, Tofighi S. Iranian clinical trials: an analysis of registered trials in International Clinical Trial Registry Platform. (ICTRP). *J Evid Based Med*. (2017) 10:91–6. doi: 10.1111/jebm.12248
14. Chen L, Su Y, Quan L, Zhang Y, Du L. Clinical trials focusing on drug control and prevention of ventilator-associated pneumonia: a comprehensive analysis of trials registered on ClinicalTrials.gov. *Front Pharmacol*. (2018) 9:1574. doi: 10.3389/fphar.2018.01574
15. Chen, L., Wang, M., Yang, Y., Shen, J., and Zhang, Y. (2020). Registered interventional clinical trials for old populations with infectious diseases on clinicaltrials.gov: a cross-sectional study. *Front. Pharmacol*. 11:942. doi: 10.3389/fphar.2020.00942
16. Su JR, Li Z, Shao XJ, Ji CR, Ji R, Zhou RC, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). *Gastrointest Endosc*. (2020) 91:415–24 e414. doi: 10.1016/j.gie.2019.08.026
17. Bibault JE, Chaix B, Guillemasse A, Cousin S, Escande A, Perrin M, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. *J Med Internet Res*. (2019) 21:e15787. doi: 10.2196/15787
18. Hyun SH, Ahn MS, Koh YW, Lee SJ. A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer. *Clin Nucl Med*. (2019) 44:956–60. doi: 10.1097/RLU.0000000000002810
19. Song SH, Park H, Lee G, Lee HY, Sohn I, Kim HS, et al. Imaging phenotyping using radiomics to predict micropapillary pattern within lung adenocarcinoma. *J Thorac Oncol*. (2017) 12:624–32. doi: 10.1016/j.jtho.2016.11.2230
20. Jacobs C, Van Ginneken B. Google's lung cancer AI: a promising tool that needs further validation. *Nat Rev Clin Oncol*. (2019) 16:532–3. doi: 10.1038/s41571-019-0248-7
21. Luo X, Zang X, Yang L, Huang J, Liang F, Rodriguez-Canales J, et al. Comprehensive computational pathological image analysis predicts lung cancer prognosis. *J Thorac Oncol*. (2017) 12:501–9. doi: 10.1016/j.jtho.2016.10.017
22. Yu KH, Zhang C, Berry GJ, Altman RB, Re C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun*. (2016) 7:12474. doi: 10.1038/ncomms12474
23. Gu J, Lu C, Guo J, Chen L, Chu Y, Ji Y, et al. Prognostic significance of the IASLC/ATS/ERS classification in Chinese patients-A single institution retrospective study of 292 lung adenocarcinoma. *J Surg Oncol*. (2013) 107:474–80. doi: 10.1002/jso.23259
24. Hung JJ, Yeh YC, Jeng WJ, Wu KJ, Huang BS, Wu YC, et al. Predictive value of the international association for the study of lung cancer/American Thoracic Society/European Respiratory Society classification of lung adenocarcinoma in tumor recurrence and patient survival. *J Clin Oncol*. (2014) 32:2357–64. doi: 10.1200/JCO.2013.50.1049
25. Tsao M-S, Marguet S, Teuff GL, Lantuejoul S, Shepherd FA, Seymour L, et al. Subtype classification of lung adenocarcinoma predicts benefit from adjuvant chemotherapy in patients undergoing complete resection. *J Clin Oncol*. (2015) 33:3439–46. doi: 10.1200/JCO.2014.58.8335

26. Van Rijn JC, Reitsma JB, Stoker J, Bossuyt PM, Van Deventer SJ, Dekker E. Polyp miss rate determined by tandem colonoscopy: a systematic review. *Am J Gastroenterol*. (2006) 101:343–50. doi: 10.1111/j.1572-0241.2006.00390.x
27. Ahmad OF, Soares AS, Mazomenos E, Brandao P, Vega R, Seward E, et al. Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. *Lancet Gastroenterol Hepatol*. (2019) 4:71–80. doi: 10.1016/S2468-1253(18)30282-6
28. Mori Y, Kudo S-E, Berzin TM, Misawa M, Takeda K. Computer-aided diagnosis for colonoscopy. *Endoscopy*. (2017) 49:813–9. doi: 10.1055/s-0043-109430
29. Byrne MF, Shahidi N, Rex DK. Will computer-aided detection and diagnosis revolutionize colonoscopy? *Gastroenterology. (*2017) 153:1460–4.e1461. doi: 10.1053/j.gastro.2017.10.026
30. Chen P-HC, Gadepalli K, Macdonald R, Liu Y, Kadowaki S, Nagpal K, et al. An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. *Nat Med*. (2019) 25:1453–7. doi: 10.1038/s41591-019-0539-7
31. Liu Y, Kohlberger T, Norouzi M, Dahl GE, Smith JL, Mohtashamian A, et al. Artificial intelligence-based breast cancer nodal metastasis detection: insights into the black box for pathologists. *Arch Pathol Lab Med*. (2019) 143:859–68. doi: 10.5858/arpa.2018-0147-OA
32. Dalmis MU, Gubern-Merida A, Vreemann S, Bult P, Karssemeijer N, Mann R, et al. Artificial intelligence-based classification of breast lesions imaged with a multiparametric breast MRI protocol with ultrafast DCE-MRI, T2, and DWI. *Invest Radiol*. (2019) 54:325–32. doi: 10.1097/RLI.0000000000000544
33. El Hajjar, A., and Rey, J. F. (2020). Artificial intelligence in gastrointestinal endoscopy: general overview. *Chin Med J (Engl)*. 133, 326–34. doi: 10.1097/CM9.0000000000000623
34. Zarin DA, Ide NC, Tse T, Harlan WR, West JC, Lindberg D, et al. Issues in the registration of clinical trials. *JAMA*. (2007) 297:2112–20. doi: 10.1001/jama.297.19.2112

