A deep learning approach with subregion partition in MRI image analysis for metastatic brain tumor

Jiaxin Shi[1†], Zilong Zhao[2†], Tao Jiang[1], Hua Ai[1], Jiani Liu[3], Xinpu Chen[1], Yahong Luo[3], Huijie Fan[4]* and Xiran Jiang[1]*

[1]Department of Biomedical Engineering, School of Intelligent Medicine, China Medical University, Shenyang, China, [2]Department of Neurosurgery, The First Affiliated Hospital of China Medical University, Shenyang, China, [3]Department of Radiology, Cancer Hospital of China Medical University, Liaoning Cancer Hospital and Institute, Shenyang, China, [4]State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China

Purpose: To propose a deep learning network with subregion partition for predicting metastatic origins and EGFR/HER2 status in patients with brain metastasis.

Methods: We retrospectively enrolled 140 patients with clinico-pathologically confirmed brain metastasis originated from primary NSCLC ($n = 60$), breast cancer (BC, $n = 60$) and other tumor types ($n = 20$). All patients underwent contrast-enhanced brain MRI scans. The brain metastasis was subdivided into phenotypically consistent subregions using patient-level and population-level clustering. A residual network with a global average pooling layer (RN-GAP) was proposed to calculate deep learning-based features. Features from each subregion were selected with least absolute shrinkage and selection operator (LASSO) to build logistic regression models (LRs) for predicting primary tumor types (LR-NSCLC for the NSCLC origin and LR-BC for the BC origin), EGFR mutation status (LR-EGFR) and HER2 status (LR-HER2).

Results: The brain metastasis can be partitioned into a marginal subregion (S1) and an inner subregion (S2) in the MRI image. The developed models showed good predictive performance in the training (AUCs, LR-NSCLC vs. LR-BC vs. LR-EGFR vs. LR-HER2, 0.860 vs. 0.909 vs. 0.850 vs. 0.900) and validation (AUCs, LR-NSCLC vs. LR-BC vs. LR-EGFR vs. LR-HER2, 0.819 vs. 0.872 vs. 0.750 vs. 0.830) set.

Conclusion: Our proposed deep learning network with subregion partitions can accurately predict metastatic origins and EGFR/HER2 status of brain metastasis, and hence may have the potential to be non-invasive and preoperative new markers for guiding personalized treatment plans in patients with brain metastasis.

KEYWORDS

deep learning, brain metastasis, MRI, NSCLC, breast cancer

Introduction

Brain metastasis is a common complication in patients with malignant tumors and represent one of the most frequent neurological complications of systemic cancer with incidence rates of 20–50% depending on the type of primary tumor (Brastianos et al., 2015; Cagney et al., 2017). In clinical practice, the presence of brain metastasis is much more common than primary brain tumors, and has been a major cause of morbidity and mortality (Achrol et al., 2019). In recent years, the incidence of brain metastasis has increased because of advances in neuroimaging procedures (Hardesty and Nakaji, 2016). Among different types of primary tumors, non-small cell lung cancer (NSCLC, 41–56%) and breast cancer (BC, 15–30%) are the most common origins of brain metastasis (Bekaert et al., 2017). Early and accurate identification of the brain metastatic origin before surgery could change the personal treatment plan, and hence of great clinical significance (Kniep et al., 2019).

For NSCLC patients with brain metastasis, the epidermal growth factor receptor (EGFR) gene mutation status is essential for treatment strategies, such as EGFR-tyrosine kinase inhibitors (EGFR-TKIs) therapy (Guo et al., 2020). For BC patients with brain metastasis, the HER2 status is crucial for selecting therapeutic methods (Loibl and Gianni, 2017). This is because patients with HER2 positive are routinely treated with targeted antibody therapy (Koboldt et al., 2012; Lam et al., 2014) and usually have a poor prognosis (Cameron et al., 2017). Considering tissue acquisition from the primary tumor is not always clinically possible, the metastases can be an important alternate to reflect the characteristics and gene status of the primary tumor (Krawczyk et al., 2014). However, radiologists can hardly assess metastatic origins or gene status of primary tumor through visual examination on MRI data due to the absence of specific markers (Kniep et al., 2019).

Advances in computer-aided diagnosis (CAD) and artificial intelligence have played an increasingly important role in the field of medical imaging (Gore, 2020; Barragán-Montero et al., 2021). Radiomics, as an emerging field refers to systematic calculate and analysis big amount of quantitative features from medical imaging (Lambin et al., 2017). However, there were limited studies reported CAD-based identification of the metastatic origin of brain metastasis (Ortiz-Ramon et al., 2017; Béresová et al., 2018). Previous reports both assumed that the brain metastasis is homogeneous throughout the tumor volume. However, recent researches demonstrated that solid tumors are heterogeneous with some tumor regions are more biologically aggressive and may reflect different biological processes (Gatenby et al., 2013). This was known as intratumor heterogeneity (ITH), and was indicated to have significant implications that represent distinct tumor progression (O'Connor et al., 2015). Subregion-based radiomics algorithms were suggested to divide the whole tumor area into intratumor subregions, and hence allow to capture valuable information from the subregions (Lin et al., 2021). To data,

subregion radiomics analyses have been conducted in breast cancer (Fan et al., 2018), lung cancer (Shen et al., 2021) and esophageal squamous (Xie et al., 2019), and demonstrated to significantly improve the diagnostic performance of radiomics methods. To our knowledge, subregion radiomics has not been investigated in brain metastasis.

In this study, we subdivided the brain metastasis into phenotypically consistent subregions based on patient- and population-level clustering, and evaluated handcrafted features and deep learning-based features from a proposed residual network with a global average pooling layer (RN-GAP) for predicting metastatic origins of brain metastasis and assessing EGFR and HER2 status.

Materials and methods

Patients

We included 140 patients harboring brain metastasis from January 2017 to September 2020. The patients met the following criteria (i) with the diagnosis of brain metastasis confirmed by pathological examination; (ii) 3.0T brain MRI scan before treatment; and (iii) complete records of the EGFR gene mutation status by gene sequencing. The exclusion criteria included: (i) not received any anti-cancer treatment (including radiotherapy, chemotherapy or biotherapy) before the MRI examination; (ii) poor image quality or severe motion artifacts; and (iii) incomplete clinical information. To determine EGFR mutation status for the patients, genetic sequencing analysis was performed based puncture biopsy of the primary tumor before treatment. To determine HER2 status, immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) analysis were used (Ruan et al., 2017; Fehrenbacher et al., 2020).

MRI scans and tumor segmentation

For each patient, the T1CE MR were obtained with a 3.0T MRI scanner (Siemens Magnetom Verioio, Erlangen, Germany). The GA-DTPA was used as contrast agent when performing the T1CE MRI scanning. The parameters of the T1CE MRI sequence were shown in **Table 1**. To obtain the whole tumor region, a radiologist with 4 years of experience manually outlined edges of the whole tumor region layer by layer. The segmented region of interest (ROI) was stored in a.NII format and used for computer-aided analyses.

