Physical and Engineering Sciences in Medicine (2023) 46:1309–1320
https://doi.org/10.1007/s13246-023-01300-0

SCIENTIFIC PAPER

MRI radiomics for brain metastasis sub-pathology classification from non-small cell lung cancer: a machine learning, multicenter study

Fuxing Deng[1,2,3] · Zhiyuan Liu[1,2,3] · Wei Fang[4] · Lishui Niu[1,2,3] · Xianjing Chu[1,2,3] · Quan Cheng[3,5] · Zijian Zhang[1,2,3] · Rongrong Zhou[1,2,3] · Guang Yang[6,7,8]

Received: 23 March 2023 / Accepted: 4 July 2023 / Published online: 17 July 2023

Abstract

The objective of this study is to develop a machine-learning model that can accurately distinguish between different histologic types of brain lesions in patients with non-small cell lung cancer (NSCLC) when it is not safe or feasible to perform a biopsy. To achieve this goal, the study utilized data from two patient cohorts: 116 patients from Xiangya Hospital and 35 patients from Yueyang Central Hospital. A total of eight machine learning algorithms, including Xgboost, were compared. Additionally, a 3-dimensional convolutional neural network was trained using transfer learning to further evaluate the performance of these models. The SHapley Additive exPlanations (SHAP) method was developed to determine the most important features in the best-performing model after hyperparameter optimization. The results showed that the area under the curve (AUC) for the classification of brain lesions as either lung adenocarcinoma or squamous carcinoma ranged from 0.60 to 0.87. The model based on single radiomics features extracted from contrast-enhanced T1 MRI and utilizing the Xgboost algorithm demonstrated the highest performance (AUC: 0.85) in the internal validation set and adequate performance (AUC: 0.80) in the independent external validation set. The SHAP values also revealed the impact of individual features on the classification results. In conclusion, the use of a radiomics model incorporating contrast-enhanced T1 MRI, Xgboost, and SHAP algorithms shows promise in accurately and interpretably identifying brain lesions in patients with NSCLC.

Keywords Magnetic resonance imaging · Brain metastasis · Radiotherapy · Radiomics · Explain ability · Machine learning

Zijian Zhang and Rongrong Zhou contributed equally as co-senior authors.

✉ Zijian Zhang
wanzzj@csu.edu.cn

✉ Rongrong Zhou
zhourr@csu.edu.cn

[1] Department of Oncology, Xiangya Hospital, Central South University, Changsha 410008, China

[2] National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, Hunan Province 410008, China

[3] Xiangya Lung Cancer Center, Xiangya Hospital, Central South University, Changsha 410008, China

[4] Department of Radiology, Yueyang Central Hospital, Yueyang 414000, China

[5] Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha 410008, China

[6] National Heart and Lung Institute, Imperial College London, London, UK

[7] Cardiovascular Research Centre, Royal Brompton Hospital, London, UK

[8] School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK

Introduction

Brain metastases represent a significant and growing public health challenge, with an estimated 10–40% of solid tumors patients developing metastases to the brain over the course of their disease [1]. The median survival time for patients with brain metastases is 12 months or less, according to prognostic estimates based on the SEER database [2]. The primary origin of cerebral metastases is frequently observed to be from advanced lung carcinoma, presenting with a varied range of symptoms depending on the magnitude, placement, and frequency of these lesions [3]. NSCLC can be classified as adenocarcinoma (ADC), squamous cell carcinoma (SCC), large cell carcinoma (LCC), and other relatively rare subtypes [4]. The orientation of the brain in NSCLC varies according to histological subtype [5, 6]. Recently, developments in immunotherapy and targeted therapy have shown improved overall survival in BM patients in subgroups selected by primary tumor molecular characteristics [7–9]. The combination of immunotherapy with radiation therapy for different types of pathology can effectively reduce the risk of radiation toxicity, such as radiation necrosis [10]. Therefore, the identification of the primary lesion and its pathological types in patients with BM is of great clinical importance. Biopsy is the standard method for histological classification and genetic status confirmation, but it is an invasive procedure that can result in complications such as bleeding and infection, and it may only represent a small portion of the tumor or a single lesion, making it difficult to fully reflect the tumor's characteristics [11]. The study explores the establishment of a pathological subtype-based machine-learning model of BMs in NSCLC, which lays a foundation for the further prediction of cancer driver gene expression based on patient radiomics data.

"Radiomics" refers to the extraction of many quantitative features from standard clinical imaging data, and then convert these to higher-dimensional mineable data [12, 13]. Several studies have demonstrated the potential to classify histological subtypes and genetic status, using radiomics approaches based on medical images [14–17]. Also, some studies have evaluated the relationship between radiomics features or deep learning and brain pathology based on head nuclear magnetic resonance imaging (MRI) [18–20]. Liu et al. conducted a study that developed a model to classify glioblastoma, the most common primary brain tumor in adults, and solitary brain metastasis [21]. Although Kniep's study model can distinguish between these five different pathology sources, the classification accuracy of lung cancer pathological subtypes was no more than 75% [22]. The model combined features from deep learning and radiomics features can differentiate brain metastases from other brain tumors, but they focused on the single tumor and did not differentiate between different sources of NSCLC [23].

However, few studies have investigated the application of radiomics in metastatic lesions and as with many machine learning-derived correlation methods, their interpretability of the model-based radiomics features is often a challenge. Standard machine learning, which comprises artificial intelligence is black-box predictions that cannot readily be interpreted by clinicians. To our best knowledge, there is currently no established model for accurately predicting the histological type of brain metastatic lesions from lung cancer based on head MRI [22]. Several interpretability methods in machine learning can be used to understand and explain the predictions made by a model, such as Local interpretable model-agnostic explanations (LIME) [24], Partial dependence plots (PDP) [25], and SHapley Additive exPlanations (SHAP) [26]. The LIME method is only locally interpretable. This means that it can only provide explanations for individual predictions, rather than for the model as a whole. One disadvantage of PDP is that they can only show the relationship between a single feature and the model's predictions. It can be difficult to understand the relationship between multiple features and the model's predictions using PDPs alone. To overcome these limitations, the calculation of the Shapley value based on sampling may help doctors to understand the importance of individual features in all or single samples. The SHAP values are a perturbation-based feature attribution method that borrows from the game theory. The SHAP values are the average of the marginal contributions across all permutations, which can disclose the individual contribution of each feature on the output of the model for each observation or example [27].

Hence, in this paper, based on MRI, we exploratively developed a radiomics model with new methods to identify the histological types of brain metastatic lesions from NSCLC. In particular, in order to interpret the classifiers, we compute global and individual SHAP values of the radiomics features and assess how these features differentiate the subtypes.

Methods

Patient and data source

Patients with brain lesions from non-small cell lung cancer were collocated from January 2016 to December 2020 at the electrical health record system in Xiangya Hospital. For external validation, MRI data were extracted from Yueyang Central Hospital. Pathological diagnoses were confirmed using a pathology report in the electronic health record system. The selection criteria and exclusions were pre-defined.

