



Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology

Weimiao Wu^{1,2*†}, Chintan Parmar^{1,3,4†}, Patrick Grossmann^{1,3,5}, John Quackenbush^{2,5}, Philippe Lambin⁴, Johan Bussink⁶, Raymond Mak¹ and Hugo J. W. L. Aerts^{1,3,5*}

¹ Department of Radiation Oncology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, ² Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA, ³ Department of Radiology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, ⁴ Research Institute GROW, Maastricht University, Maastricht, Netherlands, ⁵ Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA, ⁶ Department of Radiation Oncology, Radboud University Medical Center, Nijmegen, Netherlands

OPEN ACCESS

Edited by:

Issam El Naqa,
McGill University, Canada

Reviewed by:

Joshua Silverman,
New York University Medical Center,
USA

Derek Merck,
Rhode Island Hospital, USA

*Correspondence:

Weimiao Wu
weimiao.wu@mail.harvard.edu;
Hugo J. W. L. Aerts
hugo@jimmy.harvard.edu

[†]Weimiao Wu and Chintan Parmar
contributed equally.

Specialty section:

This article was submitted to
Radiation Oncology, a section of the
journal *Frontiers in Oncology*

Received: 10 November 2015

Accepted: 14 March 2016

Published: 30 March 2016

Citation:

Wu W, Parmar C, Grossmann P,
Quackenbush J, Lambin P, Bussink J,
Mak R and Aerts HJWL (2016)
Exploratory Study to Identify
Radiomics Classifiers for Lung Cancer
Histology.
Front. Oncol. 6:71.
doi: 10.3389/fonc.2016.00071

Background: Radiomics can quantify tumor phenotypic characteristics non-invasively by applying feature algorithms to medical imaging data. In this study of lung cancer patients, we investigated the association between radiomic features and the tumor histologic subtypes (adenocarcinoma and squamous cell carcinoma). Furthermore, in order to predict histologic subtypes, we employed machine-learning methods and independently evaluated their prediction performance.

Methods: Two independent radiomic cohorts with a combined size of 350 patients were included in our analysis. A total of 440 radiomic features were extracted from the segmented tumor volumes of pretreatment CT images. These radiomic features quantify tumor phenotypic characteristics on medical images using tumor shape and size, intensity statistics, and texture. Univariate analysis was performed to assess each feature's association with the histological subtypes. In our multivariate analysis, we investigated 24 feature selection methods and 3 classification methods for histology prediction. Multivariate models were trained on the training cohort and their performance was evaluated on the independent validation cohort using the area under ROC curve (AUC). Histology was determined from surgical specimen.

Results: In our univariate analysis, we observed that fifty-three radiomic features were significantly associated with tumor histology. In multivariate analysis, feature selection methods ReliefF and its variants showed higher prediction accuracy as compared to other methods. We found that Naive Bayes classifier outperforms other classifiers and achieved the highest AUC (0.72; p -value = 2.3×10^{-7}) with five features: Stats_min, Wavelet_HLL_rigl_lowGrayLevelRunEmphasis, Wavelet_HHL_stats_median, Wavelet_HLL_stats_skewness, and Wavelet_HLH_glcM_clusShade.

Conclusion: Histological subtypes can influence the choice of a treatment/therapy for lung cancer patients. We observed that radiomic features show significant association with the lung tumor histology. Moreover, radiomics-based multivariate classifiers were independently validated for the prediction of histological subtypes. Despite achieving lower than optimal prediction accuracy (AUC 0.72), our analysis highlights the impressive potential of non-invasive and cost-effective radiomics for precision medicine. Further research in this direction could lead us to optimal performance and therefore to clinical applicability, which could enhance the efficiency and efficacy of cancer care.

Keywords: quantitative imaging, radiomics, lung cancer histology, computational science, feature selection

INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths worldwide with 150,000 deaths per year in US (1). Lung cancer is clinically divided into two groups: small cell lung cancer (SCLC, ~25%) and non-small cell lung cancer (NSCLC, ~75%) (2). The most common histological subtypes of NSCLC are adenocarcinoma (~38%) and squamous cell carcinoma (~20%) (3). These two subtypes have distinct histologic features, i.e., squamous cell carcinoma is associated with intercellular bridging and individual cell keratinization, whereas glandular architecture is prominent for adenocarcinoma (3, 4). Histological classification of lung cancer provides important information about tissue characteristics and anatomical location. Adenocarcinoma often develops at the periphery of the lungs, while squamous carcinoma is generally located more centrally (5). Differences have also been found in expression of glycolysis and hypoxia-related markers between histological subtypes, which suggests histology-specific glucose metabolism in NSCLC (5, 6). In addition, histological tumor classification could determine the optimal treatment and/or therapy strategies (7). Recent advancement in the therapy for lung cancer is characterized by discovery of targetable mutations and histology-based therapeutic regimen selection (7, 8). For example, Pemetrexed is the preferred treatment for stage IV lung adenocarcinoma, whereas bevacizumab is not recommended for squamous carcinoma due to the risk of pulmonary hemorrhage observed in phase II trials (8–10). It has also been shown that treatment for stage III lung cancer patients with squamous carcinoma has significant improvement in survival with cisplatin/gemcitabine versus cisplatin/pemetrexed, but for adenocarcinoma patients, the latter treatment provides superior survival rate (11, 12). More importantly, histology classification increases the likelihood of identifying patients with targetable mutations like EGFR mutations, which occurs primarily in adenocarcinoma (4).

In routine clinical practice, the most common way of classifying tumor histology is through the histopathological analysis of tumor tissues via biopsy. This is clinically limited by the inherent risk of invasive procedures as well as poor time and cost efficiency (8). Therefore, automatic, non-invasive, and cost-effective alternatives are desired. Medical imaging provides promising opportunities in this regard. It assesses the tumor tissue characteristics non-invasively. Furthermore, it is relatively cost-effective and is already a routinely used clinical practice for oncologic diagnosis, staging, and treatment guidance (13–15).

Radiomics, a high throughput data mining approach, can exploit the non-invasive medical image data (14). It focuses on extracting a large number of quantitative imaging features, which can provide a detailed and comprehensive characterization of the tumor phenotype (8). Several studies have shown the prognostic/predictive power of radiomic features in different cancer types by using different medical imaging modalities like CT (16, 17), MRI (18), PET (15, 19, 20), and US (21). It has been shown that radiomic features are associated with several clinically relevant factors, such as tumor stage (22), tumor metabolism (23), overall patient survival (17, 24), metastasis (13), treatment response (25), and the underlying gene expression profiles (26, 27). These associations can be leveraged to build efficient and effective prediction/prognostic models. Therefore, radiomics is a promising field providing a non-invasive and cost-effective way for personalized medicine.

A limited number of studies have investigated the association of radiomic features and NSCLC tumor histology (22, 28). Most of them used a clustering-based unsupervised approach for associating radiomic features with tumor histological subtypes. However, in order to achieve higher prediction accuracies, supervised methods are generally preferred over unsupervised approaches if labeled data is available. Furthermore, like any other high-throughput data mining approach, radiomics also falls prey to the curse of dimensionality, which means we would need to get an enormous amount samples due to high dimensional radiomic features (29). Feature/variable selection is one of the solutions to this problem (30). Many feature selection methods have been proposed in machine learning literature, which should be applied for radiomics-based predictive analyses (31).