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 15 September 2020
doi: 10.3389/fonc.2020.01151

Comparison of Radiomics-Based Machine-Learning Classifiers in Diagnosis of Glioblastoma From Primary Central Nervous System Lymphoma

Chaoyue Chen[1,2†], *Aiping Zheng*[3†], *Xuejin Ou*[1,3], *Jian Wang*[4] *and Xuelei Ma*[1,5*]

[1] State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, Collaborative Innovation Center for Biotherapy, Chengdu, China, [2] Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu, China, [3] West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China, [4] School of Computer Science, Nanjing University of Science and Technology, Nanjing, China, [5] Department of Biotherapy, Cancer Center, West China Hospital, Sichuan University, Chengdu, China

OPEN ACCESS

Edited by:
Tsair-Fwu Lee,
National Kaohsiung University of Science and Technology, Taiwan

Reviewed by:
Remco Molenaar,
Amsterdam University Medical Center, Netherlands
Yu-Jie Huang,
Kaohsiung Chang Gung Memorial Hospital, Taiwan

***Correspondence:**
Xuelei Ma
drmaxuelei@gmail.com

[†]These authors share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 17 July 2019
Accepted: 08 June 2020
Published: 15 September 2020

Citation:
Chen C, Zheng A, Ou X, Wang J and Ma X (2020) Comparison of Radiomics-Based Machine-Learning Classifiers in Diagnosis of Glioblastoma From Primary Central Nervous System Lymphoma. Front. Oncol. 10:1151. doi: 10.3389/fonc.2020.01151

Purpose: The purpose of the current study was to evaluate the ability of magnetic resonance (MR) radiomics-based machine-learning algorithms in differentiating glioblastoma (GBM) from primary central nervous system lymphoma (PCNSL).

Method: One-hundred and thirty-eight patients were enrolled in this study. Radiomics features were extracted from contrast-enhanced MR images, and the machine-learning models were established using five selection methods (distance correlation, random forest, least absolute shrinkage and selection operator (LASSO), eXtreme gradient boosting (Xgboost), and Gradient Boosting Decision Tree) and three radiomics-based machine-learning classifiers [linear discriminant analysis (LDA), support vector machine (SVM), and logistic regression (LR)]. Sensitivity, specificity, accuracy, and areas under curves (AUC) of models were calculated, with which the performances of classifiers were evaluated and compared with each other.

Result: Brilliant discriminative performance would be observed among all classifiers when combined with the suitable selection method. For LDA-based models, the optimal one was Distance Correlation + LDA with AUC of 0.978. For SVM-based models, Distance Correlation + SVM was the one with highest AUC of 0.959, while for LR-based models, the highest AUC was 0.966 established with LASSO + LR.

Conclusion: Radiomics-based machine-learning algorithms potentially have promising performances in differentiating GBM from PCNSL.

Keywords: glioblastoma, primary central nervous system lymphoma, magnetic resonance imaging, radiomics, machine learning

INTRODUCTION

Glioblastoma (GBM) and primary central nervous system lymphoma (PCNSL) are considered as the common primary brain tumors, which share similar radiological characteristics but diverse in therapeutic strategies (1–3). The standard of treatment for a GBM is total resection, followed by daily radiation and chemotherapy (like temozolomide) for 6.5 weeks, then a 6-month regimen

of oral chemotherapy given 5 days a month, while the first-line treatment for PCNSL is systemic chemotherapy (like high-dose methotrexate regimen) (4). In most cases, the morphological description of two types of tumors on MRI is characteristic enough for adequate discrimination (5, 6). However, misdiagnosis could still incur in some cases because the images of atypical GBM and atypical PCNSL could mimic each other (7). Advanced MRI technology could be useful in the differentiation. However, the urgency of novel radiological methods focused on conventional MR sequences has still been highlighted given that the advanced MRI cannot be performed as the routine examination for every patient.

Texture analysis (TA) refers to a number of a set of mathematical methods describing the features of images, with which non-visual information could be represented with analyzable pixel intensities and the spatial distributions (8, 9). It has been applied as the radiological imaging biomarkers to evaluate tumor heterogeneity, and showed promising ability in as tumor diagnosis, presurgical grading, as well as gene mutation prediction (10–12). Moreover, with quantified analyses of images, it has also been incorporated with various novel computer technologies, such as machine learning (13–16).