Table 1 Detailed inclusion criteria

Criteria	Description
Age	≥ 18 years
Pathology	Histopathological confirmation of lung adenocarcinoma or lung squamous cell carcinoma
Size	> 3 mm
Location	No meningeal metastasis or other brain tumors
MRI Matrix	> 114 × 114
Therapy	No prior radiation therapy on the brain

Eligibility criteria consisting of inclusion criteria are in Table 1.

Magnetic resonance imaging (MRI) was performed using a 3-Tesla or 1.5-Tesla horizontal system (Siemens). Magnetic resonance spectroscopy was conducted before treatment in patients. The MRI protocol consisted of three parts: an anatomical MRI protocol for the clinical diagnosis of brain tumors, including a high-resolution T1-weighted sequence, a contrast-enhanced T1-weighted sequence, and a T2-weighted sequence. The study was approved by the hospital ethics committee, so patient informed consent was waived.

Data preprocessing

Manual segmentation was performed in Insight Toolkit (ITK-SNAP 3.8.0) which is freely available (http://www.itksnap.org/pmwiki/pmwiki.php). The brain lesions regions of interest (ROI) in each slice were manually outlined in the ITK-SNAP to include tumor tissue and infiltrating peritumoral edema, with the exclusion of cysts and necrosis areas, and verified by 2 neuroradiologists. Disagreements were resolved by discussion until agreement. Both readers were blinded to the histopathological diagnosis of the tumor.

Preprocessing was conducted using the Simple ITK package in python 3.6 [28]. Firstly, all MRI images (T1, T1C, and T2 images) were corrected for gradient nonlinearity, and intensity was corrected using the N4 bias field correction, which can also remove all of the intensity nonuniformity with a low frequency. After bias field correction, those images were resampled at a voxel size of 0.375 × 0.375 × 6.5.

Model building and validation

Machine learning with radiomics

Each unique tumor segmentation was calculated using PyRadiomics (Version 3.0.1) in python 3.6 [29]. Mathematical formulas are described on the website of the package (https://pyradiomics.readthedocs.io). Procedures and features were in accordance with the Imaging Biomarker Standardization Initiative (IBSI) [30], such as the description of the patient, acquisition of the MRI data, processing of the images, and machine learning and radiomics analysis. There were 1137 features extracted from MRI in each modality. These kinds of features were given in Table 2. A detailed description of all radiomics features was provided on the Pyradiomics webpage (https://pyradiomics.readthedocs.io/en/latest/features.html#). Normalization procedures were applied to all radiomic features to eliminate interdimensional variability. The chosen normalization strategy was the z-score normalization technique, which involved

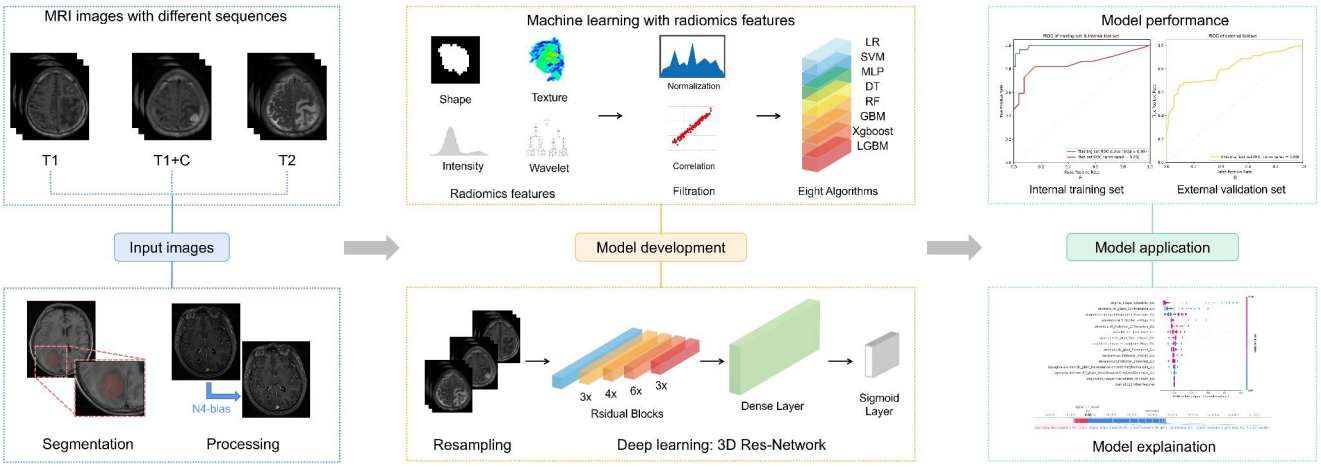

Fig. 1 The workflow illustrates development models to classify the pathology of NSCLC brain lesions. Three modality MRI images (T1, T1C, and T2) from NSCLC brain lesions patients were inputted. On the one hand, optimal features, after normalization, were selected via correlation analysis and used to construct models by eight algorithms. On the other hand, images after resampling were input into the 3-dimensional Residual Neural Network. The performances of all models were compared by AUROC. Also, test the performance of the best-performing classifiers in internal and external sets. The features of the best model were further explained by the SHAP. Abbreviation: LR, Logistic Regression; SVM, Support Vector Machine; MLP, Multilayer Perceptron; DT, Decision Tree; RF, Random Forest; GBM, gradient boosting; XgBoost, extrema gradient boosting; LGBM, light gradient boosting

Table 2 Extraction methods and feature quantification of Radiomics features

Feature class	Features	Feature class	Features
GLSZM	GrayLevelNonUniformity	GLRLM	GrayLevelNonUniformity
	GrayLevelNonUniformityNormalized		GrayLevelNonUniformityNormalized
	GrayLevelVariance		GrayLevelVariance
	HighGrayLevelZoneEmphasis		HighGrayLevelRunEmphasis
	LargeAreaEmphasis		LongRunEmphasis
	LargeAreaHighGrayLevelEmphasis		LongRunHighGrayLevelEmphasis
	LargeAreaLowGrayLevelEmphasis		LongRunLowGrayLevelEmphasis
	LowGrayLevelZoneEmphasis		LowGrayLevelRunEmphasis
	SizeZoneNonUniformity		RunEntropy
	SizeZoneNonUniformityNormalized		RunLengthNonUniformity
	SmallAreaEmphasis		RunLengthNonUniformityNormalized
	SmallAreaHighGrayLevelEmphasis		RunPercentage
	SmallAreaLowGrayLevelEmphasis		RunVariance
	ZoneEntropy		ShortRunEmphasis
	Zone%		ShortRunHighGrayLevelEmphasis
	ZoneVariance		ShortRunLowGrayLevelEmphasis
Shape	Elongation	GLDM	DependenceEntropy
	Flatness		DependenceNonUniformity
	LeastAxisLength		DependenceNonUniformityNormalized
	MajorAxisLength		DependenceVariance
	Maximum2DDiameterColumn		GrayLevelNonUniformity
	Maximum2DDiameterRow		GrayLevelVariance
	Maximum2DDiameterSlice		HighGrayLevelEmphasis
	Maximum3DDiameter		LargeDependenceEmphasis
	MeshVolume		LargeDependenceHighGrayLevelEmphasis
	MinorAxisLength		LargeDependenceLowGrayLevelEmphasis
	Sphericity		LowGrayLevelEmphasis
	SurfaceArea		SmallDependenceEmphasis
	SurfaceVolumeRatio		SmallDependenceHighGrayLevelEmphasis
	VoxelVolume		SmallDependenceLowGrayLevelEmphasis
First-order	10Percentile	GLCM	Autocorrelation
	90Percentile		ClusterProminence
	Energy		ClusterShade
	Entropy		ClusterTendency
	InterquartileRange		Contrast
	Kurtosis		Correlation
	Maximum		DifferenceAverage
	MeanAbsoluteDeviation		DifferenceEntropy
	Mean		DifferenceVariance
	Median		Id
	Minimum		Idm
	Range		Idmn
	RobustMeanAbsoluteDeviation		Idn
	RootMeanSquared		Imc1/2
	Skewness		InverseVariance
	TotalEnergy		JointAverage
	Uniformity		JointEnergy
	Variance		JointEntropy