In this study, we investigated 24 commonly used feature selection methods and 3 supervised machine-learning classification methods for the prediction of lung cancer histologic subtypes, using independent training and validation cohorts from two different institutions. We first reduced the radiomic feature space into a non-redundant subspace by using correlation-based feature elimination. Second, we applied 24 (Information Gain, Gain Ratio Gini Index, MDL, DKM, ReliefF, and their variants) univariate filter-based feature selection methods to the resultant non-redundant feature subset. We chose these filter-based methods because of their popularity in the literature and their high computational efficiency.

The main objective of this study was to investigate the association between radiomic features and lung tumor histology.

We employed machine-learning methods to build radiomics-based multivariate classifiers for the prediction of tumor histology. Non-invasive and cost-effective radiomic data could improve the histological classification and hence the treatment/therapy, which in general could have a large impact in cancer care. With improving image feature extraction techniques, a higher accuracy in classification is expected to achieve. This work will serve as a promising prognostic tool for informing treatment choice and fostering personalized therapy for lung cancer patients.

MATERIALS AND METHODS

Datasets

We used two NSCLC cohorts collected at different institutions in the Netherlands. The training dataset (Lung1) contains 198 patients (mean age 69.5, range 34–88 years) with pathologically confirmed adenocarcinoma ($n = 152$) or squamous carcinoma ($n = 51$), UICC stages I–IIIb, treated with radical radiotherapy or with chemoradiation at MAASTRO Clinic in Maastricht in the Netherlands. Classification of the tumors as either adenocarcinoma or squamous cell carcinoma was based on hemotoxylin and eosin (H&E) staining according to the World Health Organization (WHO) classification of malignant lung tumors. Experienced radiation oncologists using a standard clinical delineation protocol performed delineation based on fused PET-CT imaging.

The test dataset (Lung2) has 152 patients (mean age 65.6, range 41–86 years) with pathologically confirmed adenocarcinoma ($n = 62$) or squamous carcinoma ($n = 90$), stages (I–IVa), treated at Radboud University Medical Center (RUMC) in Nijmegen, the Netherlands, between February 2004 and October 2011. Histological classification was based on H&E staining according to the WHO classification of malignant lung tumors. Manual delineations of the CT-scans were available for all included patients in the two datasets. Further details regarding the two data sets can be found in a previous paper (32). The Institutional Review Board of the Maastricht University Medical Center (MUMC+) and the Institutional Review Board of the RUMC waved review due to the retrospective nature of this study (32).

Radiomic Features

Tumor phenotypic characteristics were quantified by extracting 440 3D radiomic features from the segmented tumor regions of pretreatment CT images (32). All radiomic features can be divided into three groups: (1) Intensity: these features quantify the density of the tumor region on the CT image from the first-order histogram of voxel intensities. (2) Shape: these features quantify the 3D geometric properties of the tumors. (3) Texture: textural features quantify the intratumor heterogeneity by using the gray level cooccurrence (GLCM) and gray level run length matrices (GLRLM). Intensity and textural features were also computed after applying 3D wavelet transformations (coiflet filters) to the original image. Matlab R2012b was used for the image analysis, and radiomic features were automatically extracted using Matlab R2012b. Details about the image analysis as well as the mathematical definition of the radiomic features can be found in previous literature (32). All radiomic data were centered and scaled before performing the analysis.

Univariate Analysis

The association between the radiomic features and histological subtypes was assessed using the area under the receiver operating characteristic curve (ROC curve) (33). We computed AUC for all the features in a univariate manner. Significance was estimated using a random permutation test with iteration of 1000. The analysis was performed using R package survcomp (34).

Feature Selection

Like any other high throughput approach, radiomics also has highly redundant feature space. So, if we just rank features based on feature relevance, it is likely that highly correlated features have similar rank and they end up together in the selected feature subset (35). Several studies have discussed this issue with respect to filter-based feature selection methods (36, 37). Besides, correlated features can cause the classifiers to be sensitive to small changes in the datasets. This could cause a severe problem in cohorts with a different structure of collinearity (38).

To address this problem, we used a simple two-stage feature selection method by combing correlation-based feature elimination and univariate feature selection. In the first stage, we eliminated highly correlated features using a correlation matrix. We calculated column-wise average absolute correlation $C = \frac{1}{n} \sum_j c_{ij}$ for each feature. A threshold M is set for elimination, that is, for each pair-wise correlation c_{ij} that exceed M , we removed the feature with higher column-wise average absolute correlation C . By eliminating those highly correlated features, we are left with “non-redundant” set of features. This process was implemented by R package “corrplot” with “findCorrelation” function (39).

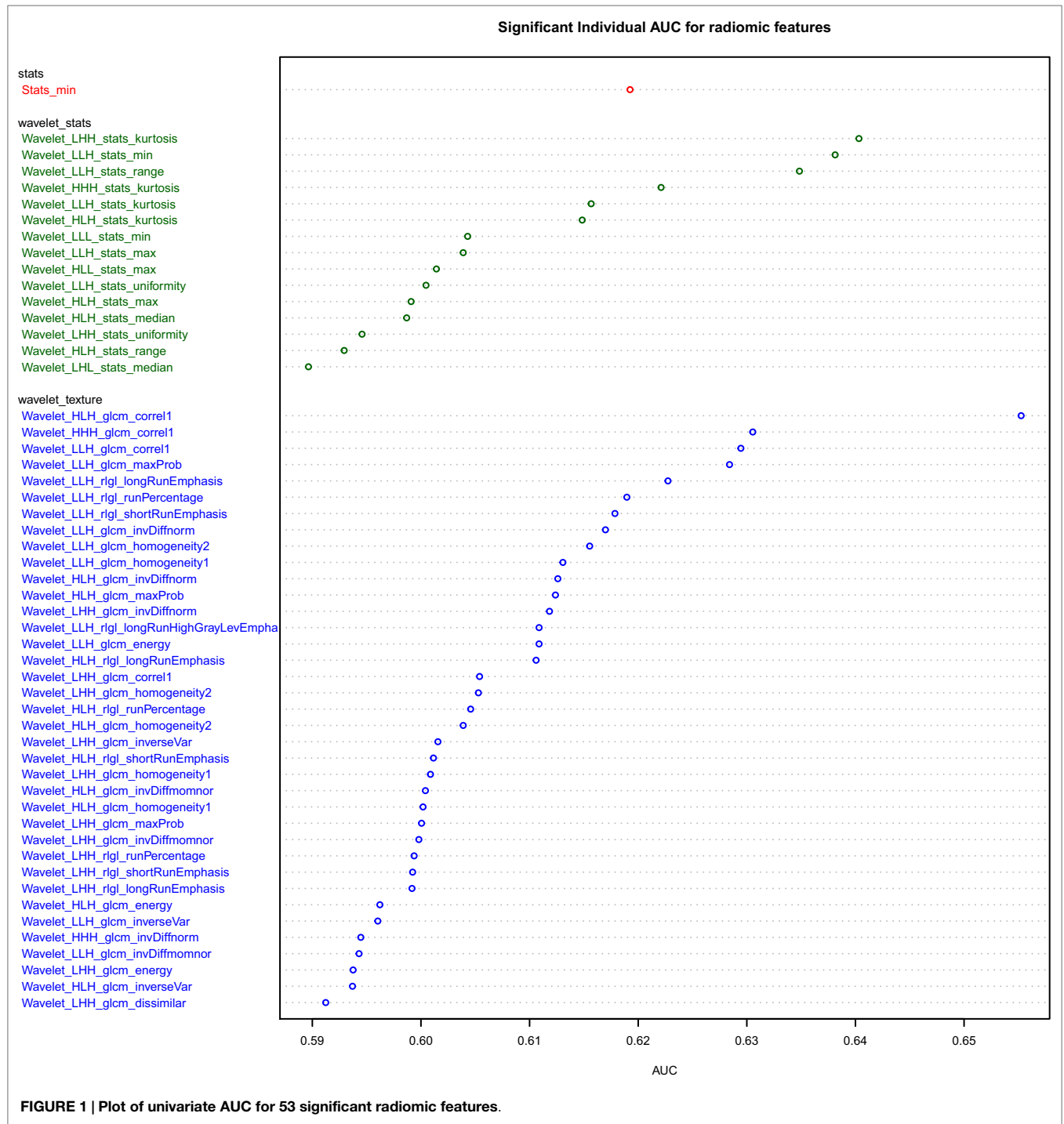
In the second stage, we applied univariate feature selection methods to the non-redundant feature set generated in the first stage and chose top-ranked features. Feature ranking methods