The purpose of the present study is to discriminate GBM from PCNSL with radiomics-based machine-learning algorithms in contrast-enhanced T1-weighted (T1C) imaging. In addition, we evaluated different combinations of selection methods and classifiers, trying to make comparison of models' performances.

METHOD

Patient Selection

The patients were selected from neurosurgery department by reviewing the electronic medical records between 2015 and 2018. The including criteria of patients were as follows: (1) pathologically confirmed on GBM or PCNSL; (2) undertook MR scan before any tumor biopsy or surgery; (3) newly diagnosed GBM or PCNSL. Some patients were excluded because of the history of intracranial surgery or irrelevant intracranial diseases. In total, 138 patients (72 men, median age 48 years; and 66 women, median age 54 years) were enrolled from the institution database, including 76 patients diagnosed with GBM and 62 diagnosed with PCNSL.

The MR images were collected from the PACS system in the radiological department. We focused on conventional MR sequences, including T1-weighted image (T1WI), contrast-enhanced T1-weighted (T1C) imaging, T2-weighted image (T2WI), and fluid-attenuated inversion recovery, considering that the advanced MR sequences were not commonly used in our institution. After the initial evaluation of images, T1C was selected as the study sequences with rather clear description of the boundary between the tumor tissues and normal brain tissue (**Figure 1**).

MRI Protocol

The preoperative MR scan was conducted with 3-T GE MRI system with an eight-channel phase-array head coil. The protocols of the contrast-enhanced T1-weighted imaging were time repetition = 2,000 ms, field of view = 240 × 240 mm^2, time echo = 30 ms, 30 axial slices, slice thickness = 5 mm (no slice gap), flip angle = 90°, and 200 volumes in each run. Gadopentetate dimeglumine (0.1 mmol/kg) were taken as the contrast agent. The multi-directional data of contrast-enhanced MRI were collected with the continuous interval time of 90–250 s.

All procedures involving human participants were in accordance with the ethical standards of the institutional and/or national research committee. The Ethics Committee of Sichuan University approved this retrospective study. Written informed consent was necessary before radiological examination (written informed consent for patients <16 years old was signed by parents or guardians) for all patients. They agreed to undertake the examination if needed and were informed that the statistics (including MR image) might be used for academic purposes in the future.

Texture Feature Extraction

Two neurosurgeons participated in the extraction of texture features by using lifeX software (http://www.lifexsoft.org) under the supervisions of senior radiologists. By manually drawing along the tumor tissue slice by slice, the software automatically retrieved 3D-based texture features from two sets of orders with default settings (17). In the first order, statistics from shape- and histogram-based matrix were retrieved. In the second order, statistics from gray-level co-occurrence matrix (GLCM), gray-level zone length matrix (GLZLM), neighborhood gray-level dependence matrix (NGLDM), and gray-level run length matrix (GLRLM) were retrieved. The images were excluded of which the volume of interest did not reach 64 voxels to avoid the interference of the lower image matrix resolution.

Mann–Whitney *U*-test was employed to explore if there is significant statistical difference between the data extracted by two researchers. The results suggested that none of the features were significantly different, implying that the results could be considered reliable and reproducible (shown in **Supplementary Material 1**).

Classification Algorithm Application

The patients were randomly divided into the training group and the validation group on the proportion of 4:1. For machine-learning classifiers, the optimal texture features were selected first for classifiers to reduce the number of input variables to improve the performance of the model and to both reduce the computational cost. Considering the optimal selection method was controversial for different classifiers, five methods were conducted separately, including distance correlation, random forest (RF), least absolute shrinkage and selection operator (LASSO), eXtreme gradient boosting (Xgboost), and Gradient Boosting Decision Tree (GBDT).

The purpose of machine learning was to establish and train the models to discriminate GBM from PCNSL with radiomics features extracted from T1C imaging. Three classifiers were tested, including linear discriminant analysis (LDA), support vector machine (SVM), and logistic regression (LR). Thus, 15 diagnostic models were evaluated with different combinations of selection methods and classifiers. The models were trained with