Table 2 (continued)

Feature class	Features	Feature class	Features
NGTDM	Busyness		MCC
	Coarseness		MaximumProbability
	Complexity		SumAverage
	Contrast		SumEntropy
	Strength		SumSquares

Abbreviation: First-order: First-order statistics describe the distribution of voxel intensities within the image region defined by the mask through commonly used and basic metrics. GLSZM: A Gray Level Size Zone (GLSZM) quantifies gray level zones in an image. A gray level zone is defined as a the number of connected voxels that share the same gray level intensity. A voxel is considered connected if the distance is 1 according to the infinity norm (26-connected region in a 3D, 8-connected region in 2D). NGTDM: A Neighboring Gray Tone Difference Matrix quantifies the difference between a gray value and the average gray value of its neighbors within distance δ. Shape: In this group of features we included descriptors of the two-dimensional size and shape of the ROI. These features are independent from the gray level intensity distribution in the ROI and are therefore only calculated on the non-derived image and mask. GLDM: A Gray Level Dependence Matrix (GLDM) quantifies gray level dependencies in an image. A gray level dependency is defined as a the number of connected voxels within distance δ that are dependent on the center voxel. GLRLM: A Gray Level Run Length Matrix (GLRLM) quantifies gray level runs, which are defined as the length in number of pixels, of consecutive pixels that have the same gray level value. GLCM: A Gray Level Co-occurrence Matrix (GLCM) of size Ng × Ng describes the second-order joint probability function of an image region constrained by the mask and is defined as P(i,j|δ,θ). Seeing https://pyradiomics.readthedocs.io/en/latest/features.html# for the detailed formula for each feature among various features classes.

transforming the data so that the resulting mean was zero and the standard deviation was one. To reduce the collinearity of the extracted features and remove redundant information to avoid the overfitting issue, the collinearity and redundant feature were removed from each modality MRI image by calculating the 'Spearman' correlation coefficient [31]. For features with correlation coefficients greater than 0.9, one of them is retained to build the model. Based on selected radiomics features, eight popular machine learning methods were utilized to build the predictive models. These methods included Generalized Linear Model (GLM) [32], Random Forest [33], Support Vector Machine (SVM) [34], Decision Tree [35], Gradient Boosting Machine (GBM) [36], Multi-Layer Perceptron (MLP) [37], eXtrema Gradient Boosting (Xgboost) [38], Light Gradient Boosting Machine (LGBM) [39]. In order to evaluate the model's stability, a cross-validation procedure technique was used, and 5-fold cross-validation was performed. The discrimination ability of the model was assessed with the area under the curve (AUC) value. In addition, the best model among the models, after using the grad search to tune parameters was presented in a confusion matrix, accuracy, sensitivity, specificity, and the F-score. Model constructions were performed on the Python platform (python version 3.6) [38–40]. Based on the tree-like model, feature importance ranking will be used to display important features by Shapley additive explanations (SHAP) values, which are used to aid in the interpretation of the performance. SHAP values are computed and visualized with the SHAP python package (https://shap.readthedocs.io/en/latest/index.html).

Deep learning without radiomics

Recent advances using deep convolutional neural networks allow automatic feature extraction and end-to-end machine learning models, so we explored whether we could directly use delineation data to calcific pathological subtypes of NSCLC in an end-to-end way. This study is one step less than the extraction of radiomics features through machine learning, and the deep learning network was directly utilized to extract both surface and deep features from MRI scans. The 3D Res-Net-34 end-to-end model was built using the PyTorch framework and transfer learning. The channel is set to 3, corresponding to three different MRI modes. The delineated tumor area was used as the region of interest (ROI) to extract the whole tumor and the ROI was then normalized to $2\times112\times112$. A transfer learning approach was applied using the pretrained Res-Net CNN that was trained on the medical images by Tencent (https://github.com/Tencent/MedicalNet). The model utilized the Adam adaptive learning rate method, with a learning rate set to 1e-5. The batch size for the training process was set to 8. The cross-entropy function was utilized as the loss function. The model is trained for 300 epochs. The flow chart of our analyses is depicted in Fig. 1. The development model code and radiomics features are open source and available on GitHub (https://github.com/BboyT/BM_NSCLC_subpathology).

Results

Baseline characteristics

In this study, 116 patients with brain lesions from lung cancer who were examined by MRI using the research protocol were searched at Xiangya Hospital from January 2018 to December 2020. The number of patients with a single brain lesion is 98 and the number of patients with multiple brain lesions is 18 in Xiangya hospital. The dataset from Xiangya hospital (adenocarcinoma lesions: 96, squamous cell

carcinoma lesions:50) was split into a training set with 70% of the patients (81 patients, 98 lesions) and an internal test set of size 30% (35 patients 48 lesions). In the external test set, 35 patients with 116 brain lesions from non-small cell lung cancer MRI data were extracted from Yueyang Central Hospital between January 2016 and December 2021. The number of patients with a single brain lesion is 12 and the number of patients with multiple brain lesions is 23 at Yueyang Central hospital. A total of 261 brain lesions included from untreated 151 brain tumor patients, were selected for training in the ML classifier and validation queue. The maximum tumor diameters were available for all patients, and the mean tumor diameters were 23.1 mm (range, 14.5 – 28.8 mm in Xiangya hospital) and 17.3 mm (range: 10.0 – 21.5 mm, in Yueyang central hospital) (Fig. 2).

The selected Radiomic features

From the 1137 features that were extracted from each MRI data set, respectively (Table 2). After data dimensional reduction, t1 has 281 remaining features, t2 has 264 remaining features, and t1 has contrast-enhanced the remaining 225 features.