TABLE 1 | Feature filtering methods and corresponding scoring schemes.

Feature filtering methods	Scoring function
Information gain (44)	$IN(A_i) = \sum_{i=1}^c p(\tau_i) \log p(\tau_i) - \sum_{j=1}^{a_{mi}} \sum_{i=1}^c p(\tau_i a_{i,j}) \log p(\tau_i a_{i,j})$
Gain ratio (45)	$GR(A_i) = \frac{\sum_{i=1}^c p(\tau_i) \log p(\tau_i) - \sum_{j=1}^{a_{mi}} \sum_{i=1}^c p(\tau_i a_{i,j}) \log p(\tau_i a_{i,j})}{\sum_{j=1}^{a_{mi}} p(a_{ij}) \log p(a_{ij})}$
Gini index (46)	$GI(A_i) = \sum_{i=1}^c p^2(\tau_i) - \sum_{j=1}^{a_{mi}} \sum_{i=1}^c p^2(\tau_i a_{i,j})$
MDL (47)	$MDL(A_i) = \frac{1}{n} \log_2 \left(\begin{matrix} n \\ n_1, \dots, n_c \end{matrix} \right) - \sum_{j=1}^{a_{mi}} \left(\begin{matrix} n_j \\ n_j, \dots, n_j \end{matrix} \right) + \log_2 \left(\begin{matrix} n+c+1 \\ c-1 \end{matrix} \right) - \sum_{j=1}^{a_{mi}} \left(\begin{matrix} n_j+c-1 \\ c-1 \end{matrix} \right)$
DKM (40)	$DKM(A_i) = \sum_{i=1}^c 2\sqrt{p(\tau_{\max})(1-p(\tau_{\max}))} - \sum_{j=1}^{a_{mi}} \sum_{i=1}^c \sqrt{p(\tau_{\max} a_{ij})(1-p(\tau_{\max} a_{ij}))}$

where n_j is number of samples from class i th with j th value of features

where $p(\tau_{\max}) = \max_{i=1to c} p(\tau_i)$ represents most probable class value



that we have used in this study are Gini index, Information Gain, Gain ratio, MDL, DKM and ReliF, and their variants (40, 41). Detailed description of these methods can be found in documentation of R package “CORElearn” (42).

Most of the used feature selection methods rank features based on their discriminating abilities between classes. Assume we have a set of m dimensional feature vectors $A = (A_1, A_2, \dots, A_m)$ and c number of labeled classes $\tau = (\tau_1, \tau_2, \dots, \tau_c)$. They evaluate each

feature by the purity gained by adding the split on $A_i = a_{i,j}$. Split is the partitioning of samples according the values of feature at evaluation. $p(\tau_i)$ is the probability of class τ_i , and $p(\tau_i|a_{i,j})$ is the probability of class τ_i conditioned on the feature A_i has value $a_{i,j}$ (43). These information theory-based feature selection approaches and their corresponding scoring functions are described in **Table 1**. The derivatives and equations are cited and summarized from the Ref. (40, 43).

ReliefF

ReliefF (44) evaluates partitioning power of features based on how well their values distinguish between very similar instances. Given a randomly selected instance R_i , it searches k -nearest neighbors from the same class, and calls them “nearest hits” H , and also K -nearest neighbors from the different class, and called them “nearest miss” M . It then updates the weight W_v for all attributes depending on R_i (40). The process is repeated for m times and result is averaged over m iterations, the function for iteration v is:

$$W_v = W_v - \frac{1}{m} \text{con}(A_v, R_i, H) + \frac{1}{m} \sum_{t=1, t \neq R_i}^c \frac{p(\tau_t) \text{con}(A_v, R_i, M(C))}{1 - p(R_i, \tau)}$$

where $\text{con}(A_v, R_i, H)$ and $\text{con}(A_v, R_i, M)$ are the contribution functions of nearest neighbors (hits and misses). For example, a simple contribution function can be averaging differences of feature’s value for K -Neighbors: $\text{con}(A_v, R_i, H) = \frac{1}{k} \sum_{j=1}^k \text{diff}(A_v, R_i, H_j)$. ReliefF can efficiently evaluate features when there are strong dependencies among them (40), but like other feature filtering methods, it could not discriminate between redundant features.

For each feature ranking method, we varied the selection size from 5 to 45 (5, 10, 15, . . . , 45) and fit three classifier on the selected feature subset using training dataset. We then evaluate those classifiers by evaluating their prediction performance on validation cohort. Performance of each classifier is measured

by AUC. All feature ranking part is performed with R package “CORElearn” in R 3.2.0 (42).

Classifier Models

Three classifiers: random Forests, Naive Bayesian, and K -nearest neighbors were evaluated in this study.

Random Forest is an ensemble learning method for classification, which consists of a collection of decision trees (48). It uses weighted average of those trees for the final decision. This classifier is robust to noises and outliers, and can handle high dimensional spaces fast, but it has been observed to have overfitting problem (48). In this experiment, we set the number of decision trees to 100 and the number of nearest instances for weighted classification to 30.

Naive Bayes is another efficient learning algorithms for machine learning. It is a probabilistic classifier based on Bayes’s rule and strong conditional independence assumption among features. The probability of an example $E = (x_1, x_2, \dots, x_n)$ belonging to class c is defined as $p(c|E) = \frac{p(E|c)p(c)}{p(E)}$. E is classified to a positive class if and only if $f_b(E) = \frac{p(C=+|E)}{p(C=-|E)} \geq 1$. Assume all features are independent and given the class $p(E|c) = p(x_1, x_2, \dots, x_n|c) = \prod_{i=1}^n \frac{p(x_i|C=+)}{p(x_i|C=-)}$, the Bayes’s classifier is: $f_{nb}(E) = \frac{p(C=+)}{p(C=-)} \prod_{i=1}^n \frac{p(x_i|C=+)}{p(x_i|C=-)}$.

Naive Bayes has advantage of requiring small amount of training data to estimate each parameter. Although Naive Bayes has strong independency assumptions, which is hardly to meet in practice, it has been shown to be effective even in cases of completely deterministic dependency among features (49).