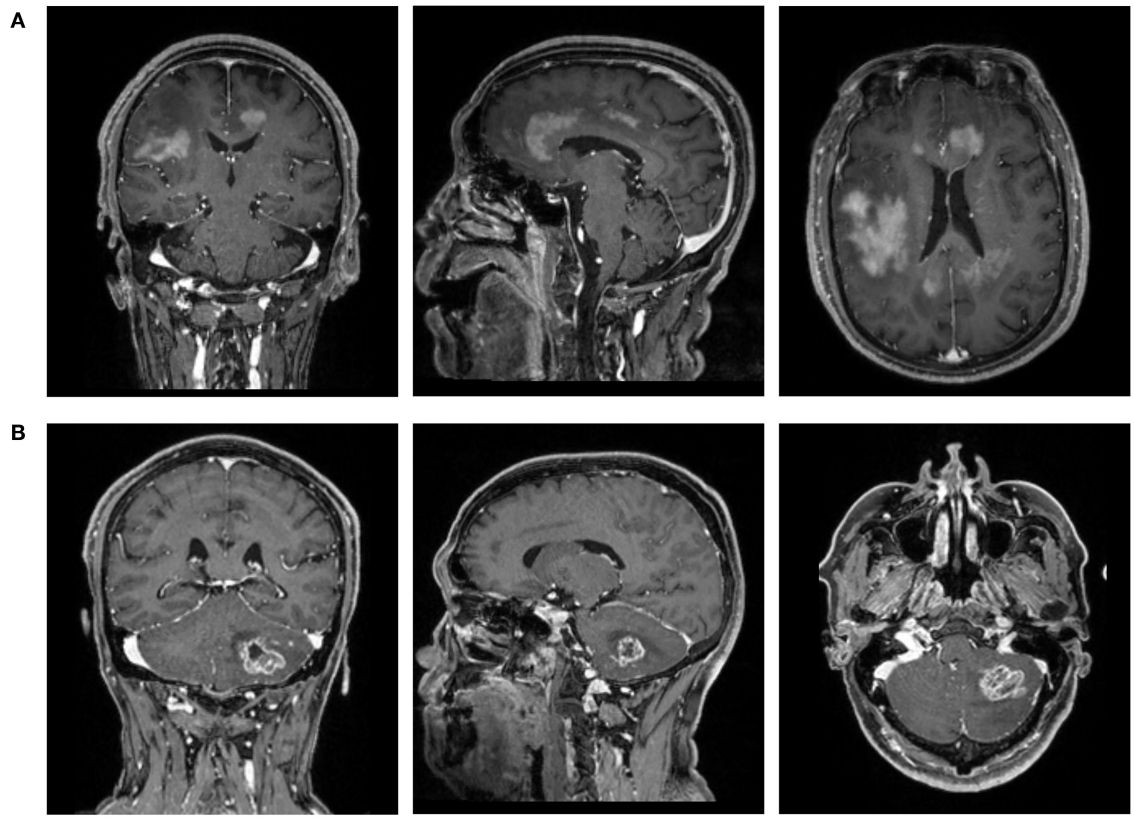


FIGURE 1 | The magnetic resonance images (T1C) of patients with **(A)** primary central nervous system lymphoma (PCNSL) or **(B)** glioblastoma (GBM).

the statistics of the training group and tested in the validation group. Sensitivity, specificity, area under the receiver operating characteristic curve (AUC), and accuracy of each model were recorded for evaluation. On application of each model, the cycle of training-validation was performed 100 times to obtain the realistic distribution of classification accuracies. The flow chart of the study is represented in **Figure 2**.

The models were programmed using Python Programming Language in this study. The models were directly established with default hyperparameter settings of scikit-learn packages (https://scikit-learn.org/stable/).

RESULT

The selected features with different methods are represented in **Table 1**. Four features, GLRLM_LGRE, GLRLM_HGRE, GLRLM_SRHGE, and GLZLM_HGZE, were almost selected even using different methods, suggesting that they were the most significant features in discrimination compared with the others. The other selected features should be reasonably considered as relevant in discrimination, but was hard to tell how much they influenced the algorithms' performances.

The performances of models are listed in **Table 2**. As mentioned previously, the models were established with different combinations of selection methods and classifiers. The results indicated that all three classifiers represented impressive differential ability when using suitable selected features, and the LDA classifier showed much better compatibility compared with other classifiers. Over-fitting was observed in six models, including RF + SVM, Xgboost + SVM, GBDT + SVM, and RF + LR, Xgboost + LR, and GBDT + LR. For LDA-based models, the AUCs in the validation group were 0.978, 0.964, 0.977, 0.750, and 0.956; for the SVM-based models, the AUCs were 0.959 and 0.822; and for LR-based models, the AUCs were 0.933 and 0.975.

In the current study, the optimal model was Distance Correlation + LDA. In the training group, the predictive model showed the discriminative ability with AUC of 0.992, accuracy of 0.993, sensitivity of 0.996, and specificity of 0.990. In the validation group, the performance of the model was rather good, with AUC of 0.978, accuracy of 0.979, sensitivity of 0.982, and specificity of 0.976. The association between discriminative functions from models is represented in **Figure 3**. **Figure 4** represents the examples of distribution of the direct LDA function diagnosis of GBM and PCNSL for one cycle.