ML Classifiers in the training and validation cohort

The ROC of the ML classification model based on the best modality MRIs is presented in Fig. 3. See Figures in online appendices [Supplementary Information (SI)] for ROC plots of other modalities and union MRIs. The best-performing model overall, however, is the model trained on a single modality contrast-enhanced T1 MRI. The eight resulting models used contrasted-enhanced T1 MRI demonstrate in Table 3. Figure 3; Table 3 demonstrate that the performance of distinguishing adenocarcinoma and squamous cell carcinoma in non-small cell lung cancer using contrasted-enhanced T1 MRI radiomics feature based Xgboost algorithm (AUC: 0.87) has the highest performance. Therefore, we further evaluation to the Xgboost model after optimizing hyperparameter values using the grid search method. Figure 4 A. shows the AUROC of Xgboost (0.85 in the internal test set; 0.99 in the training set), and accuracy is 81.25%;

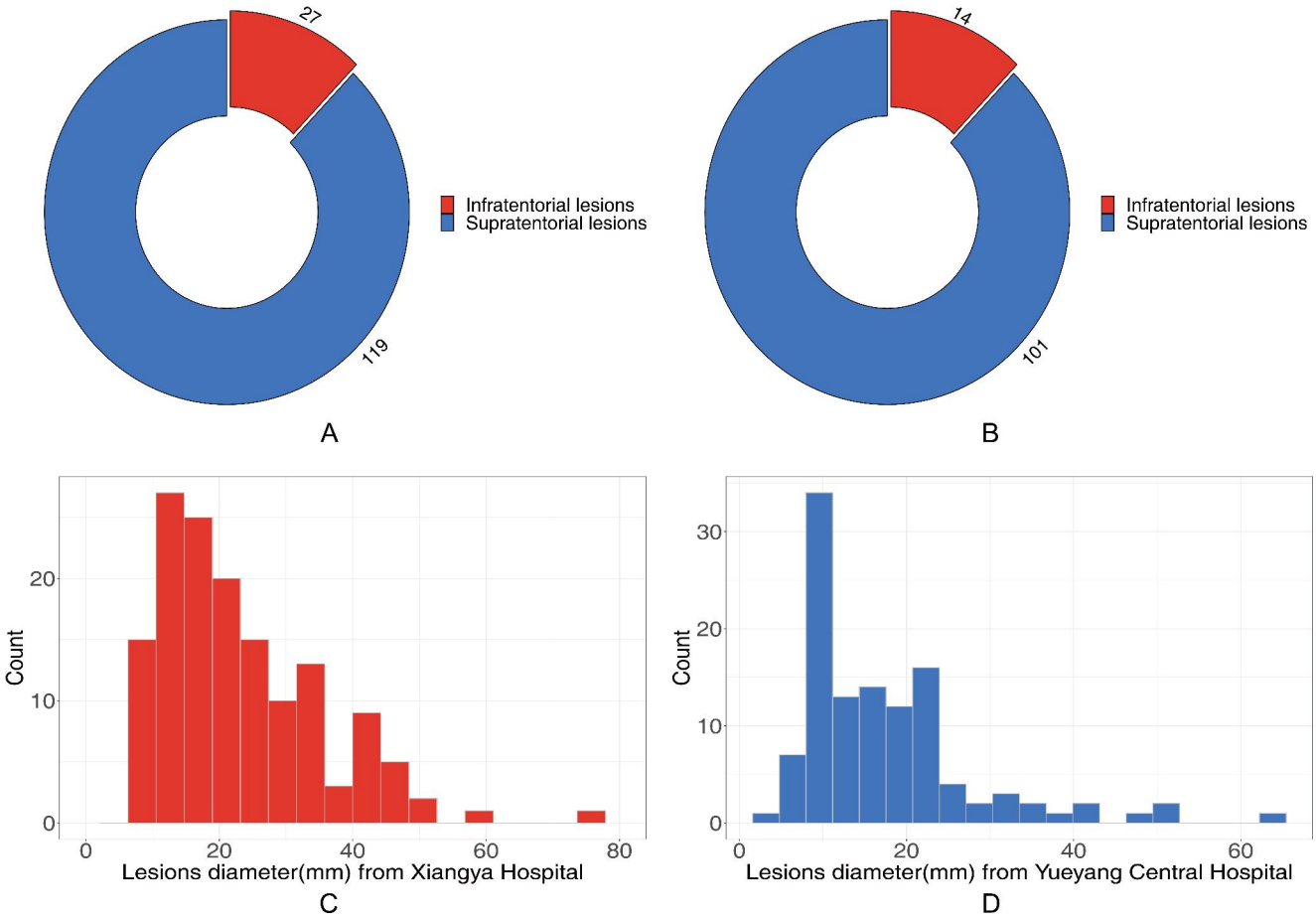

Fig. 2 **A**, The pie chart shows the number of lesions locations of Supratentorial or Infratentorial lesions in Xiangya Hospital patients; **B**, The pie chart shows the number of lesions locations of Supratentorial or Infratentorial lesions in Yueyang Central hospital; **C,D**, The histogram is the distribution of lesion diameters in the two hospitals

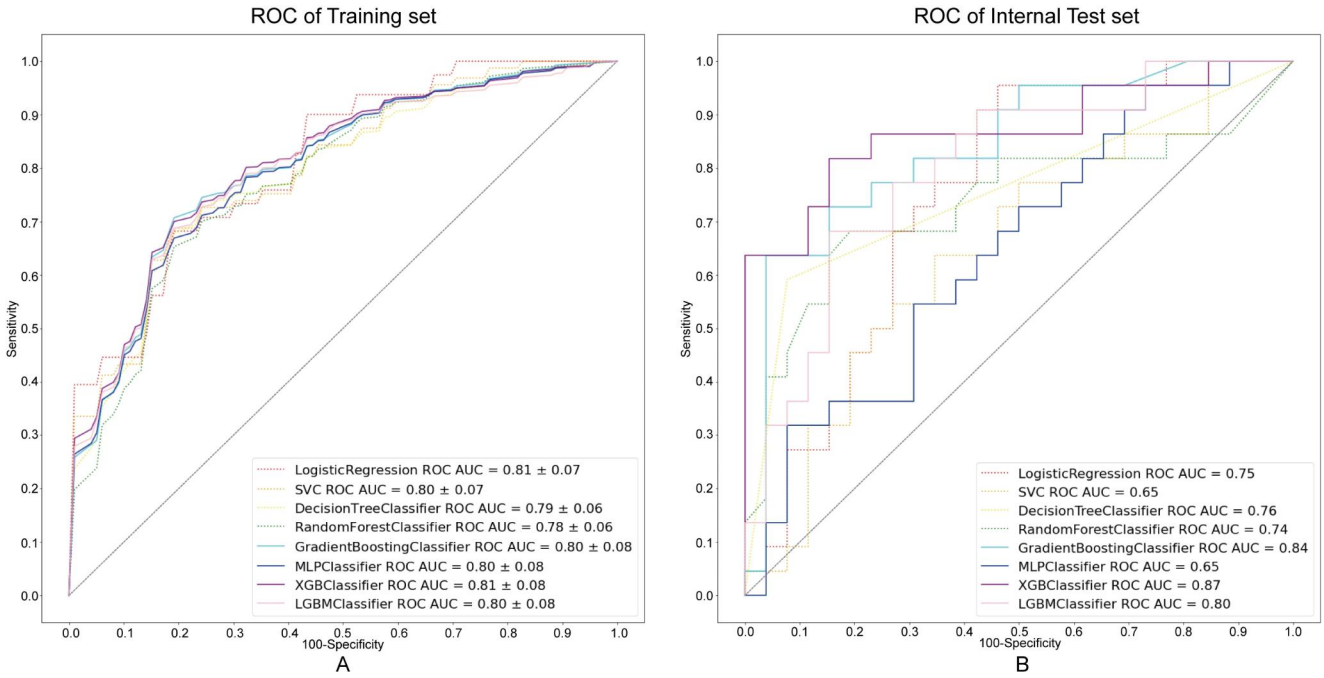

Fig. 3 ROC curves are eight models for comparing the pathological classification of NSCLC brain lesions. **A**, Training set AUC, **B**, Internal Test set AUC of images only using T1C radiomics