K -nearest neighbors is another non-parametric method used for classification. It is one of the simplest learning methods. The advantage of K -nearest neighbors is that it makes no assumption about the characteristic of the features. One major problem is that it is easily misled by irrelevant features to the classification

TABLE 2 | Classification accuracy of the optimal classifier.

Optimal cut-off	Sensitivity	Specificity	PPV	NPV	Accuracy
0.271	0.55	0.8	0.72	0.65	0.70

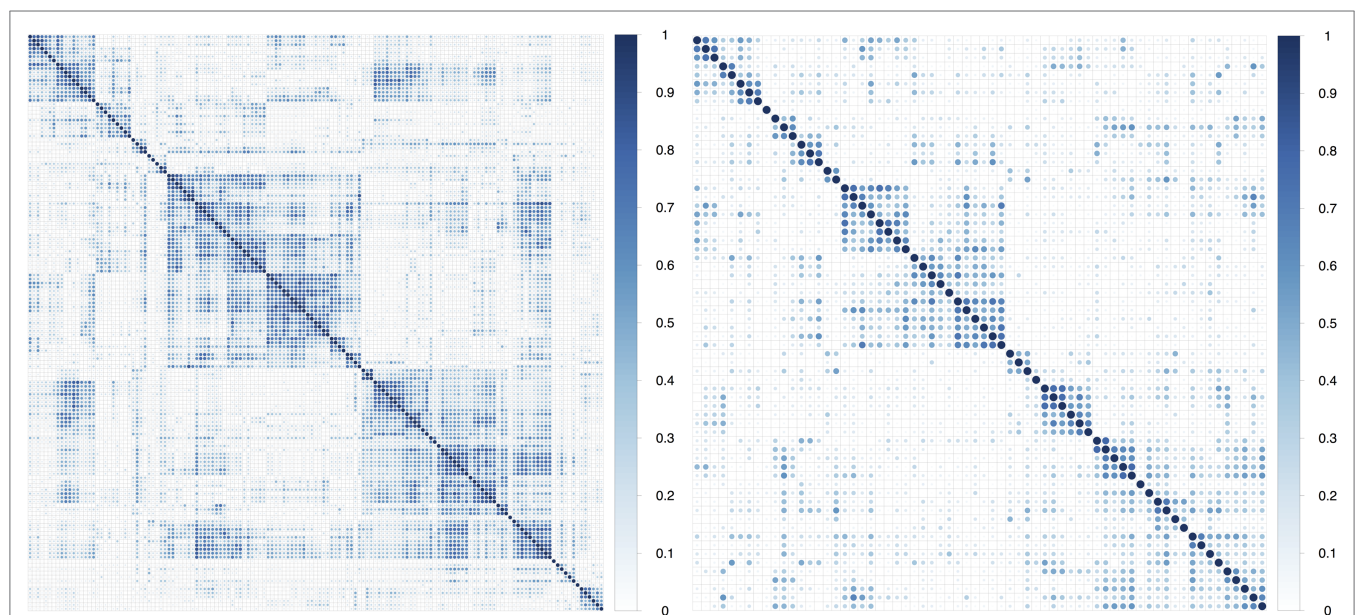


FIGURE 2 | Plot of absolute value of pair-wise Pearson correlations of radiomic features in the training dataset before and after correlation filtering.

(noise) and highly susceptible to curse of dimensionality. Therefore feature selection is important before fitting this model (50). Moreover, it is computationally intensive classification method. We used $k = 9$ for the implementation of this method.

RESULTS

A total of 440 radiomic features were investigated in terms of their association with and power to predict tumor histology. Feature selection and classification training was done using the training cohort Lung1 ($n = 198$), and the performance was evaluated in the independent validation cohort Lung2 ($n = 152$).

Univariate analysis of the training dataset showing 53 features have significant predictive power (5% FDR corrected), nearly all of which are wavelet transformed features (Figure 1). Wavelet_HLH_glc_m_correl1 had the highest AUC of 0.66 (CI:

0.57–0.74, p value 0.003). The adenocarcinoma subgroup has a higher value than the squamous carcinoma subgroup for nine gray-level cooccurrence matrix (GLCM)-based texture features (HLH and LLH wavelet transformed Energy, Homogeneity1, Homogeneity2, Inverse Variance, LLH wavelet transformed Maximum Probability), and two Gray-Level Run-Length texture features (LLH wavelet transformed Long-Run Emphasis and Long-Run High-Gray Level Emphasis). On the other hand, the squamous carcinoma subgroup has a higher value for four RLGL features (HLH and LLH wavelet-transformed Run Percentage and Short-Run Emphasis) and one statistic feature (LLH wavelet transformed Kurtosis).

In multivariate analysis, we observed that about 75% of the features had absolute pair-wise Pearson correlations higher than 0.8, and 67% were over 0.9 (Table 2). To reduce redundancy, we removed features having high absolute pair-wise correlation