DISCUSSION

In the current study, we performed research in differentiating GBM from PCNSL with the radiomics-based machine-learning technology. Radiomics parameters were extracted from T1C

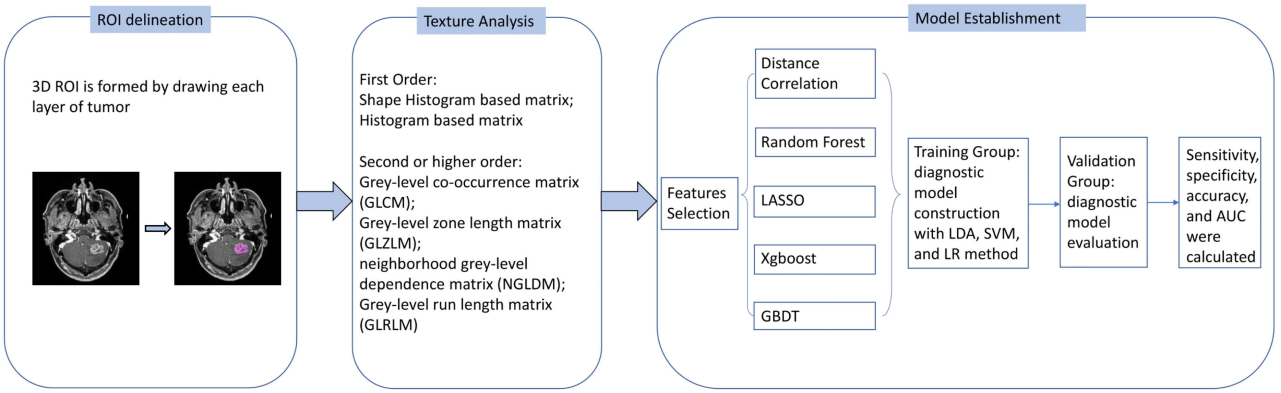


FIGURE 2 | Flow chart of image processing and machine learning.

TABLE 1 | The features selected with different methods.

Selection method	Selected features
Distance correlation	GLRLM_LGRE; GLRLM_HGRE; GLRLM_SRLGE; GLRLM_SRHGE; GLRLM_LRLGE; GLZLM_LGZE; GLZLM_HGZE; GLZLM_SZLGE
RF	GLRLM_LGRE; GLRLM_HGRE; GLRLM_SRLGE; GLRLM_SRHGE; GLRLM_LRHGE; GLZLM_HGZE
LASSO	minValue; meanValue; stdValue; SHAPE_Volume; GLCM_Contrast; GLRLM_HGRE; GLRLM_SRHGE; GLRLM_LRHGE; GLRLM_GLNU; GLRLM_RLNU; GLZLM_LZE; GLZLM_HGZE; GLZLM_SZHGE; GLZLM_LZHGE; GLZLM_GLNU; GLZLM_ZLNU
XgBoost	GLRLM_LGRE
GBDT	GLRLM_LGRE; GLRLM_HGRE; GLRLM_SRLGE; GLRLM_SRHGE; GLRLM_LRHGE; GLZLM_LGZE; GLZLM_HGZE; GLZLM_SZLGE; GLZLM_SZHGE

RF, random forest; LASSO, least absolute shrinkage and selection operator; Xgboost, eXtreme gradient boosting; GBDT, Gradient Boosting Decision Tree.

TABLE 2 | Results of the discriminative model in distinguishing GBM from PCNSL in the training and validation group.

Classifier	Selection method	Training group				Validation group			
		AUC	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity
LDA	Distance correlation	0.992	0.993	0.996	0.990	0.978	0.979	0.982	0.976
	RF	0.970	0.968	0.935	0.990	0.964	0.957	0.906	0.990
	LASSO	0.997	0.996	0.992	0.995	0.977	0.971	0.955	0.989
	Xgboost	0.791	0.810	0.995	0.740	0.750	0.789	0.995	0.735
	GBDT	0.972	0.970	0.939	0.996	0.956	0.950	0.892	0.995
SVM	Distance correlation	0.957	0.962	0.998	0.934	0.959	0.964	0.997	0.943
	RF (over-fitting)	1	1	1	1	0.5	0.585	1	0.943
	LASSO	0.843	0.835	0.747	0.966	0.822	0.789	0.671	0.965
	Xgboost (over-fitting)	0.5	0.541	0.747	0.967	0.5	0.586	0.671	0.965
	GBDT (over-fitting)	1	1	1	1	0.5	0.586	0.670	0.965
LR	Distance correlation	0.977	0.956	0.961	0.949	0.933	0.927	0.941	0.911
	RF (over-fitting)	1	0.547	1	0.592	0.511	0.515	0.551	0.596
	LASSO	0.959	0.988	0..942	0.981	0.975	0.966	0.975	0.964
	Xgboost (over-fitting)	0.959	0.988	0.942	0.981	0.5	0.5	0.542	0.586
	GBDT (over-fitting)	0.951	0.562	0.954	0.592	0.538	0.515	0.577	0.596

AUC, area under curve; RF, random forest; LASSO, least absolute shrinkage and selection operator; Xgboost, eXtreme gradient boosting; GBDT, Gradient Boosting Decision Tree; LDA, linear discriminant analysis; SVM, support vector machine; LR, logistic regression.

images to detect non-visual information of two types of tumors. The models were established with five selection methods and three classifiers and tested to find the optimal model. The result showed that the radiomics-based machine-learning classifier represented excellent performance in all classifiers with AUC more than 0.900. The optimal model was the combination of Distance Correlation + LDA with AUC of 0.978, accuracy of 0.979, sensitivity of 0.982, and specificity of 0.976. Given that the T1C image was routine examination for GBM and PCNSL, our results suggested that radiomics was a feasible solution for clinical application without requiring additional fees or platform.