Table 3 The AUC of the eight models in the Internal Test set

Data Source	LR	SVM	DT	RF	GBM	MLP	XgBoost	LGBM
T1	0.77	0.76	0.74	0.76	0.82	0.77	0.84	0.72
T1C	0.75	0.65	0.72	0.71	0.84	0.66	0.87	0.80
T2	0.73	0.63	0.78	0.65	0.74	0.73	0.75	0.82
Union	0.76	0.70	0.65	0.76	0.76	0.76	0.76	0.81

Abbreviation: LR, Logistic Regression; SVM, Support Vector Machine; MLP, Multilayer Perceptron; DT, Decision Tree; RF, Random Forest; GBM, gradient boosting; XgBoost, extrema gradient boosting; LGBM, light gradient boosting

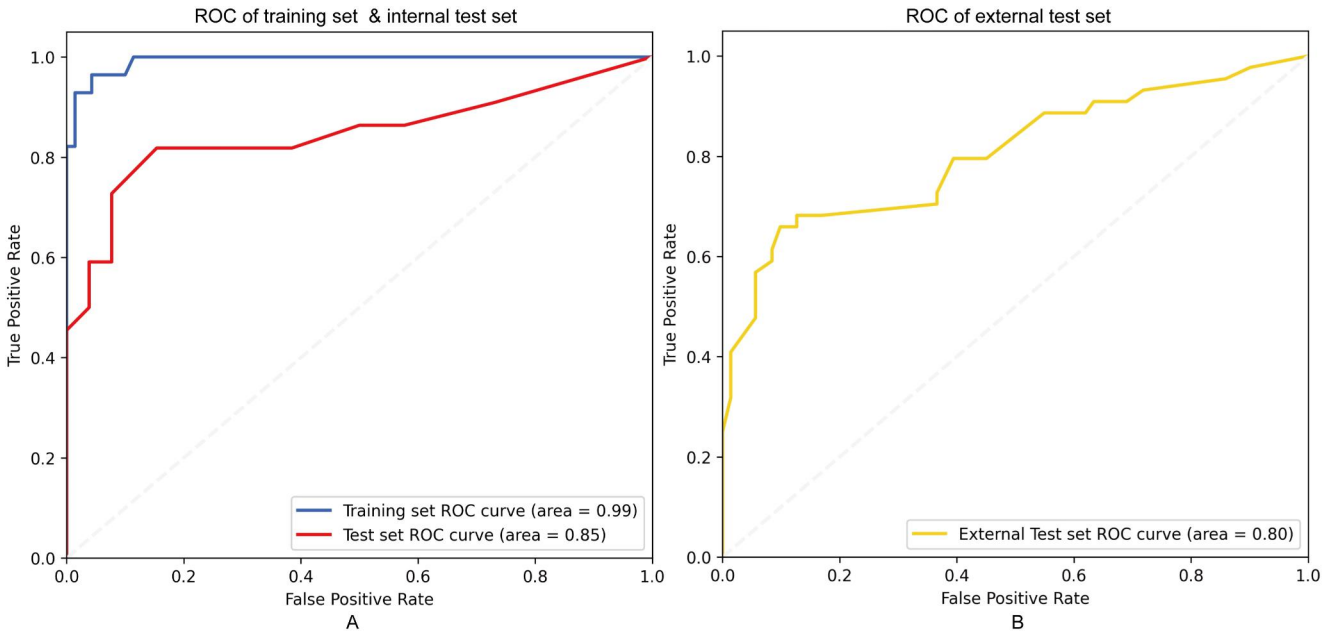

Fig. 4 AUC plots were using contrast-enhanced T1 and Xgboost after adjusting parameters. A, The AUC is from training and internal test set; B, The AUC is from external test set

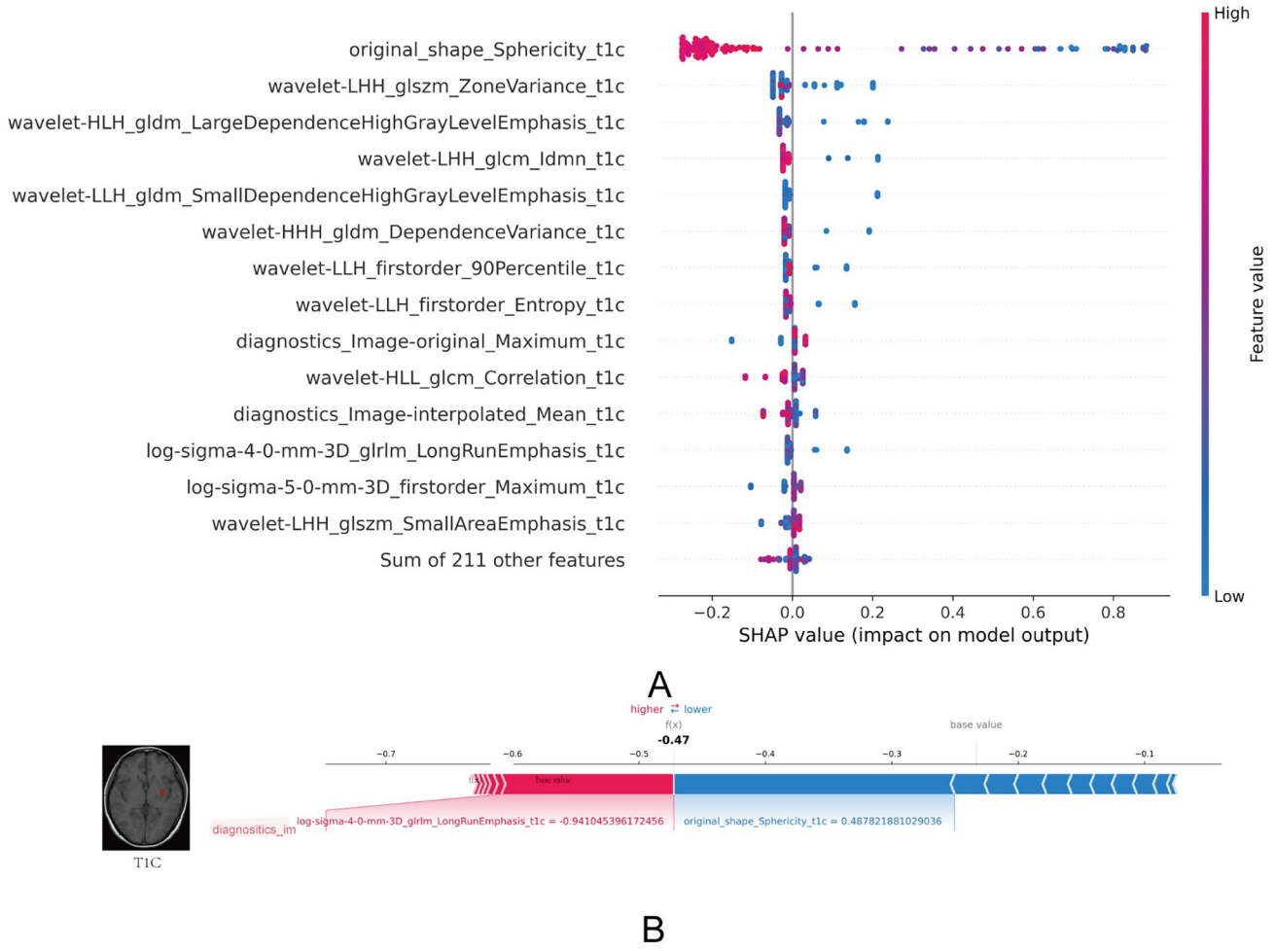

Fig. 5 A, Bubble plot is a SHAP value versus feature importance. B, SHAP force plots explained how the radiomics model discriminates in the pathology of an NSCLC brain lesions patient

specificity is 92.30%; sensitivity is 68.18%; recall is 77.42%; F score is 84.21%. Figure 4B shows the AUROC(0.85) of Xgboost in the external test set and the accuracy is 77.39%. The accuracy of the 3D-ResNet 34 to identify lung adenocarcinoma or squamous carcinoma is 0.52, which is worse than the ML classifier models.

Model Explanation

The Xgboost can be used to rank feature importance using SHAP values (Fig. 5). From a global explanatory perspective, the top features ranked by important radiomics feature analysis using Xgboost are displayed in Fig. 5A. For the squamous cell carcinoma vs. adenocarcinoma in the lung classification task, a positive SHAP value (pink) indicates the squamous cell carcinomas label and a negative SHAP value (blue) indicates the adenocarcinoma label. From a local explanatory perspective, Fig. 5B could interpret the assessment of a single patient. It visualized the SHAP value of a feature as a force to increase or decrease the assessment, and every prediction began with the base value (-0.646) which was the average SHAP value of all predictions. The arrow's length accounted for how much (in percentage) a particular feature contributed to the SHAP value. The arrow's color represented whether the contributions were positive (red) or negative (blue). As shown in Fig. 5B, the SHAP value of this patient was − 0.66, which was smaller than the base value (-0.646), thereby indicating that we could assess this patient in the adenocarcinoma group.

Discussion

In this study, it was first demonstrated that the combination of machine learning techniques with SHAP methods and high-dimensional radiological features from brain MRI data. The model, finally built with the Xgboost algorithm and enhanced-contrast T1 MRI, has a good performance

(AUROC in external independent test cohort: 0.80). The model is interpreted by SHAP values to identify which features correlate most strongly with improved stability, and to find that the most important properties classifying the pathology in the brain metastasis. Therefore, this feature of tumor images reliably predicts pathology and could provide solid foundations for the subdivided into different gene expressions to assist clinical decisions in an understandable way for both patients and clinicians.

Radiomics can provide more detailed characterization than possible by the human eye. Exclusively, previous studies have focused on the identification of the actual source of brain metastasis on the organ level [22]. Other scholars assessed the discriminative ability of T1-weighted MRI to distinguish between brain metastasis and glioblastoma [41–43]. Thus far, sub-pathology classification studies in NSCLC brain metastasis have not been performed, which means different radiomic patterns or phenotyping approaches in brain lesions from NSCLC are needed to accurately describe. The establishment of auto-classification models of pathology in BMs from NSCLC is a prerequisite for further subdivisions of driver genes, which can facilitate the selection of model parameters or even transfer learning. We found that several radiomics features were able to build a model to diagnose pathology in brain lesions, and the results revealed potential associations between radiomics analysis and sub-pathology in these brain lesions from NSCLC. Here, we used a machine learning algorithm (Xgboost) to link multiple MRI radiomics features with the pathology of non-small cell lung cancer subtypes. Some prediction models' performances are affected by the radiomics platform. In the study of the model for predicting the clinical outcomes in patients treated with radiosurgery for brain metastases from NSCLC, it was shown that the radiomic platform could yield several discrepancies in computed radiomic feature values [44]. This inspires us that when applying the prediction model proposed in the previous studies, the extracted radiomics platform should be detailed and described, including the version consistency of the language. Therefore, we ensured the reproducible and generalizable application of our model and made the following contributions. First, we described the package and programming language version in the Method part. Secondly, the hyperparameters for extracting features are given in the GitHub (the 'yaml' file). Thirdly, the hyperparameters optimized for the model are presented on the Git-hub.

Differences in the model design also play a role and reflect the various areas of focus among the algorithms and the representation of radiomics features of cancer pathology. Logistic regression, the most common model inferred on the parameters directly from data, is a linear model. There are the radiomics features with high-dimensional, highly correlated nature of excessive features in MRI that the linear model cannot handle. Support vector machines (SVM) use kernels and therefore do not work directly on the features, but rather on reduced and tractable representations. The results of SVM do not show advantages, and the reason may be that the sample size is not enough to support the vectors in higher dimensions and find the key hyperplane. The principal weakness of supervised learning is overfitting, and the multi-layer perceptron is presented as the most severe overfitting. Multi-layer Perceptron is a simplified version of deep learning. Previous original deep learning algorithms studies have shown that the sample size is thousands due to the tens or even hundreds of millions of parameters being iteratively optimized. The tree-like models are fairly robust in structure because those modeling algorithms always take the optimal decision for splitting at each node and keep this decision. Models developed by Xgboost showed the best performance, which is low computational time complexity and good generalization. In summary, to establish a classification model using radiomics features, it is necessary to consider the patient sample size and disease characteristics comprehensively. It is recommended to use the tree-like model as the primary modeling method, and then perform grid optimization and adjust parameters.

Xgboost performed well and was starting to have less intuitive connections between important features and brain lesions pathology. Therefore, we also interpret the results of the "Black Box" machine learning models with a recently developed novel tool: SHAP. In interpretable research, the SHAP value of each feature is calculated. The top-ranked features are original shape sphericity, zone variance, diagnostic image interpolated maximum, first-order entropy, and so on. shape sphericity is a measure of the roundness of the shape of the lesion's region relative to a sphere. The zone variance measures the variance in zone size volumes for the zones. These two radiomic features may be directly visible on MRI. There is a difference between the roundness of the brain lesions regions and the pixel value of the brain lesions regions in different pathology. The tumor roundness is related to the tumor-infiltrating region, necrotic region, and edema region [41, 45].

In clinical applications, this diagnostic prediction model can be applied to patients who cannot undergo invasive pathological testing. When patients have contraindications for percutaneous biopsy in the lung or brain, the pathology classification model integrates these non-invasive MRI data for the attending physician to create a management plan. The first step is to predict the results of different pathological types of lung cancer based on a non-invasive examination, and the second step is to predict whether different driver genes are positive. The two-step approach can be used to build the model with high accuracy, which can save a lot of

examination costs and waiting time. On the hand, Biopsy is the criterion method for histological classification and genetic status confirmation, which may lead to neurological dysfunction. On the other hand, a biopsy may only represent a small proportion of the tumor, which is difficult to reflect the entire characteristics of the whole neoplasm[11]. Therefore, it is important to find a noninvasive method for auxiliary diagnoses of histological subtypes of brain lesions, such as radiomics, which has achieved impressive success in different cancer types.

We would like to address some limitations of our present study. First, this study included a relatively non-large number of patients for both the training and test cohorts. We tried our best to collect patients with brain lesions from non-small cell lung cancer in the past 5 years. Although there are still not many, the robustness of our model is good. Secondly, unconventional and special MR imaging, including Diffusion Weighted Imaging (DWI), Diffusion Tensor Imaging (DTI), Dynamic Susceptibility Contrast (DSC)-MRI, and Dynamic Contrast Enhanced (DCE)-MRI, are still very rarely used in clinical routines, but these kinds of images have essentially important information for tumor microvascular architecture, oxygen metabolism and so on. Thirdly, compared to deep learning approaches, traditional machine learning shows high stability, is transparent, and is easy to understand due to its low complexity. Because the tumor has a volume and is 3-dimensional in space, we tried to use a 3D neural network for modeling, but the performance is not as good as traditional machine learning, probably because the neural network requires a mass of data and demands advanced computer hardware. Therefore, our future work will focus on the neural networks for brain metastasis tumor classification and prediction of the prognosis of those patients, with a growing number of brain metastasis in our database over time. A crucial goal for the future could be to build a sophisticated and organized precision oncology decision support system for precision medicine and individualized treatment of human cancers.

Conclusion

In this study, we have built a novel and explicable model for the first time to distinguish between lung squamous cell carcinoma brain lesions and lung adenocarcinoma brain lesions, using radiomics from contrast-enhanced T1 MRI integrated with an extreme gradient boosting algorithm in a clinical setting. Our results have shown that machine learning models have superior performance to 3D Res-Net models when working with small sample sizes. Additionally, we have demonstrated that the SHAP method can assist clinicians in directing personalized treatment strategies, providing a clear understanding of the different pathologies.

Acknowledgements This work was supported in part by the High Performance Computing Center of Central South University.

Author contributions Fuxing Deng: Statistical analysis, machine learning model and writing the paper. Zhiyuan Liu: Identifying cases, experiments. Wei Fang; Identifying cases. Lishui Niu: Discussion and editing the paper. Xianjing Chu: Paper editing and identifying cases. Quan Cheng: Identifying cases and discussion. Zijian Zhang: study design, analysis and editing the paper. RongRong Zhou: Study design, paper editing, financial support and overall study supervision. Zijian Zhang and RongRong Zhou are senior and corresponding authors who contributed equally to this study. Guang Yang: Supervision and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding The work was supported in part by the Science Foundation of Hunan Province (grant number: 2022JJ30992); The work was supported in part by the Project Program of the National Clinical Research Center for Geriatric Disorders(Xiangya Hospital, Grant No. 2022LNJJ10); The work was supported in part by the ERC IMI (101005122), the H2020 (952172), the MRC (MC/PC/21013), the Royal Society (IEC/NSFC/211235), the NVIDIA Academic Hardware Grant Program, NIHR Imperial Biomedical Research Centre (RDA01), Imperial–Nanyang Technological University Collaboration Fund, UKRI MRC with MSIT and NRF Fund, and the UKRI Future Leaders Fellowship (MR/V023799/1).

Data Availability The code is open source and freely available at https://github.com/BboyT/BM_NSCLC_subpathology. Data supporting the findings of this study can be provided upon journal request.

Declarations

Competing Interests The authors declare no conflict of interest.

Ethics approval The multicenter study was conducted in accordance with the Declaration of Helsinki and was based on retrospectively identified or prospectively acquired brain metastasis MRI data collection in the participating centers, according to the local institutional review boards guidelines and the Central South University, Xiangya hospital institution ethical committees' approvals. (Referencenumber:202210235).

Consent for publication Not applicable.

References

1. Lamba N, Wen PY, Aizer AA (2021) Epidemiology of brain metastases and leptomeningeal disease. Neuro Oncol 23(9):1447–1456
2. Cagney DN, Martin AM, Catalano PJ, Redig AJ, Lin NU, Lee EQ, Wen PY, Dunn IF, Bi WL, Weiss SE et al (2017) Incidence and prognosis of patients with brain metastases at diagnosis of systemic malignancy: a population-based study. Neuro Oncol 19(11):1511–1521
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71(3):209–249

4. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA (2008) Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin Proc 83(5):584–594
5. Relli V, Trerotola M, Guerra E, Alberti S (2019) Abandoning the notion of Non-Small Cell Lung Cancer. Trends Mol Med 25(7):585–594
6. Kim HS, Mitsudomi T, Soo RA, Cho BC (2013) Personalized therapy on the horizon for squamous cell carcinoma of the lung. Lung Cancer 80(3):249–255
7. Planchard D, Besse B, Groen HJM, Souquet PJ, Quoix E, Baik CS, Barlesi F, Kim TM, Mazieres J, Novello S et al (2016) Dabrafenib plus trametinib in patients with previously treated BRAF(V600E)-mutant metastatic non-small cell lung cancer: an open-label, multicentre phase 2 trial. Lancet Oncol 17(7):984–993
8. Stein MK, Pandey M, Xiu J, Tae H, Swensen J, Mittal S, Brenner AJ, Korn WM, Heimberger AB, Martin MG (2019) Tumor Mutational Burden is Site Specific in Non-Small-Cell Lung Cancer and is highest in Lung Adenocarcinoma Brain Metastases. JCO Precis Oncol 3:1–13
9. Sperduto PW, De B, Li J, Carpenter D, Kirkpatrick J, Milligan M, Shih HA, Kutuk T, Kotecha R, Higaki H et al (2022) Graded Prognostic Assessment (GPA) for patients with Lung Cancer and Brain Metastases: initial report of the small cell Lung Cancer GPA and Update of the Non-Small Cell Lung Cancer GPA including the effect of programmed death ligand 1 and other prognostic factors. Int J Radiat Oncol Biol Phys 114(1):60–74
10. Goldberg SB, Schalper KA, Gettinger SN, Mahajan A, Herbst RS, Chiang AC, Lilenbaum R, Wilson FH, Omay SB, Yu JB et al (2020) Pembrolizumab for management of patients with NSCLC and brain metastases: long-term results and biomarker analysis from a non-randomised, open-label, phase 2 trial. Lancet Oncol 21(5):655–663
11. Herbst RS, Morgensztern D, Boshoff C (2018) The biology and management of non-small cell lung cancer. Nature 553(7689):446–454
12. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D (2012) Radiomics: the process and the challenges. Magn Reson Imaging 30(9):1234–1248
13. Zhang L, Wang Y, Peng Z, Weng Y, Fang Z, Xiao F, Zhang C, Fan Z, Huang K, Zhu Y et al (2022) The progress of multimodal imaging combination and subregion based radiomics research of cancers. Int J Biol Sci 18(8):3458–3469
14. Wang H, Wang Y, Zhang H, Han Y, Li Q, Ye Z (2020) Preoperative CT features for prediction of ALK gene rearrangement in lung adenocarcinomas. Clin Radiol 75(7):562e521–562e529
15. Sibille L, Seifert R, Avramovic N, Vehren T, Spottiswoode B, Zuehlsdorff S, Schäfers M (2020) (18)F-FDG PET/CT uptake classification in Lymphoma and Lung Cancer by using deep convolutional neural networks. Radiology 294(2):445–452
16. He R, Yang X, Li T, He Y, Xie X, Chen Q, Zhang Z, Cheng T (2022) A Machine Learning-Based Predictive Model of Epidermal Growth Factor Mutations in Lung Adenocarcinomas. Cancers (Basel) 14(19)
17. Li Y, Liu Y, Liang Y, Wei R, Zhang W, Yao W, Luo S, Pang X, Wang Y, Jiang X et al (2022) Radiomics can differentiate high-grade glioma from brain metastasis: a systematic review and meta-analysis. Eur Radiol 32(11):8039–8051
18. Wood DA, Kafiabadi S, Busaidi AA, Guilhem E, Montvila A, Lynch J, Townend M, Agarwal S, Mazumder A, Barker GJ et al (2022) Accurate brain-age models for routine clinical MRI examinations. NeuroImage 249:118871
19. Su C, Jiang J, Zhang S, Shi J, Xu K, Shen N, Zhang J, Li L, Zhao L, Zhang J et al (2019) Radiomics based on multicontrast MRI can precisely differentiate among glioma subtypes and predict tumour-proliferative behaviour. Eur Radiol 29(4):1986–1996
20. Cao R, Pang Z, Wang X, Du Z, Chen H, Liu J, Yue Z, Wang H, Luo Y, Jiang X (2022) Radiomics evaluates the EGFR mutation status from the brain metastasis: a multi-center study. Phys Med Biol 67(12)
21. Liu Z, Jiang Z, Meng L, Yang J, Liu Y, Zhang Y, Peng H, Li J, Xiao G, Zhang Z et al (2021) Handcrafted and Deep Learning-Based Radiomic Models Can Distinguish GBM from Brain Metastasis. J Oncol. 2021:5518717
22. Kniep HC, Madesta F, Schneider T, Hanning U, Schonfeld MH, Schon G, Fiehler J, Gauer T, Werner R, Gellissen S (2019) Radiomics of Brain MRI: utility in prediction of metastatic tumor type. Radiology 290(2):479–487
23. Joo B, Ahn SS, An C, Han K, Choi D, Kim H, Park JE, Kim HS, Lee SK (2022) Fully automated radiomics-based machine learning models for multiclass classification of single brain tumors: Glioblastoma, lymphoma, and metastasis. J Neuroradiol
24. Zhao X, Huang W, Huang X, Robu V, Flynn D (2021) Baylime: bayesian local interpretable model-agnostic explanations. In: de Campos, C, Maathuis, MH (eds) Proceedings of the thirty-seventh Conference on Uncertainty in Artificial Intelligence. PLMR, p 887-896
25. Johnson PM, Barbour W, Camp JV, Baroud H (2022) Using machine learning to examine freight network spatial vulnerabilities to disasters: a new take on partial dependence plots. Transp Res Interdisciplinary Perspect 14:100617
26. Castro J, Gómez D, Tejada J (2009) Polynomial calculation of the Shapley value based on sampling. Comput Oper Res 36(5):1726–1730
27. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 30
28. Yaniv Z, Lowekamp BC, Johnson HJ, Beare R (2018) SimpleITK Image-Analysis Notebooks: a collaborative environment for Education and Reproducible Research. J Digit Imaging 31(3):290–303
29. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin J-C, Pieper S, Aerts HJWL (2017) Computational Radiomics System to Decode the Radiographic phenotype. Cancer Res 77(21):e104–e107
30. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga RJ, Boellaard R et al (2020) The image Biomarker Standardization Initiative: standardized quantitative Radiomics for High-Throughput Image-based phenotyping. Radiology 295(2):328–338
31. Spearman C (1904) The Proof and Measurement of Association between two things. Am J Psychol 15(1):72–101
32. Zheng B, Agresti A (2000) Summarizing the predictive power of a generalized linear model. Stat Med 19(13):1771–1781
33. Paul A, Mukherjee DP, Das P, Gangopadhyay A, Chintha AR, Kundu S (2018) Improved random forest for classification. IEEE Trans Image Process 27(8):4012–4024
34. Suthaharan S (2016) Support vector machine. In: Machine learning models and algorithms for big data classification: thinking with examples for effective learning. Springer US, p 207–235
35. Swain PH, Hauska H (1977) The decision tree classifier: design and potential. IEEE Trans Geoscience Electron 15(3):142–147
36. Natekin A, Knoll A (2013) Gradient boosting machines, a tutorial. Front Neurorobotics 7:21
37. Kim T, Adali T (2002) Fully complex multi-layer perceptron network for nonlinear signal processing. J VLSI signal Process Syst signal image video Technol 32(1):29–43
38. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K (2015) Xgboost: extreme gradient boosting. R package version 04 – 2 1(4):1–4
39. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) Lightgbm: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 30

40. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
41. Artzi M, Bressler I, Ben Bashat D (2019) Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. J Magn Reson Imaging 50(2):519–528
42. Qian Z, Li Y, Wang Y, Li L, Li R, Wang K, Li S, Tang K, Zhang C, Fan X et al (2019) Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers. Cancer Lett 451:128–135
43. Ortiz-Ramon R, Ruiz-Espana S, Molla-Olmos E, Moratal D (2020) Glioblastomas and brain metastases differentiation following an MRI texture analysis-based radiomics approach. Phys Med 76:44–54
44. Carloni G, Garibaldi C, Marvaso G, Volpe S, Zaffaroni M, Pepa M, Isaksson LJ, Colombo F, Durante S, Lo Presti G et al (2022) Brain metastases from NSCLC treated with stereotactic radiotherapy: prediction mismatch between two different radiomic platforms. Radiother Oncol 178:109424
45. Young RJ, Knopp EA (2006) Brain MRI: tumor evaluation. J Magn Reson Imaging 24(4):709–724