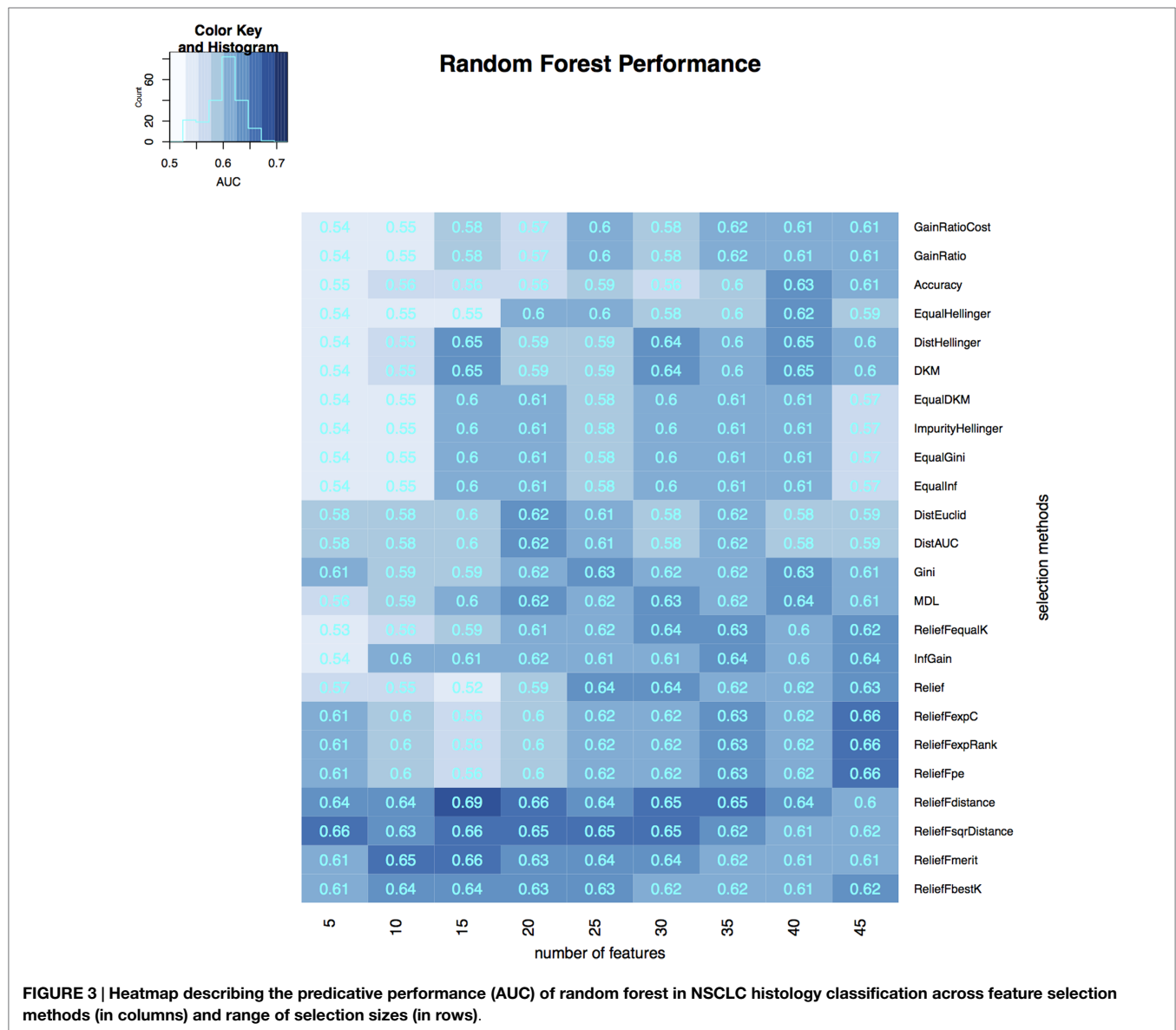


FIGURE 3 | Heatmap describing the predicative performance (AUC) of random forest in NSCLC histology classification across feature selection methods (in columns) and range of selection sizes (in rows).

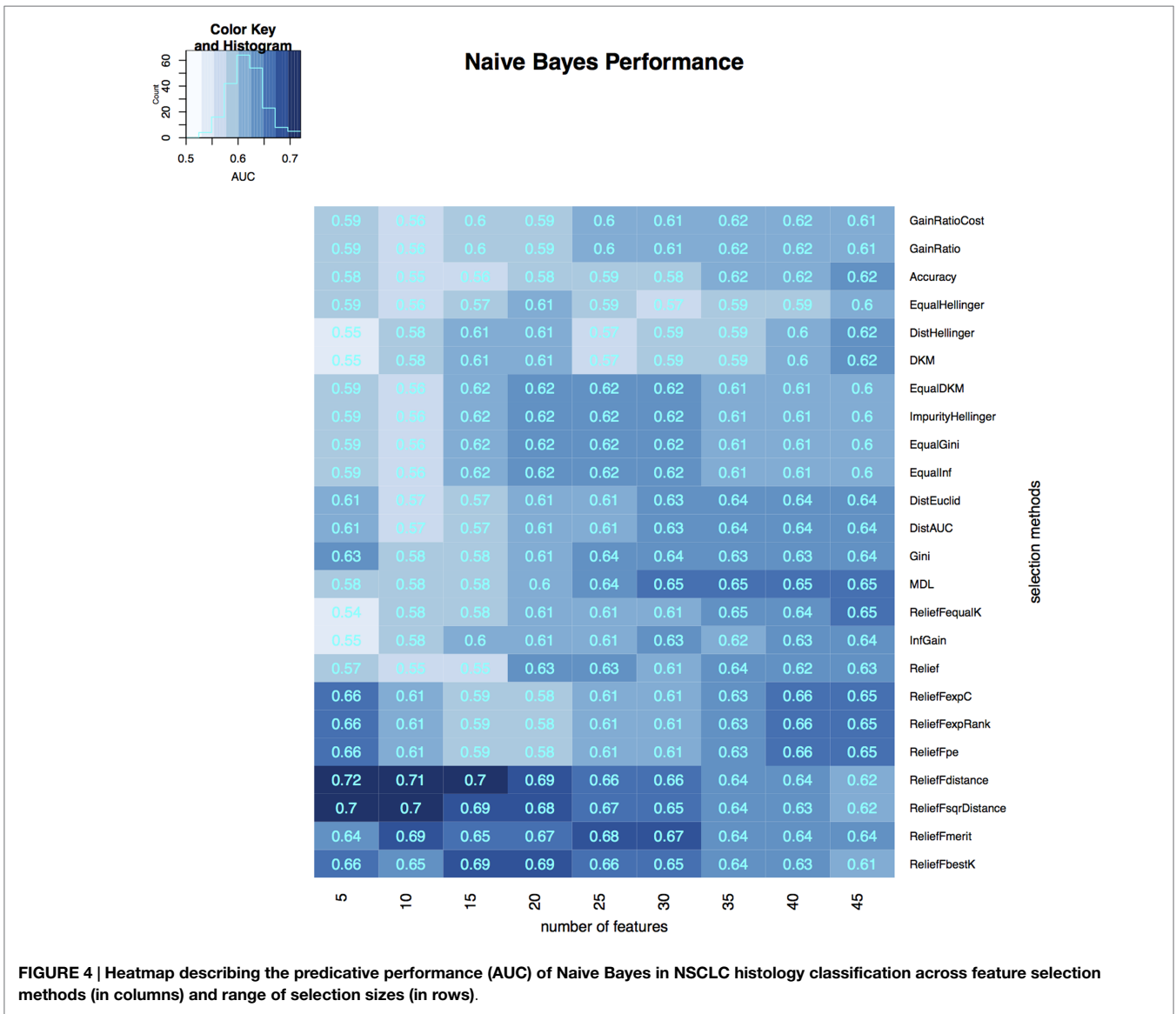


FIGURE 4 | Heatmap describing the predicative performance (AUC) of Naive Bayes in NSCLC histology classification across feature selection methods (in columns) and range of selection sizes (in rows).

($C = 0.8$). The feature number was reduced from 440 to 67 after eliminating redundant features (Figure 2). The mean of the absolute pair-wise correlations was reduced from 0.36 to 0.18 and the interquartile range (IQR) for the correlations was reduced from 0.39 to 0.20.

To select the most relevant features, we applied 24 filtering methods to the reduced feature sets. For each filtering method, we incrementally selected 5–45 features with an increment of 5 features (5, 10, 15, . . . , 45). Three classifiers were then developed on selected features in the training dataset, and the classification accuracy of each classifier on the validation dataset (Figures 3–5). The model with the best performance ($AUC = 0.72$, p value = 2.3×10^{-7}) was Naive Bayes, with five predictors selected by ReliefDistance. We obtained the optimal cutoff on the ROC curve of training cohort and used that cutoff of probability score on validation cohort to measure other prediction measures (Table 2).

As far as feature selection method is concerned, ReliefDistance showed the highest predictive performance with all three classifiers: random Forest ($AUC = 0.69$), Naive Bayes ($AUC = 0.72$), and K-nearest neighbors ($AUC = 0.64$). Feature selection method ImpurityHellinger for Random forest ($AUC = 0.61$), Gain Ratio for K-Nearest Neighbor ($AUC = 0.55$), and EqualHellinger ($AUC = 0.62$) for Naive Bayes showed lowest predictive performance. It can be observed that features evaluated using Relief variants had the most favorable performances for all three classifiers (see Figures 3–5).

In order to compare overall performance the classifiers, we used the median AUC across all 24 feature selection method as a representative AUC of each classifier. Naive Bayes had the highest performance while K-Nearest Neighbor showed the lowest performance. Random Forest is the least sensitive to feature selection methods as it showed very little standard deviation in AUC (Table 3).

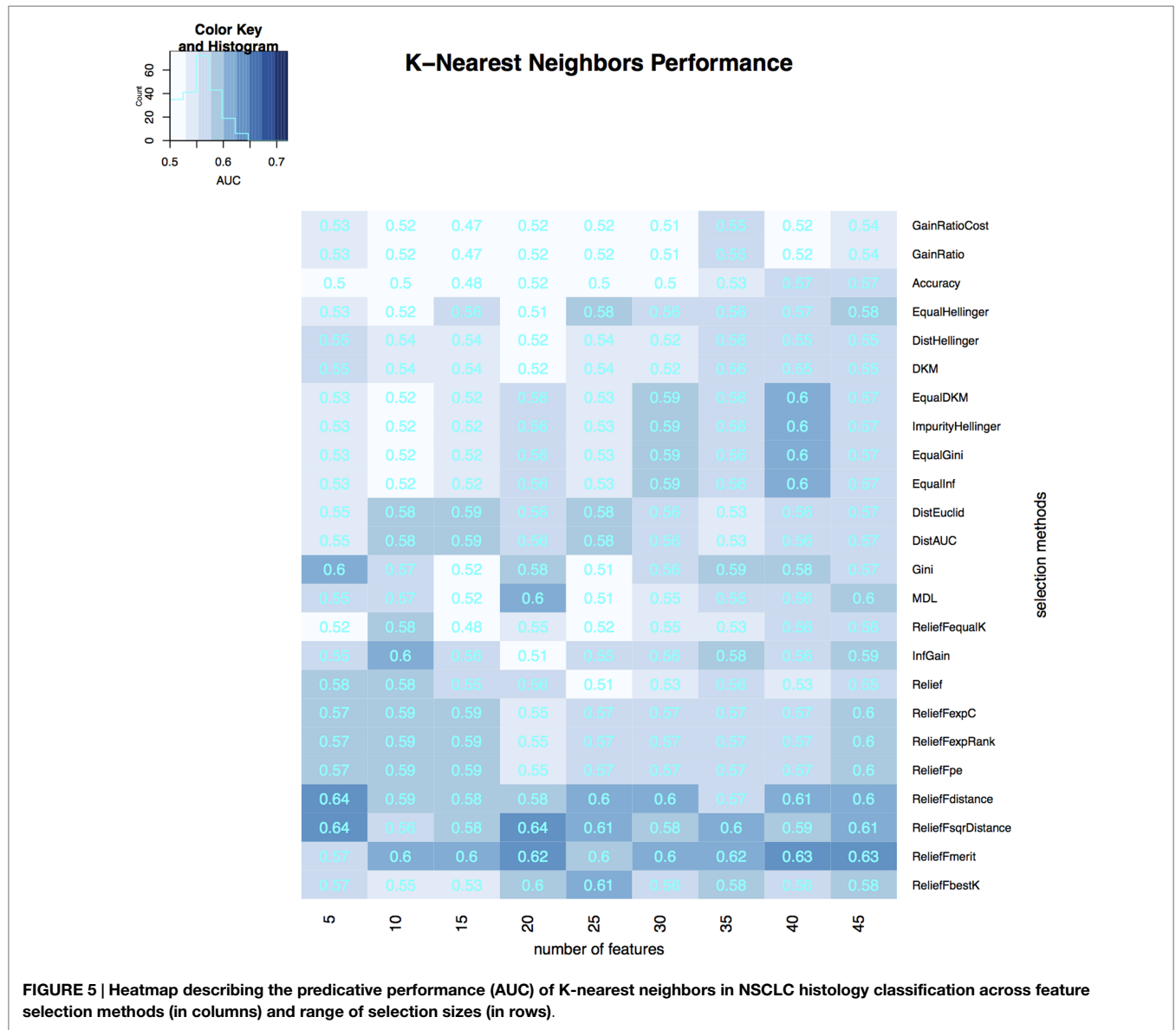


FIGURE 5 | Heatmap describing the predicative performance (AUC) of K-nearest neighbors in NSCLC histology classification across feature selection methods (in columns) and range of selection sizes (in rows).

DISCUSSION

Medical imaging has the capacity to assess tissues characteristics non-invasively, and therefore it is routinely used for diagnostic and treatment purposes in cancer care. An emerging field radiomics quantifies phenotypic characteristic of tumor tissues on medical images. In this study, we investigated the association of radiomic features and the NSCLC histological subtypes. In our univariate analysis, we observed 53 features having significant association with histological subtypes. Despite the difference in the class distributions between training and validation dataset, the multivariate machine learning models achieved high prediction accuracy, which suggests the effectiveness of these advanced machine-learning approaches as well as the strong association of radiomic features and NSCLC histology.

Our study showed that ReliefF and its variants were optimal among the 24 feature selection methods assessed.

TABLE 3 | Table describing the median value of AUC, the optimal number of features, and AUC for best/worst features selection methods.

	AUC (median ± SD)	Optimal feature number	Best/worst feature selection method (AUC)
Naive Bayes	0.64 ± 0.028	5	ReliefFdistance (0.72)/ EqualHellinger (0.62)
Random forest	0.63 ± 0.012	15	ReliefFdistance (0.69)/ ImpurityHellinger (0.61)
K-Nearest neighbor	0.60 ± 0.23	20	ReliefFdistance (0.64)/ gain ratio (0.55)

Particularly, ReliefFbestK, ReliefFmerit, ReliefFdistance, and ReliefFsqrDistance were consistently effective for all three classifiers. One reason for this is that the ReliefF-family does not assume the independence of features, unlike many other feature

selection methods. The ReliefF algorithms are able to detect context information among features and thus more accurately deals with situations where dependencies are present (51). However, like other feature selection methods, ReliefF is also unable to detect redundant features which tend to have similar scores for evaluation (52). We took care of this problem by performing correlation-based feature elimination before feature selection stage.

For the performance of three classifiers, Naive Bayes performs better than Random Forest and K-nearest neighbors. Although the median performance across feature selection methods was about the same for Naive Bayes and Random Forest, the best model achieved by Naive Bayes (AUC: 0.72; p value = 2.3×10^{-7}) was higher than Random Forest (AUC: 0.68; p value = 1.38×10^{-5}). K-nearest neighbors has the lowest performance among the three classifiers, and ReliefF could only slightly improve its accuracy. This may be because K-nearest neighbors is very sensitive to noise (50).

A shortcoming of our study is that the cut-off $M = 0.8$ is arbitrarily chosen in the correlation elimination part. One could set a more stringent or relaxed threshold based on the degree of redundancy in the dataset. A better approach would use a range of thresholds, combine them with second stage feature selection, and choose one with the most favorable result. Additionally, this two-stage feature selection is expected to fail when the interaction of two non-informative features has strong predictive power. In this case, a multivariate feature selection method like mRMR is more suitable (35).

Recently, Parmar et al. (22) identified and validated cancer specific radiomic feature clusters using consensus clustering, which provided an important tool to enhance the feature selection process. Their study found radiomic features' association with histology using an unsupervised method and achieved AUC = 0.64 for prediction. In our study by using supervised feature selection methods, we achieved higher AUCs. In another study, Basu et al. used decision trees and support vector machines for tumor classification, and results showed that ReliefF outperformed wrapper methods for 2D radiomic features (28). However, unlike our study, their results were based on a smaller cohort ($n = 74$) and lacked independent validation due to the unavailability of an independent validation cohort. Lastly Zhang et al. presented a two-stage feature selection method by combining ReliefF and mRMR (31). They showed that such a hybrid feature selection approach could improve the effectiveness of gene selection, and this could provide better discrimination for biological subtypes. This new algorithm's ability in selecting radiomic features for histology classification should be evaluated in further studies.

REFERENCES

1. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* (2001) **98**:13790–5. doi:10.1073/pnas.191502998
2. Risch A, Plass C. Lung cancer epigenetics and genetics. *Int J Cancer* (2008) **123**:1–7. doi:10.1002/ijc.23605
3. Travis WD. Pathology of lung cancer. *Clin Chest Med* (2011) **32**:669–92. doi:10.1016/j.ccm.2011.08.005

It should also be noted that retrospective cohorts-based radiomic studies like this, face many challenges. Radiomic features are sensitive to the variability of imaging scanners and scanning parameters (53, 54), tumor delineation methods (55, 56), image reconstruction (57), discretization techniques (58), etc. These different factors are not controlled for retrospective cohorts, which maybe one of the reason for not so impressive performance. Prospective cohorts created by controlling these factors could provide higher performance. In future studies, the performance of classifiers can be enhanced if we incorporate genetic signatures and clinical features like tumor grade, location, smoking history, and obesity (59, 60). For example, a recent study showed that body mass index was inversely associated with squamous carcinoma, yet for adenocarcinoma the association was positive (60). It is also important to take genetic heterogeneity into account. Recent studies examining gene expression profiles have identified several genes distinguishing adenocarcinoma and squamous carcinoma (61–63). Hence, future study incorporating clinic characteristics and genomic data will improve classification accuracy.

In conclusion, radiomic features have strong predictive power for classification of NSCLC histology. With expanding cohorts and improving image feature extraction techniques, we expect higher accuracy in classification using radiomics. This work will serve as a promising diagnostic tool for identifying lung cancer histology in a non-invasive way and thus informing treatment choices and personalized therapy for lung cancer patients.

AUTHOR CONTRIBUTIONS

HA, WW, and CP conceived of the project, analyzed the data, and wrote the paper. PG, JQ, JB, PL, and RM provided expert guidance, data, or analysis tools, and reviewed the manuscript.

ACKNOWLEDGMENTS

Authors acknowledge financial support from the National Institute of Health (NIH-USA U24CA194354 and NIH-USA U01CA190234). This research is also supported by the Dutch technology Foundation STW (grant no 10696 DuCAT). Authors also acknowledge financial support from EU 7th framework program (ARTFORCE – no 257144, REQUITE – no 601826), SME Phase 2 (EU proposal 673780 – RAIL), the European Program H2020-PHC-2015 (BD2decide, no 210274050), Alpe d'HuZes-KWF (DESIGN).

4. Rekhtman N, Ang DC, Sima CS, Travis WD, Moreira AL. Immunohistochemical algorithm for differentiation of lung adenocarcinoma and squamous cell carcinoma based on large series of whole-tissue sections with validation in small specimens. *Mod Pathol* (2011) **24**:1348–59. doi:10.1038/modpathol.2011.92
5. Schuurbiens OC, Meijer TW, Kaanders JH, Looijen-Salamon MG, de Geus-Oei LE, van der Drift MA, et al. Glucose metabolism in NSCLC is histology-specific and diverges the prognostic potential of 18FDG-PET for adenocarcinoma and squamous cell carcinoma. *J Thorac Oncol* (2014) **9**:1485–93. doi:10.1097/JTO.0000000000000286

6. Meijer TW, Schuurbiens OC, Kaanders JH, Looijen-Salamon MG, de Geus-Oei LF, Verhagen AF, et al. Differences in metabolism between adeno- and squamous cell non-small cell lung carcinomas: spatial distribution and prognostic value of GLUT1 and MCT4. *Lung Cancer* (2012) **76**:316–23. doi:10.1016/j.lungcan.2011.11.006
7. Manegold C. Treatment algorithm in 2014 for advanced non-small cell lung cancer: therapy selection by tumour histology and molecular biology. *Adv Med Sci* (2014) **59**:308–13. doi:10.1016/j.advms.2014.08.008
8. Cufer T, Ovarcicek T, O'Brien ME. Systemic therapy of advanced non-small cell lung cancer: major-developments of the last 5-years. *Eur J Cancer* (2013) **49**:1216–25. doi:10.1016/j.ejca.2012.11.021
9. Travis WD. Classification of lung cancer. *Semin Roentgenol* (2011) **46**:178–86. doi:10.1053/j.ro.2011.02.003
10. Barash O, Peled N, Tisch U, Bunn PA, Hirsch FR, Haick H. Classification of lung cancer histology by gold nanoparticle sensors. *Nanomedicine* (2012) **8**:580–9. doi:10.1016/j.nano.2011.10.001
11. Scagliotti GV, Parikh P, Pawel von J, Biesma B, Vansteenkiste J, Manegold C, et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol* (2008) **26**:3543–51. doi:10.1200/JCO.2007.15.0375
12. Scagliotti G, Hanna N, Fossella F, Sugarman K, Blatter J, Peterson P, et al. The differential efficacy of pemetrexed according to NSCLC histology: a review of two phase III studies. *Oncologist* (2009) **14**:253–63. doi:10.1634/theoncologist.2008-0232
13. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RT, 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
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, van Stiphout RG, Starmans MH, Rios Velazquez E, Nalbantov G, Aerts HJ, et al. Predicting outcomes in radiation oncology–multifactorial decision support systems. *Nat Rev Clin Oncol* (2013) **10**:27–40. doi:10.1038/nrclinonc.2012.196
16. Hayano K, Yoshida H, Zhu AX, Sahani DV. Fractal analysis of contrast-enhanced CT images to predict survival of patients with hepatocellular carcinoma treated with sunitinib. *Dig Dis Sci* (2014) **59**:1996–2003. doi:10.1007/s10620-014-3064-z
17. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *Eur Radiol* (2012) **22**:796–802. doi:10.1007/s00330-011-2319-8
18. Tiwari P, Kurhanewicz J, Madabhushi A. Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS. *Med Image Anal* (2013) **17**:219–35. doi:10.1016/j.media.2012.10.004
19. Chen HH, Su W, Hsueh W, Wu Y, Lin F. Summation of F18-FDG uptakes on PET/CT images predicts disease progression in non-small cell lung cancer. *Int J Radiat Oncol* (2010) **78**:S504. doi:10.1016/j.ijrobp.2010.07.1178
20. Nair VS, Gevaert O, Davidzon G, Napel S, Graves EE, Hoang CD, et al. Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer. *Cancer Res* (2012) **72**:3725–34. doi:10.1158/0008-5472.CAN-11-3943
21. Fleshman JW, Myerson RJ, Fry RD, Kodner IJ. Accuracy of transrectal ultrasound in predicting pathologic stage of rectal cancer before and after pre-operative radiation therapy. *Dis Colon Rectum* (1992) **35**:823–9. doi:10.1007/BF02047866
22. Parmar C, Leijenaar RT, 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**:11044. doi:10.1038/srep11044
23. Ganeshan B, Abaleke S, Young RC, Chatwin CR, Miles KA. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. *Cancer Imaging* (2010) **10**:137–43. doi:10.1102/1470-7330.2010.0021
24. Hall WA, Djallilian HR, Sperduto PW, Cho KH, Gerbi BJ, Gibbons JP, et al. Stereotactic radiosurgery for recurrent malignant gliomas. *J Clin Oncol* (1995) **13**:1642–8.
25. Vaidya M, Creach KM, Frye J, Dehdashti F, Bradley JD, Naqa El I. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother Oncol* (2012) **102**:239–45. doi:10.1016/j.radonc.2011.10.014
26. Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nat Biotechnol* (2007) **25**:675–80. doi:10.1038/nbt1306
27. Zinn PO, Mahajan B, Majadan B, Sathyan P, Singh SK, Majumder S, et al. Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme. *PLoS One* (2011) **6**:e25451. doi:10.1371/journal.pone.0025451
28. Basu S, Hall LO, Goldgof DB, Gu Y, Kumar V, Choi J, et al. Developing a classifier model for lung tumors in CT-scan images. *IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2011*. Anchorage, AK: IEEE (2011). p. 1306–12.
29. Friedman JH. On bias, variance, 0/1 – loss, and the curse-of-dimensionality. *Data Min Knowl Discov* (1997) **1**:55–77. doi:10.1023/A:1009778005914
30. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* (2003) **3**:1157–82.
31. Zhang Y, Ding C, Li T. Gene selection algorithm by combining reliefF and mRMR. *BMC Genomics* (2008) **9**(Suppl 2):S27. doi:10.1186/1471-2164-9-S2-S27
32. 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
33. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* (1982) **143**:29–36. doi:10.1148/radiology.143.1.7063747
34. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* (2011) **27**:3206–8. doi:10.1093/bioinformatics/btr511
35. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol* (2005) **3**:185–205. doi:10.1142/S0219720005001004
36. Li W, Yang Y. How many genes are needed for a discriminant microarray data analysis. *Methods of Microarray Data Analysis*. Boston, MA: Springer US (2002). p. 137–49.
37. Xiong M, Fang X, Zhao J. Biomarker identification by feature wrappers. *Genome Res* (2001) **11**:1878–87. doi:10.1101/gr.190001
38. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* (2013) **36**:27–46. doi:10.1111/j.1600-0587.2012.07348.x
39. Wei T. *Corrplot: Visualization of a Correlation Matrix*. (2013). R package version 0.60. Available from: <http://CRAN.R-project.org/package=corrplot>
40. Robnik-Šikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Mach Learn* (2003) **53**:23–69. doi:10.1023/A:1025667309714
41. Dhillon PS, Foster D, Ungar LH. Minimum description length penalization for group and multi-task sparse learning. *J Mach Learn Res* (2011) **12**:525–64.
42. Robnik-Šikonja M, Savicky P. *CORElearn: CORElearn-Classification, Regression, Feature Evaluation and Ordinal Evaluation* (2013). R package version 0.9 42.
43. Robnik-Šikonja M. Experiments with cost-sensitive feature evaluation. In: Lavrač N, Gamberger D, Todorovski L, Blockeel H, editors. *Machine Learning: ECML*. Berlin: Springer (2003). p. 325–36. Lecture Notes in Computer Science.
44. Kononenko I. Estimating attributes: Analysis and extensions of RELIEF. In: Bergadano F, De Raedt L, editors. *Machine Learning: ECML-94*. Berlin: Springer (1994). p. 171–82. Lecture Notes in Computer Science.
45. Quinlan JR. *CA. 5: Programming for Machine Learning*. Burlington, MA: Morgan Kaufmann (2014).
46. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Belmont: Wadsworth (1984).
47. Kononenko I. On biases in estimating multi-valued attributes. *IJCAI*. Montreal: Morgan Kaufmann Publishers Inc. (1995). p. 1034–40.
48. Breiman L. Random forests. *Mach Learn* (2001) **45**:5–32. doi:10.1023/A:1010933404324
49. Rish I. *An Empirical Study of the Naive Bayes Classifier*. New York: IBM (2001). p. 41–6.

50. Cunningham P, Delany SJ. k-Nearest neighbour classifiers. *Multiple Classifier Syst* (2007), p. 1–17.
51. Kira K, Rendell LA. *The Feature Selection Problem: Traditional Methods and a New Algorithm*. San Jose, CA: AAAI Press (1992), p. 129–34.
52. Wu T, Xie K, Nie C, Song G. An adaptation of relief for redundant feature elimination. In: Wang J, Yen GG, Polycarpou M, editors. *Advances in Neural Networks – ISNN*. Berlin: Springer (2012), p. 73–81. Lecture Notes in Computer Science.
53. Mattonen SA, Huang K, Ward AD, Senan S, Palma DA. New techniques for assessing response after hypofractionated radiotherapy for lung cancer. *J Thorac Dis* (2014) **6**:375–86. doi:10.3978/j.issn.2072-1439.2013.11.09
54. Alobaidli S, McQuaid S, South C, Prakash V, Evans P, Nisbet A. The role of texture analysis in imaging as an outcome predictor and potential tool in radiotherapy treatment planning. *Br J Radiol* (2014) **87**:20140369. doi:10.1259/bjr.20140369
55. 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
56. Zhao B, Tan Y, Tsai WY, Schwartz LH, Lu L. Exploring variability in CT characterization of tumors: a Preliminary Phantom Study. *Transl Oncol* (2014) **7**:88–93. doi:10.1118/1.4793409
57. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol* (2010) **49**:1012–6. doi:10.3109/0284186X.2010.498437
58. Leijenaar RT, Nalbantov G, Carvalho S, van Elmpt WJ, Troost EG, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep* (2015) **5**:11075. doi:10.1038/srep11075
59. Buffill JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med* (1990) **113**:779–88. doi:10.7326/0003-4819-113-10-779
60. Zein El M, Parent ME, Nicolau B, Koushik A, Siemiatycki J, Rousseau MC. Body mass index, lifetime smoking intensity and lung cancer risk. *Int J Cancer* (2013) **133**:1721–31. doi:10.1002/ijc.28185
61. Boelens MC, van den Berg A, Vogelzang I, Wesseling J, Postma DS, Timens W, et al. Differential expression and distribution of epithelial adhesion molecules in non-small cell lung cancer and normal bronchus. *J Clin Pathol* (2007) **60**:608–14. doi:10.1136/jcp.2005.031443
62. Kettunen E, Anttila S, Seppänen JK, Karjalainen A, Edgren H, Lindström I, et al. Differentially expressed genes in nonsmall cell lung cancer: expression profiling of cancer-related genes in squamous cell lung cancer. *Cancer Genet Cytogenet* (2004) **149**:98–106. doi:10.1016/S0165-4608(03)00300-5
63. Pietanza MC, Ladanyi M. Bringing the genomic landscape of small-cell lung cancer into focus. *Nat Genet* (2012) **44**:1074–5. doi:10.1038/ng.2415

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.

Copyright © 2016 Wu, Parmar, Grossmann, Quackenbush, Lambin, Bussink, Mak and Aerts. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.