Generally, contrast-enhanced T1imaging is a routine radiological examination for patients with GBM or PCSNL. A previous study indicated that at the time of initial presentation for many cases, routine morphological MRI is capable enough in differentiating between GBM and PCNSL lesions. The image patterns are correlated with the tumor characteristics, such as intratumoral hemorrhage, angiogenesis, and necrotic or cystic components. Specifically, heterogeneous enhancement was present in 98.1% of GBM cases and homogenous enhancement in 64.8% of PCNSL cases; necrosis was observed in 88.9% of GBM lesions and 5.6% of PCNSL lesions; multiple lesions were shown in 51.9% of PCNSL cases and 35.2% of GBM cases. Signs of bleeding were uncommon in PCNSL (5.6%) and frequent in GBM (44.4%) (18). Advanced imaging techniques, such as apparent diffusion coefficient (ADC), diffusion-tensor imaging (DTI), dynamic susceptibility-weighted contrast-enhanced MRI, and perfusion weighted imaging, were also additionally performed in discriminating GBM and PCNSL if necessary (19–21). Surgeons could obtain the information on characteristics of tumors to make diagnostic and treatment decisions. However, even with these researches, the differential diagnosis between GBM and PCNSL was still a challenge in some cases, especially given that the conventional MR sequence could only make limited discrimination between two types of tumors and that advanced imaging techniques were not available for all patients.

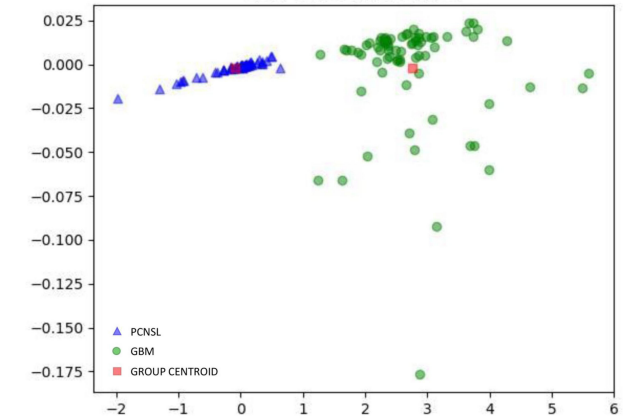


FIGURE 3 | Relationship between the discriminant functions for discriminating GBM from PCNSL.

Comparing with GBM, permeable neovascularization and higher degree of cellularity were more likely to be observed in PCNSL, which theoretically provide the mechanism of TA-based image discrimination (22–24). In our study, radiomics of T1C imaging were used to detect the microscopic differences between GBM and PCNSL, and the results suggested TA was the feasible solution in discriminating GBM and PCNSL radiologically. Radiomics has been reported to distinguish GBM from PCNSL in a previous study, and machine-learning classification model was reported to improve the performance in discrimination (6, 25). Researchers made comparison on diagnostic accuracy between radiologists and machine-learning classifiers, and they suggested that classifiers yielded better diagnostic performance than human radiologists (25). However, the sample sizes of these studies were not large enough and only a few models were tested. Our study enrolled 138 patients with rational proportion of each group and made an evaluation on 15 combinations. In a previous study, RF-based classifier represented perfect performance in discriminating atypical glioblastoma from PCNSL with AUC of 0.98 (6), and SVM-based classifier also represented non-inferior

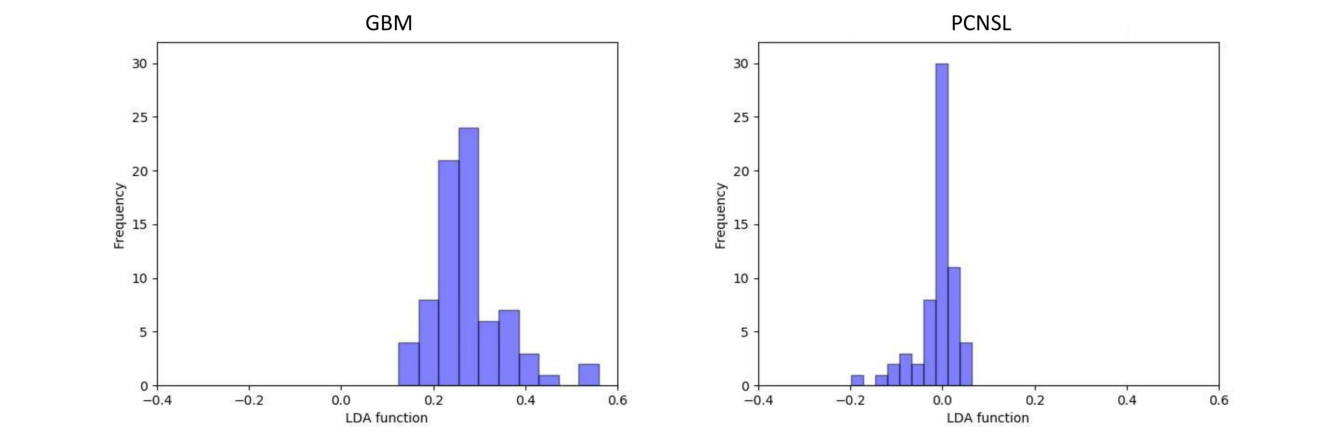


FIGURE 4 | Example of distributions of the LDA function determined for the lesions for one cycle.