Intratumor partition

We used a three-step approach for intratumor partition of the brain metastasis. Firstly, MRI local entropy was calculated for each patient's slice using a 9×9 neighborhood. Secondly,

TABLE 1 Parameters of the brain MRI screenings.

Parameters	MRI
Repetition time/echo time (TR/TE) (ms)	120/2.48
Slice thickness (mm)	5
Spacing between slices (mm)	6.5
Acquisition matrix	320×240
Pixel spacing (mm)	0.359×0.359
Field of view (mm)	600×640

Patient-level clustering was then applied to partition the tumor regions into many superpixels based on the calculated local entropy maps and MRI pixel intensity to represent the spatial heterogeneity (Wu et al., 2016). K-means clustering (Kanungo et al., 2002) was performed using the Euclidean distance between pixel intensity and entropy intensity as a distance metric. In each ROI, pixels with similar pixel intensities and entropy intensities were clustered into many superpixels by patient-level clustering based on the principle of K-means clustering. Finally, superpixels from all patients were gathered to perform the population-level clustering via exploring the similarity of inter- and intra-patients using hierarchical clustering (Johnson, 1967). The number of partitioned subregions (clusters k) was set from 2 to 10 as suggested (Pham et al., 2005). To determine the optimal number of partitioned subregions, the Calinski-Harabasz (CH) index was used as a criterion according to a previous work (Caliński and Harabasz, 1974). Differences in pixel intensity and entropy intensity of metastatic origins and gene status were evaluated with t-test.

Deep learning network development

We proposed a RN-GAP to extract deep learning-based radiomics features from the partitioned subregions of brain metastasis in MRI images. The network architecture of ResNet50 was shown in **Supplementary Table 1** (Brito et al., 2019). **Figure 1** showed the architecture of the RN-GAP network. We used 14 million natural images from ImageNet database (Deng et al., 2009) to pre-train the ResNet50 network as Subnet 1. The Subnet 2 was composed of a GAP layer (Lin et al., 2013), a fully connected layer and a softmax output layer. Global averaging pooling (GAP) layer calculated the average value of each feature map as an output node and forms all output nodes into a feature vector, reducing number of parameters. To avoid overfitting, the GAP layer was used before the fully connected layer with 512 neurons, which was followed by a softmax output layer. To fit the RN-GAP's architecture, MRI images of the largest ROIs in the partitioned subregions were resized to 224×224, and the unnecessary background was removed. The proposed RN-GAP network was constructed based on the Python v3.6 platform.

Radiomics handcrafted and deep learning-based feature calculation

Before extraction of radiomics features, pre-processing of MR images were performed with detailed descriptions can be found in **Supplementary Table 2**. We applied the open source package Pyradiomics[1] based on the Python 3.6 platform to calculate handcrafted radiomics features from brain MRI images according to reported protocols (van Griethuysen et al., 2017).

To obtain deep learning-based radiomics features, MRI images of the largest ROIs in the segmented subregions of brain metastasis were used to feed the RN-GAP. Values of the 512 neurons from the fully connected layer were recorded and used as deep learning-based features for further analysis.

---

1 https://pypi.org/project/pyradiomics/

FIGURE 1

Our deep learning network architecture: adding a global average pooling layer based on the residual network.
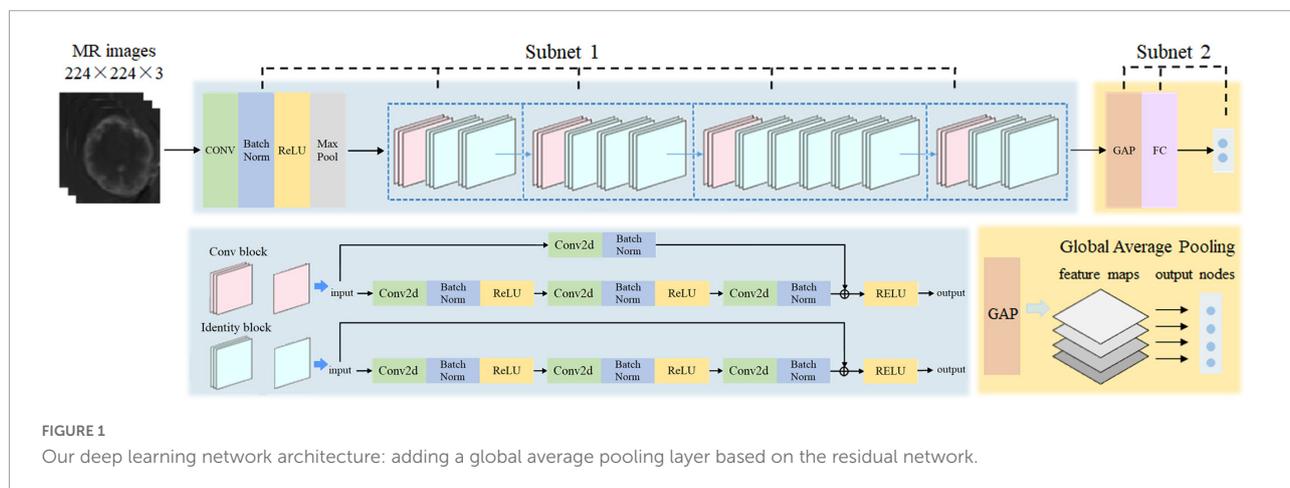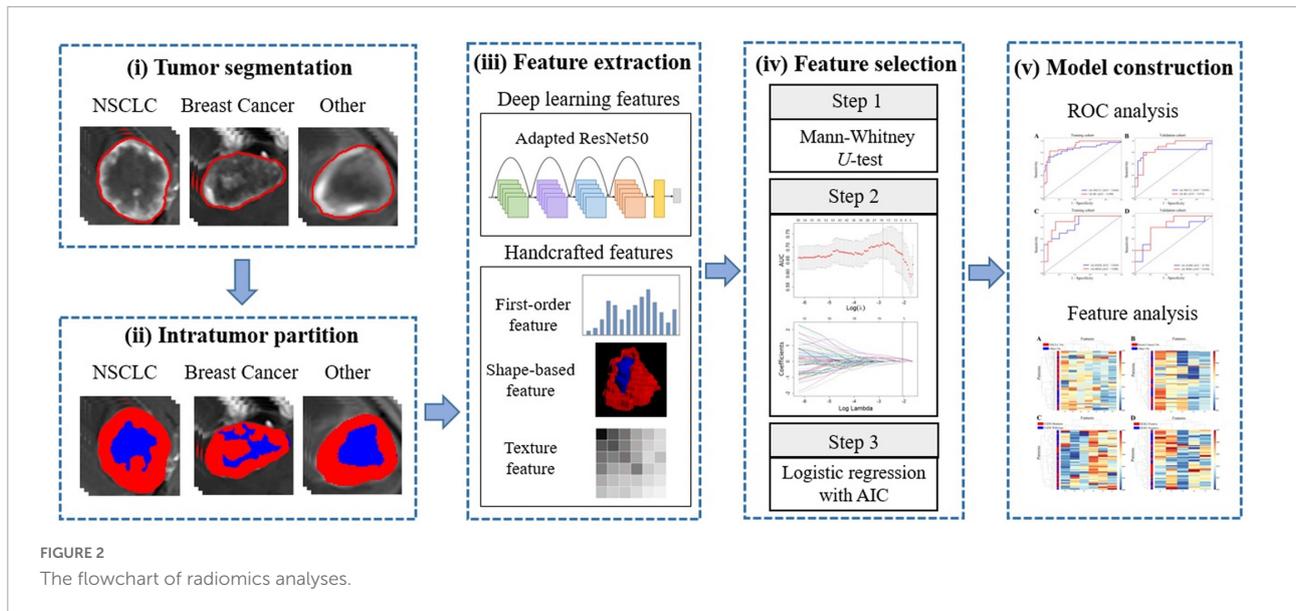
FIGURE 2
The flowchart of radiomics analyses.

Feature selection and logistic regression model construction

Inter and intra-observer reproducibilities of all features was assessed by inter-class and intar-class correlation coefficient (ICC) analyses (Leijenaar et al., 2013). Then, the Mann-Whitney U-test was performed to further select features with $P < 0.05$ as predictors with significant differences. Finally, the least absolute shrinkage and selection operator (LASSO) algorithm was used for dimensionality reduction of the calculated features with 10-fold cross-validation. The features were substituted into multivariate logistic regression analysis using the stepwise regression method to obtain the respective regression coefficients of the most predictive features. The feature selecting method was applied to both handcrafted features and deep learning-based features. Coefficients of the selected features were weighted to construct logistic regression models (LRs) for predicting metastatic origins (LR-NSCLC and LR-BC), EGFR mutation status (LR-EGFR) and HER2 status (LR-HER2). Figure 2 showed the flowchart of this study.

Results

Patient characteristics

We finally included 140 patients in the study, 41 males and 99 females, aged 29–77 years. There were 60, 60, and 20 patients harboring brain metastases from primary NSCLC, breast cancer and other types of malignant tumors, respectively. The clinical characteristics of the enrolled patients were shown in Table 2.

Intratumor partition

Figure 3 shows results of intratumoral partitions in the MRI images of brain metastasis using a patient- and population-level clustering approach. The brain metastasis can be divided into two subregions, which were denoted as S1 (marginal subregion) and S2 (inner subregion). There were obvious differences between S1 and S2. Supplementary Figure 1 showed different CH values when the clusters k was set from 2 to 10. The optimal CH index was calculated to decide the number of intratumoral subregions when k equals to 2. Figure 4 indicated that the distribution of MRI pixel intensity and local entropy of the S1 and S2 were significantly different. Compared with S1 and S2, S1 was more associated with both MRI pixel intensity and local entropy. This indicates a higher degree of tumor heterogeneity in S1.

TABLE 2   Clinical characteristics of the enrolled patients.

Tumor type	Gene status	Gender		Age (Mean ± SD)
		Female	Male	
NSCLC		30	30	57.82 ± 7.26
	EGFR mutant	20	10	56.67 ± 7.20
	EGFR wild-type	10	20	58.97 ± 7.13
Breast cancer		60	0	53.63 ± 10.53
	HER2 positive	30	0	54.03 ± 10.82
	HER2 negative	30	0	53.23 ± 10.21
Other		9	11	53.23 ± 10.21
Total		99	41	58.80 ± 13.45

NCSLC, Non-small-cell lung cancer.

FIGURE 3
Intratumoral partitions in the MRI images of patients with brain metastasis. The third column represents ROIs in the MRI image. The fourth column represents local entropy maps within the tumor. The fifth column represents the results of patient-level clustering. The sixth column represents the results of population-level clustering.
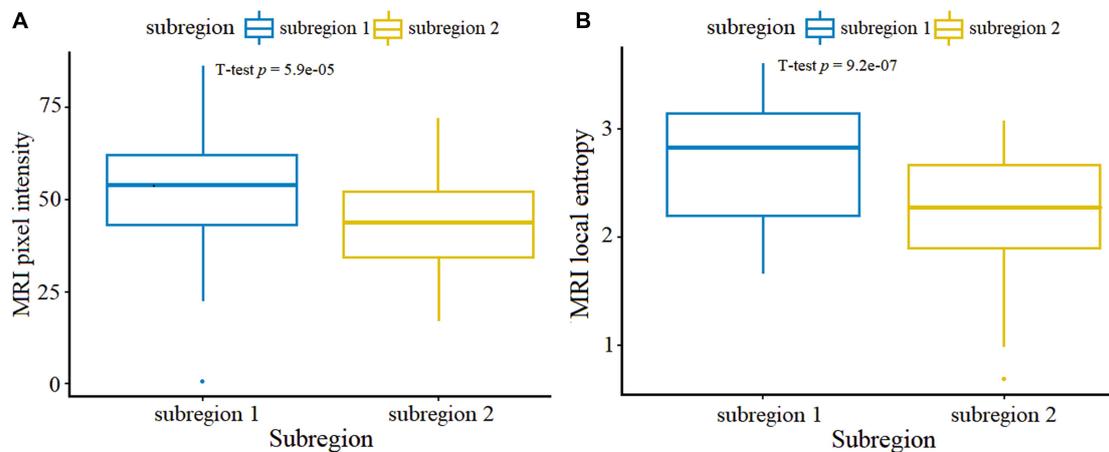
FIGURE 4
Boxplots showed MRI pixcel intensity **(A)** and local entropy **(B)** between the partitioned S1 and S2 in all patients. *P*-values were obtained using the *t*-test.

Development and evaluation of logistic regression models

All handcrafted and deep learning-based features were extracted from the partationed subregions and selected to develop logistic regression models for predicting brain metastasis originated from primary NSCLC (LR-NSCLC) and breast cancer (LR-BC), and for predicting the EGFR mutation (LR-EGFR) status and HER2 (LR-HER2) status. Formulas for each model were shown below:

LR-NSCLC = – 1.1569 + 0.2105 × wavelet-HLH_glrlm_LongRunEmphasis – 1.4567 × wavelet-LHL_glszm_SmallAreaHighGrayLevelEmphasis – 1.2819 × wavelet-HHH_firstorder_RootMeanSquared + 0.2768 × wavelet-HHH_glrlm_LongRunLowGrayLevelEmphasis + 0.9681 × lbp-3D-k_firstorder_Mean – 0.5264 × DL_215 – 0.8798 × DL_400

LR-BC = – 1.5412 + 1.9088 × wavelet-HHL_glcm_InverseVariance + 1.9784 × wavelet-HHH_firstorder_InterquartileRange – 0.9107 × DL_317 – 1.0072 × DL_144 – 0.6121 × DL_122

LR-EGFR = 0.03461 + 1.8998 × squareroot_glszm_SmallAreaHighGrayLevelEmphasi –0.8220 × log-sigma-5-mm-3D_ngtdm_Contrast + 1.6274 × wavelet-LHL_firstorder_Skewness + 0.9110 × lbp-3D-k_firstorder_MeanAbsoluteDeviation + 1.1375 × lbp-3D-m2_glszm_GrayLevelVariance + 0.0129 × DL_336

LR-HER2 = – 0.2327 + 0.5867 × wavelet-LLH_firstorder_10Percentile – 5.5909 × square_ngtdm_Busyness – 1.5990 × DL_317 + 2.4984 × wavelet-HHL_glcm_Imc1 – 0.2158 × log-sigma-3-mm-3D_glszm_LargeAreaHighGrayLevelEmphasis.

Table 3 compared prediction performance of the established LRs. Among the models, the LR-EGFR generated the lowest AUCs in both training and validation cohorts. **Figure 5** showed ROC curves of the LR-NSCLC, LR-BC, LR-EGFR and LR-HER2.

Radiomics feature analysis

The optimal radiomics features were identified from both handcrafted and deep learning-based features from brain metastasis subregions. For predicting brain metastasis originated from primary NSCLC, seven features were selected, five were handcrafted features and two were deep learning-based features. For predicting the breast cancer origin, five features were selected, two were handcrafted features and three were deep learning-based features. For predicting the EGFR mutation status, six features were selected, five were handcrafted features and one was deep learning feature. For predicting the HER2 status, five features were selected, four were handcrafted features and one was deep learning feature. As shown in **Table 4**, all features were obtained from S1. **Figure 6** depicted results of

the unsupervised cluster analysis of the selected features for all patients, which indicated similarities and differences of features across patients.

Discussion

The development of brain metastasis is a major limitation of life expectancy and quality of life for patients with malignant tumor due to neurological impairments (Quigley et al., 2013). Although early identification of the metastatic origin of brain metastasis is essential to make timely personalized treatment, it is still impossible for clinicians to make the decision by visual inspection of imaging data because there is no specific biomarker. Previous works have explored CAD-based techniques for predicting the primary tumor types (Ortiz-Ramon et al., 2017; Béresová et al., 2018), but both simply performed analysis across the whole tumor area and ignored variations of tumor cell distributions, which was pathobiologically learned as complex heterogeneity of brain metastasis (Siam et al., 2015).

In the present study, we explored a patient- and population-level clustering approach to partition the brain metastasis into two phenotypically consistent subregions, a marginal subregion (S1) and an inner subregion (S2), according to the combination of MRI pixel intensities and local entropy. The S1 has higher MRI pixel intensity and local entropy compared to S2. This was partially consistent with pathological findings that suggested the brain metastasis usually consists of a tumor active area and a necrotic area (Achrol et al., 2019). Before our study, there have been commonly used traditional intratumoral partition techniques, such as Otsu (Farhidzadeh et al., 2015) and k-means (Shang et al., 2021), which are based on the distribution of pixel intensities in the single slice of the tumor image, without considering variations and correlations between different tumor slices and individuals. The clustering approach explored in this study was better at reflecting tumor heterogeneity: (i) the patient-level clustering was applied slice by slice based on the integration of MRI pixel intensities and local entropy, which can better reflect the heterogeneity between different tumor slices; (ii) The population-level clustering was applied based on superpixels gathered from all patients, which can reflect variations and correlations between different individuals.

To reduce overfitting caused by the relatively small sample size, we proposed a RN-GAP net that used GAP to replace the traditional full-connect layers of the ResNet50 net. In the proposed RN-GAP, the Subnet 1 based on ResNet50 can effectively deal with the problem of gradient disappearance and degradation. Besides, the introduced GAP helps reduce the entire height and weight to a single vector by calculating the average of each feature map as an output point, and hence leads to dimensionality reduction without parameter optimization (Lin et al., 2013). Therefore, one advantage of the

TABLE 3 Prediction performance of the established logistic regression models.

LR model	Feature size	Cohort	AUC (95%CI)	ACC	SPE	SEN
LR-NSCLC	7	Training	0.860 (0.775–0.944)	0.796	0.925	0.725
		Validation	0.819 (0.665–0.972)	0.809	0.889	0.800
LR-BC	5	Training	0.909 (0.850–0.969)	0.839	0.925	0.825
		Validation	0.872 (0.771–0.973)	0.787	0.852	0.800
LR-EGFR	6	Training	0.850 (0.735–0.965)	0.700	0.550	1.000
		Validation	0.750 (0.514–0.987)	0.700	0.800	0.800
LR-HER2	5	Training	0.900 (0.800–1.000)	0.850	0.850	0.900
		Validation	0.830 (0.645–1.000)	0.750	0.800	0.800

FIGURE 5
Receiver operating characteristic (ROC) curves of the models for predicting metastatic origins **(A,B)** in the training **(A)** and validation **(B)** cohort, and for predicting the EGFR mutation and HER2 status **(C,D)** in the training **(C)** and validation **(D)** cohort.
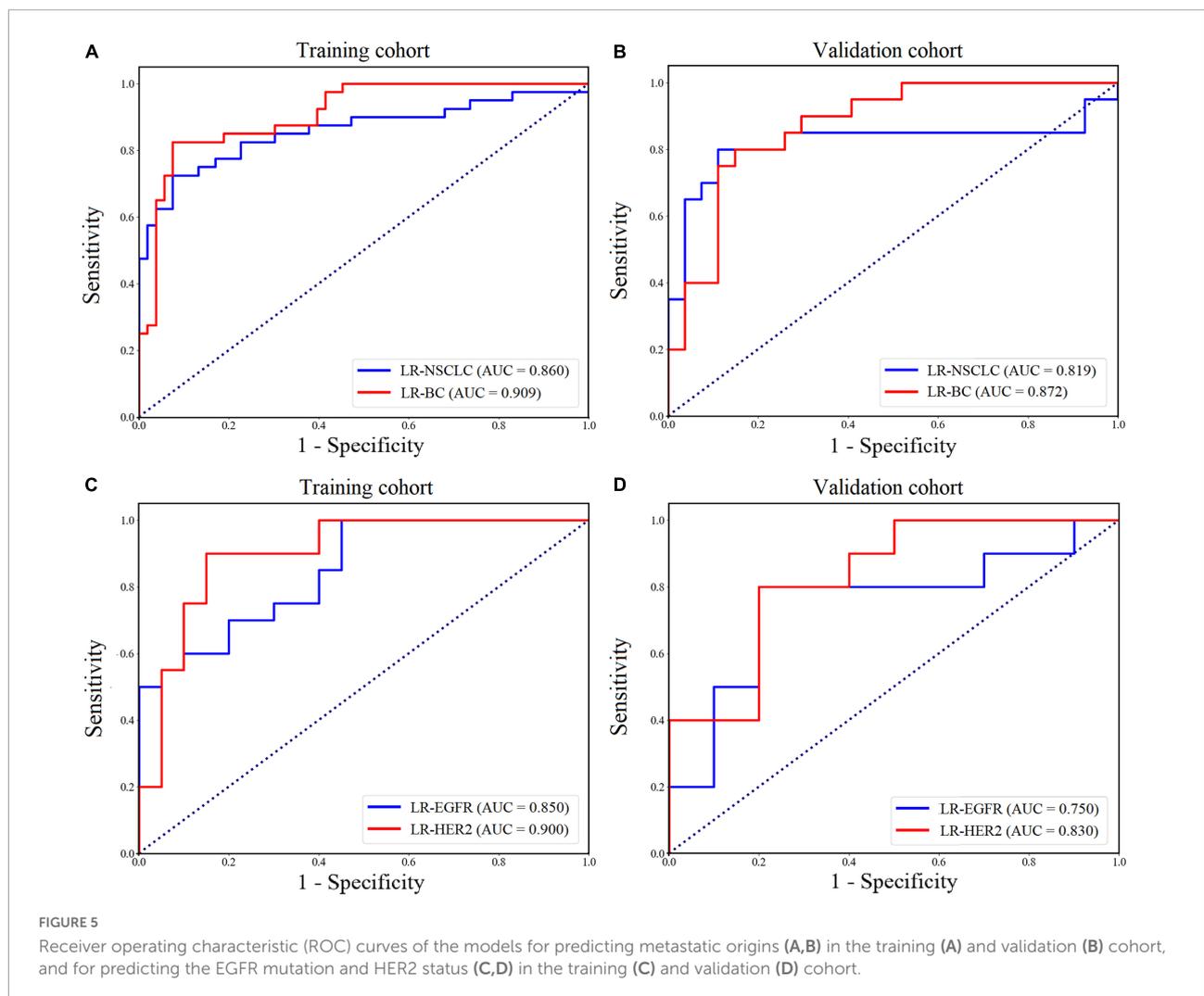
GAP is to avoid overfitting at this layer, and thus improving generalization abilities of the network. Another advantage is that the GAP strengthens the correspondence between feature maps and categories.

By combining both handcrafted and deep learning-based features, the proposed LR-NSCLC, LR-BC, LR-EGFR, and LR-HER2 achieved good prediction performance for predicting metastatic origins and EGFR/HER2 status, respectively. A total of twenty-three features were identified as the most predictive features to develop the four predictive models, sixteen were handcrafted features and seven were deep learning-based features. For predicting NSCLC and Breast cancer originated brain metastasis, our developed LR-NSCLC and LR-BC generated AUCs of 0.819–0.909, which were much higher than

TABLE 4 Prediction performance of the selected features.

Feature	Mean ± SD				AUC	P	ICC(inter−)	ICC(intra−)
	NSCLC	BC	EGFR	HER2				
Wavelet-HLH_glrlm_LongRunEmphasis (F1)	−0.003 ± 0.670	–	–	–	0.594	0.175	0.875	0.864
Wavelet-LHL_glszm_SmallAreaHighGrayLevel Emphasis (F2)	−0.004 ± 0.658	–	–	–	0.753	< 0.001	0.896	0.881
Wavelet-HHH_firstorder_RootMeanSquared (F3)	−0.344 ± 0.642	–	–	–	0767	< 0.001	0.877	0.869
Wavelet-HHH_glrlm_LongRunLowGrayLevel Emphasis (F4)	0.302 ± 0.817	–	–	–	0.617	< 0.134	0.884	0.873
Lbp-3D-k_firstorder_Mean (F5)	0.145 ± 0.689	–	–	–	0.658	0.007	0.868	0.882
DL_215 (F6)	0.035 ± 0.789	–	–	–	0.571	0.106	0.899	0.895
DL_400 (F7)	−0.050 ± 0.896	–	–	–	0.628	0.001	0.852	0.868
Wavelet-HHL_glcm_InverseVariance (F1)	–	0.202 ± 0.987	–	–	0.725	< 0.001	0.872	0.856
Wavelet-HHH_firstorder_InterquartileRange (F2)	–	0.214 ± 0.972	–	–	0.718	< 0.001	0.861	0.858
DL_317 (F3)	–	0.111 ± 1.114	–	–	0.625	0.004	0.863	0.870
DL_144 (F4)	–	0.073 ± 1.079	–	–	0.661	0.017	0.860	0889
DL_122 (F5)	–	0.080 ± 1.001	–	–	0.636	0.010	0.869	0.898
Squareroot_glszm_SmallAreaHighGrayLevel Emphasis (F1)	–		−0.102 ± 0.617	–	0.648	0.008	0.855	0.874
Log-sigma-5-mm.3D_ngtdm_Contrast (F2)	–	–	0.038 ± 0.838	–	0.623	0.110	0.873	0.894
Wavelet-LHL_firstorder_Skewness (F3)	–	–	−0.059 ± 0.869	–	0.580	0.140	0.902	0.862
Lbp-3D-k_firstorder_MeanAbsoluteDeviation (F4)	–	–	0.180 ± 0.977	–	0.602	0.119	0.883	0.896
Lbp-3D-m2_glszm_GrayLevelVariance (F5)	–	–	−0.051 ± 0.879	–	0.655	0.008	0.866	0.907
DL_336 (F6)	–	–	0.213 ± 0.656	–	0.594	0.415	0.911	0.920
Wavelet-LLH_firstorder_10Percentile (F1)	–	–		−0.056 ± 0.891	0.647	0.040	0.857	0.863
Square_ngtdm_Busyness (F2)	–	–	–	−0.028 ± 0.555	0.687	0.006	0.881	0.897
DL_317 (F3)	–	–	–	0.082 ± 0.728	0.668	0.032	0.906	0.913
Wavelet-HHL_glcm_Imc1 (F4)	–	–	–	0.031 ± 0.662	0.664	0.024	0.923	0.927
Log-sigma-3-mm-3D_glszm_LargeAreaHighGrayLevel Emphasis (F5)	–	–	–	0.187 ± 0.837	0.653	0.154	0.893	0.8

a previous study yielded AUCs of 0.63 and 0.61 (Kniep et al., 2019). This may be explainable since the previous study ignored intratumoral heterogeneity of the brain metastasis and lack of introducing deep learning techniques.

For predicting EGFR mutations, our proposed brain metastasis-based LR-EGFR (AUCs of 0.750–0.850) outperformed previously proposed traditional radiomics models based on primary lung cancer (AUCs ranged from 0.575 to 0.762) (Liu et al., 2016; Gevaert et al., 2017; Yuan et al., 2017; Zhang et al., 2018; Digumarthy et al., 2019; Pinheiro et al., 2020) and brain metastasis (AUCs ranged from 0.675 to 0.733) (Wang et al., 2021). This may be because previous studies only applied traditional handcrafted features to develop models. Our results suggested that the RN-GAP net is helpful to identify predictive deep learning-based features that provide complementary information to handcrafted features. For

predicting HER2 status, our LR-HER2 generated AUCs of 0.803–0.900, which were much higher than previous studies based on primary breast cancer and generated AUCs ranging from 0.700 to 0.860 (Li et al., 2021; Niu et al., 2021; Zhou et al., 2021). Our findings suggested information correlated to the HER2 status can also be captured from brain metastasis. The identified most predictive features related to EGFR and HER2 status filtered handcrafted features and deep-learned features, which were hidden in high-dimensional spaces and cannot be identified by radiologists with naked eyes. In clinical practice, many patients whose primary tumors were surgically removed develop brain metastasis, for which cases, a non-invasive method using brain metastasis to early assess the metastatic origins and reflect EGFR or HER2 status is an important alternative for guiding personalized treatment plan (Witzel et al., 2016).

FIGURE 6
Unsupervised cluster analysis of the selected features for predicting the NSCLC originated metastasis **(A)**, breast cancer originated metastasis **(B)**, EGFR mutation status **(C)** and HER2 status **(D)**. The *x*-axis represents the selected features with the highest predictive values. The *y*-axis represents all patients with brain metastasis.
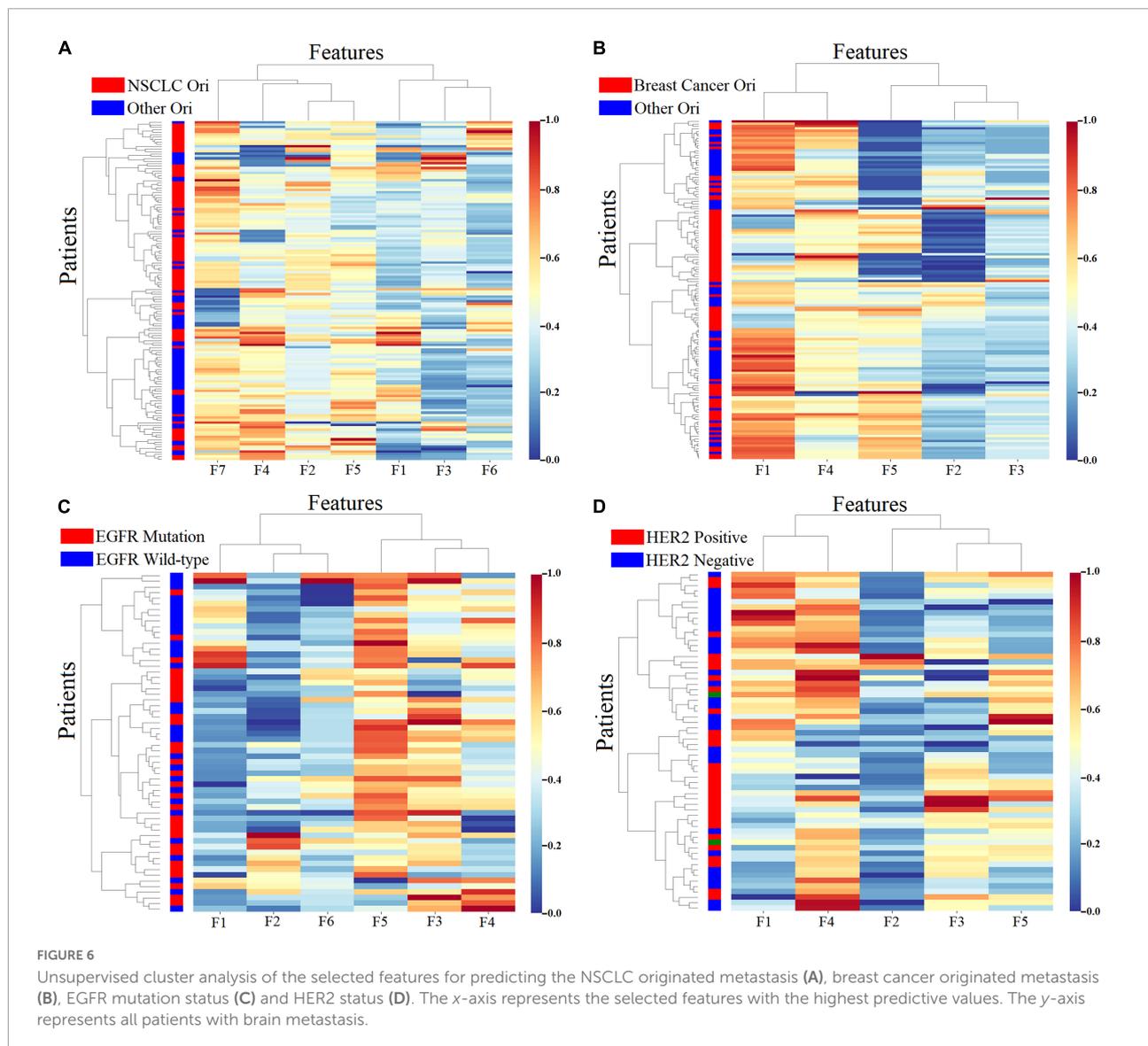
This study has several limitations. First, this is a retrospective study with a relatively small population. Validation of our findings with a larger multi-center sample is needed in our future work. Second, although our intratumoral partition method can effectively divide brain metastasis into various habitats and exhibited good predictive performance, our results can hardly be verified at the pathological level. Besides, there are still no uniform standards for intratumoral partition algorithms, for capturing tumor heterogeneity. Correlations between brain MRI-based morphological differences and corresponding histological microenvironment is warranted, for guiding the improvement of our intratumoral partition method. Third, considering different subtypes of lung cancer and breast cancer have different likelihoods of metastasizing to the brain (Soni et al., 2015), which were not studied in our work. Last, we included only the most common metastatic tumor types of

NSCLC and breast cancer (incidence rates of 56–86% for brain metastasis). Some other metastatic tumor type (e.g., melanoma, incidence rates of 6–11% for brain metastasis) was not included due to the data collection challenge in our hospital, which should be investigated in our future studies.

Conclusion

Our work indicated that subregional radiomics is of significant to predict metastatic origins and assess gene status based on brain metastasis. Our proposed handcrafted and deep learning feature combined logistic regression model may be clinically valuable to assist in early personalized treatment plan decisions.

Data availability statement

The original contributions presented in this study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

JS, ZZ, HF, and XJ contributed to conception and design of the study. JL and YL organized the database. TJ and HA performed the statistical analysis. ZZ and XJ wrote the first draft of the manuscript. JS and XC wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2022.973698/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Different CH values when the clusters k was set from 2 to 10.

References

Achrol, A. S., Rennert, R. C., Anders, C., Soffietti, R., Ahluwalia, M. S., Nayak, L., et al. (2019). Brain metastases. *Nat. Rev. Dis. Primers* 5:5. doi: 10.1038/s41572-018-0055-y

Barragán-Montero, A., Javaid, U., Valdés, G., Nguyen, D., Desbordes, P., Macq, B., et al. (2021). Artificial intelligence and machine learning for medical imaging: A technology review. *Phys. Med.* 83, 242–256. doi: 10.1016/j.ejmp.2021.04.016

Bekaert, L., Emery, E., Levallet, G., and Lechapt-Zalcman, E. (2017). Histopathologic diagnosis of brain metastases: Current trends in management and future considerations. *Brain Tumor Pathol.* 34, 8–19. doi: 10.1007/s10014-016-0275-3

Béresová, M., Larroza, A., Arana, E., Varga, J., Balkay, L., and Moratal, D. (2018). 2D and 3D texture analysis to differentiate brain metastases on MR images: Proceed with caution. *MAGMA* 31, 285–294. doi: 10.1007/s10334-017-0653-9

Brastianos, H. C., Cahill, D. P., and Brastianos, P. K. (2015). Systemic therapy of brain metastases. *Curr. Neurol. Neurosci. Rep.* 15:518. doi: 10.1007/s11910-014-0518-9

Brito, C., Machado, A., and Sousa, A. (2019). Electrocardiogram Beat-Classification Based on a ResNet Network. *Stud. Health Technol. Inform.* 264, 55–59. doi: 10.3233/SHTI190182

Cagney, D. N., Martin, A. M., Catalano, P. J., Redig, A. J., Lin, N. U., Lee, E. Q., et al. (2017). Incidence and prognosis of patients with brain metastases at diagnosis of systemic malignancy: A population-based study. *Neuro Oncol.* 19, 1511–1521. doi: 10.1093/neuonc/nox077

Caliński, T., and Harabasz, J. (1974). A dendrite method for cluster analysis. *Commun. Stat.* 3, 1–27. doi: 10.1080/03610927408827101

Cameron, D., Piccart-Gebhart, M. J., Gelber, R. D., Procter, M., Goldhirsch, A., de Azambuja, E., et al. (2017). 11 years' follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive early breast cancer: Final analysis of the

HERceptin Adjuvant (HERA) trial. *Lancet* 389, 1195–1205. doi: 10.1016/S0140-6736(16)32616-2

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: a large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition.* (Miami: IEEE), 248–255. doi: 10.1109/CVPR.2009.5206848

Digumarthy, S. R., Padole, A. M., Gullo, R. L., Sequist, L. V., and Kalra, M. K. (2019). Can CT radiomic analysis in NSCLC predict histology and EGFR mutation status? *Medicine* 98:e13963. doi: 10.1097/MD.0000000000013963

Fan, M., Cheng, H., Zhang, P., Gao, X., Zhang, J., Shao, G., et al. (2018). DCE-MRI texture analysis with tumor subregion partitioning for predicting Ki-67 status of estrogen receptor-positive breast cancers. *J. Magn. Reson. Imaging* 48, 237–247. doi: 10.1002/jmri.25921

Farhidzadeh, H., Goldgof, D. B., Hall, L. O., Gatenby, R. A., Gillies, R. J., and Raghavan, M. (2015). Texture feature analysis to predict metastatic and necrotic soft tissue sarcomas. *2015 IEEE International Conference on Systems.* (Miami: IEEE), 2798–2802. doi: 10.1109/SMC.2015.488

Fehrenbacher, L., Cecchini, R. S., Geyer, C. E. Jr., Rastogi, P., Costantino, J. P., Atkins, J. N., et al. (2020). NSABP B-47/NRG Oncology Phase III Randomized Trial Comparing Adjuvant Chemotherapy With or Without Trastuzumab in High-Risk Invasive Breast Cancer Negative for HER2 by FISH and With IHC 1+ or 2. *J. Clin. Oncol.* 38, 444–453. doi: 10.1200/JCO.19.01455

Gatenby, R. A., Grove, O., and Gillies, R. J. (2013). Quantitative imaging in cancer evolution and ecology. *Radiology* 269, 8–15. doi: 10.1148/radiol.13122697

Gevaert, O., Echegaray, S., Khuong, A., Hoang, C. D., Shrager, J. B., Jensen, K. C., et al. (2017). Predictive radiogenomics modeling of EGFR mutation status in lung cancer. *Sci. Rep.* 7:41674. doi: 10.1038/srep41674

Gore, J. C. (2020). Artificial intelligence in medical imaging. *Magn. Reson. Imaging* 68:A1–A4. doi: 10.1016/j.mri.2019.12.006

Guo, Y., Song, J., Wang, Y., Huang, L., Sun, L., Zhao, J., et al. (2020). Concurrent Genetic Alterations and Other Biomarkers Predict Treatment Efficacy of EGFR-TKIs in EGFR-Mutant Non-Small Cell Lung Cancer: A Review. *Front. Oncol.* 10:610923. doi: 10.3389/fonc.2020.610923

Hardesty, D. A., and Nakaji, P. (2016). The Current and Future Treatment of Brain Metastases. *Front. Surg.* 3:30. doi: 10.3389/fsurg.2016.00030

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika* 32, 241–254. doi: 10.1007/BF02289588

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., and Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 881–892. doi: 10.1109/TPAMI.2002.1017616

Kniep, H. C., Madesta, F., Schneider, T., Hanning, U., Schönfeld, M. H., Schön, G., et al. (2019). Radiomics of Brain MRI: Utility in Prediction of Metastatic Tumor Type. *Radiology* 290, 479–487. doi: 10.1148/radiol.2018180946

Koboldt, D. C., Fulton, R. S., McLellan, M. D., Schmidt, H., Kalicki-Veizer, J., McMichael, J. F., et al. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70. doi: 10.1038/nature11412

Krawczyk, P., Nicoś, M., Ramlau, R., Powrózek, T., Wojas-Krawczyk, K., Sura, S., et al. (2014). The incidence of EGFR-activating mutations in bone metastases of lung adenocarcinoma. *Pathol. Oncol. Res.* 20, 107–112. doi: 10.1007/s12253-013-9667-4

Lam, S. W., Jimenez, C. R., and Boven, E. (2014). Breast cancer classification by proteomic technologies: Current state of knowledge. *Cancer Treat. Rev.* 40, 129–138. doi: 10.1016/j.ctrv.2013.06.006

Lambin, P., Leijenaar, R. T. H., Deist, T. M., Peerlings, J., de Jong, E. E. C., van Timmeren, J., et al. (2017). Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* 14, 749–762. doi: 10.1038/nrclinonc.2017.141

Leijenaar, R. T., Carvalho, S., Velazquez, E. R., van Elmpt, W. J., Parmar, C., Hoekstra, O. S., et al. (2013). Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and inter-observer variability. *Acta Oncol.* 52, 1391–1397. doi: 10.3109/0284186X.2013.812798

Li, C., Song, L., and Yin, J. (2021). Intratumoral and Peritumoral Radiomics Based on Functional Parametric Maps from Breast DCE-MRI for Prediction of HER-2 and Ki-67 Status. *J. Magn. Reson. Imaging* 54, 703–714. doi: 10.1002/jmri.27651

Lin, M., Chen, Q., and Yan, S. (2013). Network in network. *arXiv.* [Preprint]. doi: 10.48550/arXiv.1312.4400

Lin, M., Wynne, J. F., Zhou, B., Wang, T., Lei, Y., Curran, W. J., et al. (2021). Artificial intelligence in tumor subregion analysis based on medical imaging: A review. *J. Appl. Clin. Med. Phys.* 22, 10–26. doi: 10.1002/acm2.13321

Liu, Y., Kim, J., Balagurunathan, Y., Li, Q., Garcia, A. L., Stringfield, O., et al. (2016). Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. *Clin. Lung Cancer* 17, 441–448. doi: 10.1016/j.cllc.2016.02.001

Loibl, S., and Gianni, L. (2017). HER2-positive breast cancer. *Lancet* 389, 2415–2429. doi: 10.1016/S0140-6736(16)32417-5

Niu, S., Jiang, W., Zhao, N., Jiang, T., Dong, Y., Luo, Y., et al. (2021). Intra- and peritumoral radiomics on assessment of breast cancer molecular subtypes based on mammography and MRI. *J. Cancer Res. Clin. Oncol.* 148, 97–106. doi: 10.1007/s00432-021-03822-0

O'Connor, J. P., Rose, C. J., Waterton, J. C., Carano, R. A., Parker, G. J., and Jackson, A. (2015). Imaging intratumor heterogeneity: Role in therapy response, resistance, and clinical outcome. *Clin. Cancer Res.* 21, 249–257. doi: 10.1158/1078-0432.CCR-14-0990

Ortiz-Ramon, R., Larroza, A., Arana, E., and Moratal, D. (2017). A radiomics evaluation of 2D and 3D MRI texture features to classify brain metastases from lung cancer and melanoma. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2017, 493–496. doi: 10.1109/EMBC.2017.8036869

Pham, D. T., Dimov, S. S., and Nguyen, C. D. (2005). Selection of K in K-means clustering. *Proc. Inst. Mech. Eng. Part C* 219, 103–119. doi: 10.1243/095440605X8298

Pinheiro, G., Pereira, T., Dias, C., Freitas, C., Hespanhol, V., Costa, J. L., et al. (2020). Identifying relationships between imaging phenotypes and lung cancer-related mutation status: EGFR and KRAS. *Sci. Rep.* 10:3625. doi: 10.1038/s41598-020-60202-3

Quigley, M. R., Fukui, O., Chew, B., Bhatia, S., and Karlovits, S. (2013). The shifting landscape of metastatic breast cancer to the CNS. *Neurosurg. Rev.* 36, 377–382. doi: 10.1007/s10143-012-0446-6

Ruan, S., Liu, Y., Tang, X., Yang, Z., Huang, J., Li, X., et al. (2017). HER-2 status and its clinicopathologic significance in breast cancer in patients from southwest China: Re-evaluation of correlation between results from FISH and IHC. *Int. J. Clin. Exp. Pathol.* 10, 7270–7276.

Shang, S., Sun, J., Yue, Z., Wang, Y., Wang, X., Luo, Y., et al. (2021). Multi-parametric MRI based radiomics with tumor subregion partitioning for differentiating benign and malignant soft-tissue tumors. *Biomed. Signal Process. Control* 67:102522. doi: 10.1016/j.bspc.2021.102522

Shen, H., Chen, L., Liu, K., Zhao, K., Li, J., Yu, L., et al. (2021). A subregion-based positron emission tomography/computed tomography (PET/CT) radiomics model for the classification of non-small cell lung cancer histopathological subtypes. *Quant. Imaging Med. Surg.* 11, 2918–2932. doi: 10.21037/qims-20-1182

Siam, L., Bleckmann, A., Chaung, H. N., Mohr, A., Klemm, F., Barrantes-Freer, A., et al. (2015). The metastatic infiltration at the metastasis/brain parenchyma-interface is very heterogeneous and has a significant impact on survival in a prospective study. *Oncotarget* 6, 29254–29267. doi: 10.18632/oncotarget.4201

Soni, A., Ren, Z., Hameed, O., Chanda, D., Morgan, C. J., Siegal, G. P., et al. (2015). Breast cancer subtypes predispose the site of distant metastases. *Am. J. Clin. Pathol.* 143, 471–478. doi: 10.1309/AJCPYO5FSV3UPEXS

van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., et al. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 77:e104–e107. doi: 10.1158/0008-5472.CAN-17-0339

Wang, G., Wang, B., Wang, Z., Li, W., Xiu, J., Liu, Z., et al. (2021). Radiomics signature of brain metastasis: Prediction of EGFR mutation status. *Eur. Radiol.* 31, 4538–4547. doi: 10.1007/s00330-020-07614-x

Witzel, I., Oliveira-Ferrer, L., Pantel, K., Müller, V., and Wikman, H. (2016). Breast cancer brain metastases: Biology and new clinical perspectives. *Breast Cancer Res.* 18:8. doi: 10.1186/s13058-015-0665-1

Wu, J., Gensheimer, M. F., Dong, X., Rubin, D. L., Napel, S., Diehn, M., et al. (2016). Robust Intratumor Partitioning to Identify High-Risk Subregions in Lung Cancer: A Pilot Study. *Int. J. Radiat. Oncol. Biol. Phys.* 95, 1504–1512. doi: 10.1016/j.ijrobp.2016.03.018

Xie, C., Yang, P., Zhang, X., Xu, L., Wang, X., Li, X., et al. (2019). Sub-region based radiomics analysis for survival prediction in oesophageal tumours treated by definitive concurrent chemoradiotherapy. *EBioMedicine* 44, 289–297. doi: 10.1016/j.ebiom.2019.05.023

Yuan, M., Pu, X. H., Xu, X. Q., Zhang, Y. D., Zhong, Y., Li, H., et al. (2017). Lung adenocarcinoma: Assessment of epidermal growth factor receptor mutation status based on extended models of diffusion-weighted image. *J. Magn. Reson. Imaging* 46, 281–289. doi: 10.1002/jmri.25572

Zhang, L., Chen, B., Liu, X., Song, J., Fang, M., Hu, C., et al. (2018). Quantitative Biomarkers for Prediction of Epidermal Growth Factor Receptor Mutation in Non-Small Cell Lung Cancer. *Transl. Oncol.* 11, 94–101. doi: 10.1016/j.tranon.2017.10.012

Zhou, J., Tan, H., Li, W., Liu, Z., Wu, Y., Bai, Y., et al. (2021). Radiomics Signatures Based on Multiparametric MRI for the Preoperative Prediction of the HER2 Status of Patients with Breast Cancer. *Acad. Radiol.* 28, 1352–1360. doi: 10.1016/j.acra.2020.05.040