performance to expert human with AUC of 0.877 (25). In our study, the results showed that all three classifiers represented perfect performance when combined with a suitable selection method. It is worth noting that the result of the optimal SVM-based model in our study was with AUC of 0.96, demonstrating much better diagnostic performance than the previous study.

The possible explanation for the improvement was the performance improvement in selection method. Radiomics analysis involved large amounts of features, but machine learning required the most suitable parameters. Previous researchers selected parameters with F-statistic approach into SVM classifier, while we selected with distance correlation, RF, LASSO, Xgboost, or GBDT approach. The combination of LASSO + SVM represented similar discriminative performance such as in the previous study with AUC of 0.822. Besides performances, we can also find that the selection methods were also important to the model stability. Over-fitting is a problem that should be avoided in designing the machine-learning models, which happens when the models catch inaccurate values in the data and the noisy data. Our results suggested that over-fitting probably occurred when using RF, Xgboost, and GBDT as selection methods. Perhaps the features selected with these methods contained too much noise and led to the over-fitting of models.

As for the classifier selection, the purpose of enrollment of three classifiers was to choose the suitable one in discriminating GBM from PCNSL. The results suggested that with suitable features, all of them could represent discriminative ability. It is worthy to note that although we chose Distance Correlation + LDA as the optimal model, some models (like LASSO + LDA and LASSO + LR) also represented pretty similar discriminative performances. The model Distance Correlation + LDA was chosen as the optimal one because it has the minimal difference between sensitivity and specificity compared with LASSO + LDA and LASSO + LR. However, given that all classifier/feature selection methods investigated seem to perform quite comparably and variance in AUC may be partially attributed to small statistical group, the additional gain in information by comparing machine-learning models was quite limited and carefully interpreted. Future investigations with larger sample sizes are required to address this problem and verify our results.

There were several limitations to our study. First, the isolated evaluation of T1C image is not representative of the real clinical work given other sequences (such as ADC, perfusion, DTI, and T2 gradient-echo) could also be useful. Second, the diagnostic performance of radiomics-based machine learning was not compared with other advanced MRI technology. Third, the study cohort is not large enough, requiring study with a large population to verify our results. Forth, the machine-learning classifier was not validated in the other dataset. Considering the considerable variability in images acquired with various MR scanner at different institutions, we cannot guarantee the diagnostic ability of our machine-learning classifier for external datasets. However, the image processing and analysis protocol were open-source packages, meaning they should be validated and reproduced with other datasets.

CONCLUSION

Radiomics with machine-learning algorithm technology represented promising ability in differentiating GBM from PCNSL.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Sichuan University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

XM participated in conceptualization and revised some intellectual content in the manuscript. CC collected MR image, participated in MRI features extraction, and drafted this manuscript. XO collected MR image and participated in MRI features extraction. JW deployed the machine-learning algorism and responsible for statistical analysis. AZ participated in the most revision work. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.01151/full#supplementary-material

REFERENCES

1. Dolecek TA, Propp JM, Stroup NE, Kruchko C. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005-2009. *Neuro Oncol.* (2012) 14(Suppl. 5):v1–49. doi: 10.1093/neuonc/nos218
2. Schlegel U. Primary CNS lymphoma. *Ther Adv Neurol Disord.* (2009) 2:93–104. doi: 10.1177/1756285608101222
3. Kickingereder P, Wiestler B, Sahm F, Heiland S, Roethke M, Schlemmer HP, et al. Primary central nervous system lymphoma and atypical glioblastoma: multiparametric differentiation by using diffusion-, perfusion-, and susceptibility-weighted MR imaging. *Radiology*. (2014) 272:843–50. doi: 10.1148/radiol.14132740
4. von Baumgarten L, Illerhaus G, Korfel A, Schlegel U, Deckert M, Dreyling M. The diagnosis and treatment of primary CNS lymphoma. *Dtsch Arztebl Int*. (2018) 115:419–26. doi: 10.3238/arztebl.2018.0419

5. Koeller KK, Smirniotopoulos JG, Jones RV. Primary central nervous system lymphoma: radiologic-pathologic correlation. *Radiographics*. (1997) 17:1497–526. doi: 10.1148/radiographics.17.6.9397461
6. Suh HB, Choi YS, Bae S, Ahn SS, Chang JH, Kang SG, et al. Primary central nervous system lymphoma and atypical glioblastoma: differentiation using radiomics approach. *Eur Radiol.* (2018) 28:3832–9. doi: 10.1007/s00330-018-5368-4
7. Al-Okaili RN, Krejza J, Woo JH, Wolf RL, O'Rourke DM, Judy KD, et al. Intraaxial brain masses: MR imaging-based diagnostic strategy–initial experience. *Radiology*. (2007) 243:539–50. doi: 10.1148/radiol.2432060493
8. Scalco E, Rizzo G. Texture analysis of medical images for radiotherapy applications. *Br J Radiol.* (2017) 90:20160642. doi: 10.1259/bjr.20160642
9. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clin Radiol.* (2004) 59:1061–9. doi: 10.1016/j.crad.2004.07.008
10. Li Y, Liu X, Qian Z, Sun Z, Xu K, Wang K, et al. Genotype prediction of ATRX mutation in lower-grade gliomas using an MRI radiomics signature. *Eur Radiol.* (2018) 28:2960–8. doi: 10.1007/s00330-017-5267-0
11. Ditmer A, Zhang B, Shujaat T, Pavlina A, Luibrand N, Gaskill-Shipley M, et al. Diagnostic accuracy of MRI texture analysis for grading gliomas. *J Neuro Oncol.* (2018) 140:583–9. doi: 10.1007/s11060-018-2984-4
12. Zhou H, Vallieres M, Bai HX, Su C, Tang H, Oldridge D, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro Oncol.* (2017) 19:862–70. doi: 10.1093/neuonc/now256
13. Lee JH, Ha EJ, Kim JH. Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT. *Eur Radiol.* (2019) 29:5452–7. doi: 10.1007/s00330-019-06098-8
14. Xie J, Liu R, Luttrell J, Zhang C. Deep learning based analysis of histopathological images of breast cancer. *Front Genet.* (2019) 10:80. doi: 10.3389/fgene.2019.00080
15. van IJzendoorn DGP, Szuhai K, Briaire-de Bruijn IH, Kostine M, Kuijjer ML, Bovee JVMG. Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. *PLoS Comput Biol.* (2019) 15:e1006826. doi: 10.1371/journal.pcbi.1006826
16. Ko SY, Lee JH, Yoon JH, Na H, Hong E, Han K, et al. Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. *Head Neck*. (2019) 41:885–91. doi: 10.1002/hed.25415
17. Nioche C, Orlhac F, Boughdad S, Reuze S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* (2018) 78:4786–9. doi: 10.1158/0008-5472.CAN-18-0125
18. Malikova H, Koubska E, Weichet J, Klener J, Rulseh A, Liscak R, et al. Can morphological MRI differentiate between primary central nervous system lymphoma and glioblastoma? *Cancer Imaging*. (2016) 16:40. doi: 10.1186/s40644-016-0098-9
19. Toh CH, Wei KC, Chang CN, Ng SH, Wong HF. Differentiation of primary central nervous system lymphomas and glioblastomas: comparisons of diagnostic performance of dynamic susceptibility contrast-enhanced perfusion MR imaging without and with contrast-leakage correction. *AJNR*. (2013) 34:1145–9. doi: 10.3174/ajnr.A3383
20. Radbruch A, Wiestler B, Kramp L, Lutz K, Baumer P, Weiler M, et al. Differentiation of glioblastoma and primary CNS lymphomas using susceptibility weighted imaging. *Eur J Radiol.* (2013) 82:552–6. doi: 10.1016/j.ejrad.2012.11.002
21. Choi YS, Lee HJ, Ahn SS, Chang JH, Kang SG, Kim EH, et al. Primary central nervous system lymphoma and atypical glioblastoma: differentiation using the initial area under the curve derived from dynamic contrast-enhanced MR and the apparent diffusion coefficient. *Eur Radiol.* (2017) 27:1344–51. doi: 10.1007/s00330-016-4484-2
22. Kickingereder P, Sahm F, Wiestler B, Roethke M, Heiland S, Schlemmer HP, et al. Evaluation of microvascular permeability with dynamic contrast-enhanced MRI for the differentiation of primary CNS lymphoma and glioblastoma: radiologic-pathologic correlation. *AJNR*. (2014) 35:1503–8. doi: 10.3174/ajnr.A3915
23. Guo AC, Cummings TJ, Dash RC, Provenzale JM. Lymphomas and high-grade astrocytomas: comparison of water diffusibility and histologic characteristics. *Radiology*. (2002) 224:177–83. doi: 10.1148/radiol.2241010637
24. Toh CH, Castillo M, Wong AM, Wei KC, Wong HF, Ng SH, et al. Primary cerebral lymphoma and glioblastoma multiforme: differences in diffusion characteristics evaluated with diffusion tensor imaging. *AJNR*. (2008) 29:471–5. doi: 10.3174/ajnr.A0872
25. Alcaide-Leon P, Dufort P, Geraldo AF, Alshafai L, Maralani PJ, Spears J, et al. Differentiation of enhancing glioma and primary central nervous system lymphoma by texture-based machine learning. *AJNR*. (2017) 38:1145–50. doi: 10.3174/ajnr.A5173

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership