



# Radiomics models using machine learning algorithms to differentiate the primary focus of brain metastasis

Yuping Xie<sup>1#</sup>, Xuanzi Li<sup>1#</sup>, Shuai Yang<sup>2#</sup>, Fujie Jia<sup>1</sup>, Yuanyuan Han<sup>1</sup>, Mingsheng Huang<sup>1</sup>, Lei Chen<sup>3</sup>, Wei Zou<sup>1</sup>, Chuntao Deng<sup>2</sup>, Zibin Liang<sup>1</sup>

<sup>1</sup>The Cancer Center of The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, China; <sup>2</sup>Department of Radiotherapy, The Cancer Center of the Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, China; <sup>3</sup>Department of Neurosurgery, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, China

**Contributions:** (I) Conception and design: Y Xie, X Li, S Yang, C Deng, Z Liang; (II) Administrative support: Y Xie, C Deng, Z Liang; (III) Provision of study materials or patients: Z Liang, S Yang, W Zou, L Chen; (IV) Collection and assembly of data: Y Xie, Y Han, M Huang; (V) Data analysis and interpretation: Y Xie, Y Han, X Li, S Yang, C Deng, Z Liang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

**Correspondence to:** Chuntao Deng, MSc. Department of Radiotherapy, The Cancer Center of The Fifth Affiliated Hospital of Sun Yat-sen University, No. 52 Meihua East Road, Xiangzhou District, Zhuhai 519020, China. Email: dengcht@mail.sysu.edu.cn; Zibin Liang, MD, PhD. The Cancer Center of The Fifth Affiliated Hospital of Sun Yat-sen University, No. 52 Meihua East Road, Xiangzhou District, Zhuhai 519020, China. Email: liangzb@mail.sysu.edu.cn.

**Background:** Brain metastases are common brain tumors in adults. Brain metastases from different primary tumors have special magnetic resonance imaging (MRI) features. As a new technology that can extract and quantify medical image data, and with the rapid development of artificial intelligence, the machine learning model based on radiology has been successfully applied to the diagnosis and differentiation of tumors. This study aimed to develop radiomics models from post-contrast T1-weighted images using machine learning algorithms to differentiate lung cancer from breast cancer brain metastases.

**Methods:** A retrospective analysis was conducted on 118 lung cancer brain metastases patients and 62 breast cancer brain metastases patients confirmed by surgery pathology or combined clinical and imaging diagnosis at The Fifth Affiliated Hospital of Sun Yat-sen University from August 2015 to September 2023. Patients were randomly divided into a training set (126 cases) and a validation set (54 cases) at a 7:3 ratio. Enhanced T1-weighted images of all patients were imported into ITK-SNAP software to manually delineate the region of interest (ROI). Radiomic features were extracted based on the ROI and feature selection was performed using the least absolute shrinkage and selection operator. Significant features were used to develop models using logistic regression (LR), support vector machine (SVM), K-nearest neighbors (KNN), multilayer perceptron (MLP), and light gradient boosting machine (LightGBM). The diagnostic performance of the models was assessed using the receiver operating characteristic (ROC) curve.

**Results:** The LightGBM radiomics model exhibited the best diagnostic performance, with an area under the curve (AUC) of 0.875 [95% confidence interval (CI): 0.819–0.931] in the training set and 0.866 (95% CI: 0.740–0.993) in the validation set.

**Conclusions:** The enhanced MRI radiomics model, especially the LightGBM model, can accurately predict the primary lesion types of brain metastases from lung cancer and breast cancer origins.

**Keywords:** Brain metastasis; lung cancer; radiomics; contrast-enhanced T1-weighted images; magnetic resonance imaging (MRI)

Submitted Aug 03, 2024. Accepted for publication Jan 09, 2025. Published online Feb 24, 2025.

doi: 10.21037/tcr-24-1355

**View this article at:** <https://dx.doi.org/10.21037/tcr-24-1355>

<sup>^</sup> ORCID: 0009-0001-2768-3775.

## Introduction

Brain metastasis is a prevalent malignancy affecting the central nervous system, comprising 10–15% of intracranial tumors and posing a significant threat to the health and survival of cancer patients (1). It is estimated that 20–40% of individuals with malignant tumors will develop brain metastasis during the course of their disease progression. The primary sources of brain metastasis are commonly lung cancer (50%), breast cancer (15–20%), malignant melanoma (5–10%), renal cancer (7%), and colorectal cancer (4–6%) (2). The median survival time varies depending on the primary tumor type, ranging from 2 to 27 months.

Lung cancer stands as the predominant primary source of brain metastases, with its incidence and mortality rates showing a gradual increase. Long-term follow-up results from the U.S. Surveillance, Epidemiology, and End Results (SEER) database showed that for patients with non-small cell lung cancer (NSCLC), the risk of brain metastases is 11%, 6% and 12% for adenocarcinoma, squamous cell

carcinoma and large cell carcinoma, respectively (3). For small cell lung cancer (SCLC) patients, brain metastases occur in 10% at initial diagnosis, rising to 40–50% during their diagnosis and treatment, and as high as 60–80% in patients surviving more than 2 years (4).

In approximately 10% of patients with brain metastases, the central nervous system serves as the primary presentation of the underlying malignancy, presenting with symptoms including headaches, vomiting, cognitive decline, behavioral alterations, sensory impairments, motor weakness, aphasia, visual disturbances, ataxia, seizures, among others (5). In cases of brain metastases of unknown primary origin, timely and precise identification and diagnosis of the primary tumor site are crucial for prognostic evaluation and clinical decision-making.

Magnetic resonance imaging (MRI) scans are the preferred imaging modality for detecting brain metastases. It can precisely define the tumor's location, shape, size, enhancement pattern, presence of edema, hemorrhage, necrosis, and cystic changes, aiding in identifying the potential origin of the metastases. Compared with non-enhanced T1-weighted sequences, enhanced T1-weighted sequences have some advantages in MRI radiomics of brain metastases, such as enhanced lesion visualisation, vascular enhancement, well-defined lesion boundaries and enhanced dynamic monitoring of lesions, which are more useful for identifying the primary focus of brain metastases (6).

Currently, the diagnosis of the primary tumor often relies on high-risk invasive biopsies, time-consuming pathological evaluations, and genetic test results. For patients exhibiting neurological symptoms, the presence of multiple origins of secondary brain metastases and cerebral lesions that complicate cerebral infarction poses a significant diagnostic challenge. In recent years, the rapid advancement of artificial intelligence in the medical domain has led to a substantial increase in the attention towards radiomics, a discipline that integrates medical imaging and machine learning for quantitative analysis (7). The utilization of contrast-enhanced T1-weighted radiomics and machine-learning models offers a non-invasive approach for identifying the site of brain metastasis, enabling a comprehensive analysis of the primary tumor area (8). This approach circumvents the need for invasive investigations under time constraints and significantly accelerates the diagnostic and therapeutic decision-making process. Not only does this approach mitigate the financial burden on patients, but it also alleviates psychological and physical distress, thus facilitating earlier and more tailored

### Highlight box

#### Key findings

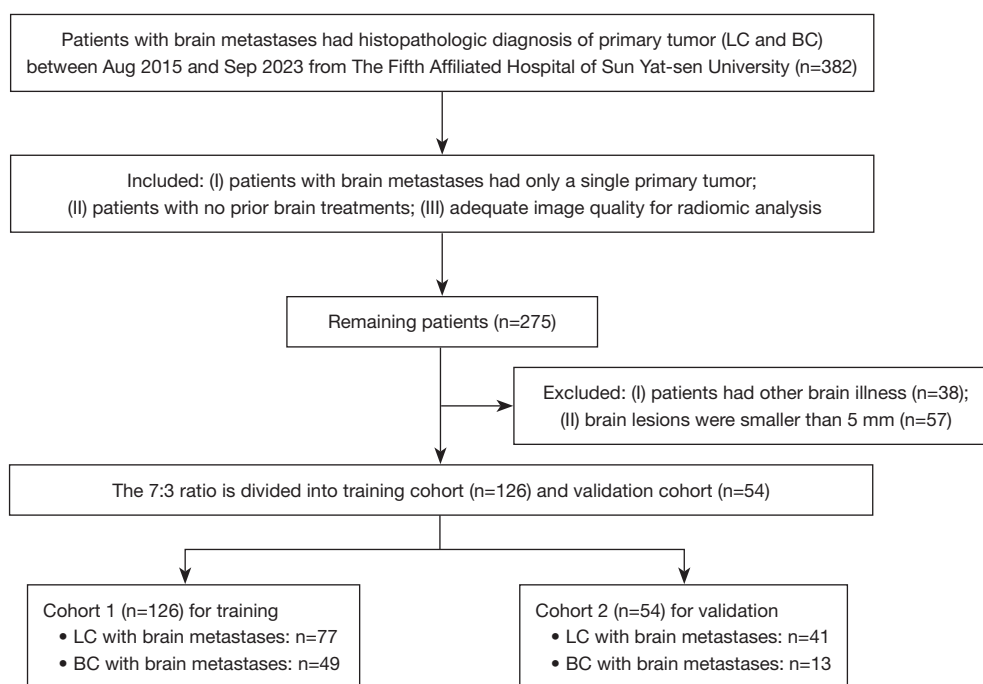
- Radiomics holds considerable potential in the prediction of primary lesions. Develop radiomics models from post-contrast T1-weighted images using machine learning algorithms to differentiate lung cancer from breast cancer brain metastases.

#### What is known and what is new?

- Brain metastases are common brain tumors in adults. Brain metastases from different primary tumors have special magnetic resonance imaging (MRI) features. As a new technology that can extract and quantify medical image data, and with the rapid development of artificial intelligence, the machine learning model based on radiology has been successfully applied to the diagnosis and differentiation of tumors.
- Radiomic features were extracted based on the region of interest (ROI) and feature selection was performed using the least absolute shrinkage and selection operator. Significant features were used to develop models using logistic regression (LR), support vector machine (SVM), K-nearest neighbors (KNN), multilayer perceptron (MLP), and light gradient boosting machine (LightGBM). The diagnostic performance of the models was assessed using the receiver operating characteristic (ROC) curve.

#### What is the implication, and what should change now?

- For patients with neurological symptoms as the first diagnosis of multiple origins and secondary brain metastases, there is an urgent need for contrast-enhanced T1-weighted imaging-based histology and machine learning models, which are expected to achieve non-invasive differential diagnosis of cerebral malignant tumours.



**Figure 1** The workflow of clinical enrollment process. LC, lung cancer; BC, breast cancer.

therapeutic management.

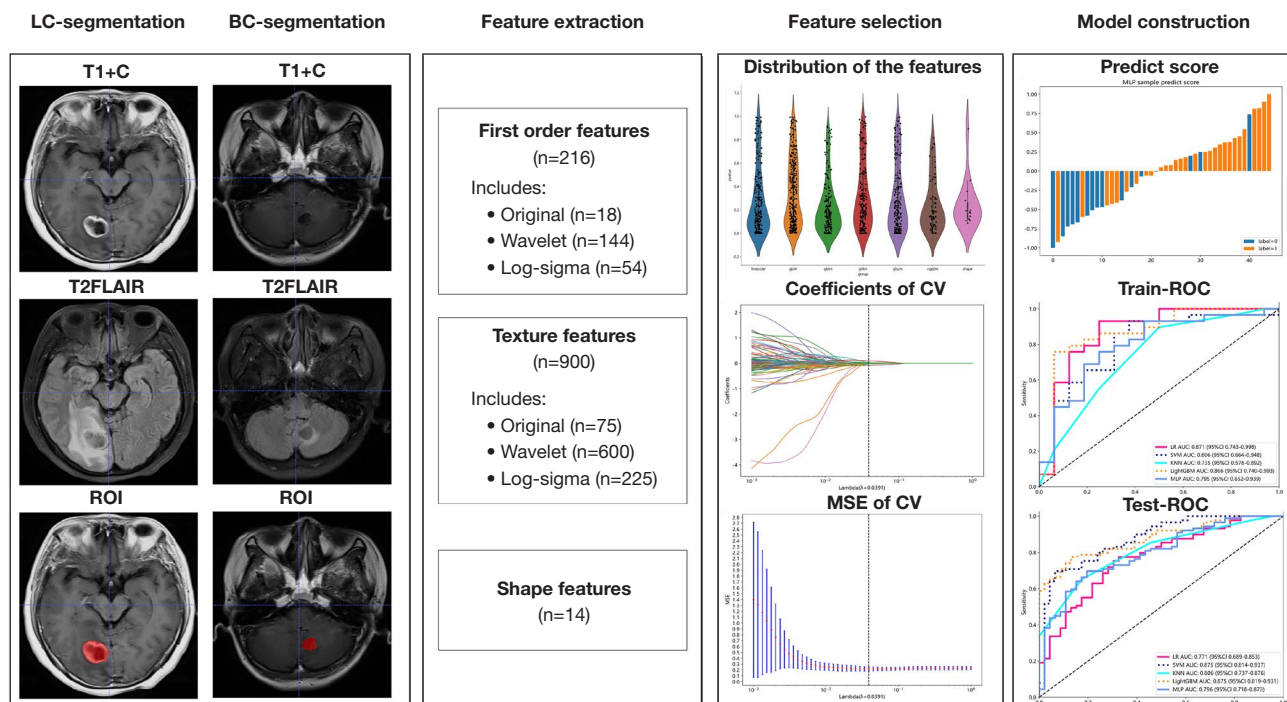
Therefore, radiomics holds considerable potential in the prediction of primary lesions. A noninvasive and reliable classification of lung cancer and non-lung cancer brain metastasis types, integrated with intratumoral, peritumoral, and morphological features, could significantly aid clinicians in making earlier and more convenient individualized treatment decisions (8). The objective of this study was to extract MRI radiological features from our hospital's brain metastasis dataset. Furthermore, we aimed to apply various established machine learning models, leveraging augmented T1-weighted radiomics, to differentiate between brain metastases originating from lung cancer and non-lung cancer sources. Currently, there is a lack of implementation of LightGBM, a cutting-edge imaging genomics algorithm, specifically tailored for differentiating the primary sites of brain metastases. Our research endeavor fills this gap by not only employing well-established algorithms that have been commonly used in the past, but also pioneering the integration of the novel LightGBM algorithm. This dual-pronged approach not only leverages the strengths of proven methodologies but also introduces a fresh perspective and enhanced performance, poised to improve the accuracy and efficiency of identifying the origins of

brain metastases. We present this article in accordance with the TRIPOD reporting checklist (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-1355/rc>).

## Methods

### General information

This study retrospectively collected MRI T1-weighted imaging (T1WI) images and clinical information of 382 patients with brain metastases treated at The Fifth Affiliated Hospital of Sun Yat-sen University from August 2015 to September 2023. Among these patients, 296 had lung cancer brain metastases, and 86 had brain metastases from breast cancer brain metastases. Inclusion criteria were: (I) brain metastases confirmed by radiological or histopathological evidence; (II) the primary tumor was a single primary cancer; (III) the primary tumor was confirmed by surgery or percutaneous biopsy; (IV) availability of MRI images suitable for radiomic analysis. Exclusion criteria included: (I) history of other malignant tumors; (II) history of cranial surgery for reasons other than metastases; (III) cranial MR enhanced images not meeting the quality standard for radiomic analysis; (IV) brain metastasis lesions smaller than 5 mm in diameter (*Figure 1*). This study followed



**Figure 2** The workflow of the radiomics model construction. LC, lung cancer; BC, breast cancer; FLAIR, fluid attenuated inversion recovery; ROI, region of interest; CV, cross validation; MSE, mean square error; ROC, receiver operating characteristic; LR, logistic regression; AUC, area under the curve; SVM, support vector machine; KNN, K-nearest neighbor; LightGBM, light gradient boosting machine; MLP, multilayer perceptron.

the Declaration of Helsinki (as revised in 2013) and was approved by the ethics committee of The Fifth Affiliated Hospital of Sun Yat-sen University [Approval No. [2019] Ethical Letter (K03-1)], and informed consent for this retrospective analysis was waived.

### Instruments and methods

MRI examinations for all patients were performed at The Fifth Affiliated Hospital of Sun Yat-sen University using various MR scanners (1.50, 3.00 T MR) to reflect image data heterogeneity, including 1.50 T Amira, 3.00 T Verio MR scanner (Siemens, Germany), 1.50T Aera, 1.50 T Ingenia Ambition MR scanner (Philips, Netherlands), and SIGNA Pioneer 3.00 T MR scanner (General Electric Company, USA). All patients were positioned supine with the scan range from the skull base to the vertex. For enhanced scanning: a double-barrel high-pressure syringe was used for rapid intravenous injection of the contrast agent Gadopentetic acid (Gd-DTPA) at a flow rate of 2 mL/s and a dosage of 0.1 mmol/kg, followed by a 20 mL

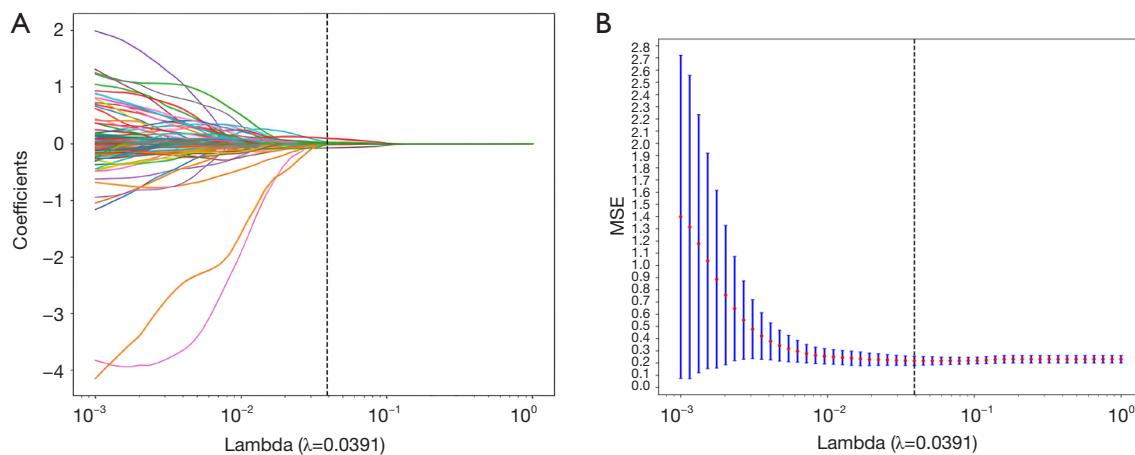
saline flush. Subsequent 3D thin-layer T1WI scanning was performed post-injection.

### Radiomic analysis workflow

The radiomic analysis process includes image segmentation, feature extraction, selection, and model building (Figure 2). A random stratified sampling method was used to divide the dataset into a training set and a validation set in a 7:3 ratio.

### Image segmentation and preprocessing

Images were obtained from the Picture Archiving and Communication System (PACS) and stored in Digital Imaging and Communications in Medicine (DICOM) format. Based on the patients' MRI diagnostic reports, the region of interest (ROI) was delineated on the axial enhanced T1WI+C images using ITK-SNAP software (version 3.8.0, www.itksnap.org), which had labelled the red areas in Figure 2. A manually labelled ROI is an image that is reviewed by a trained and experienced radiation oncologist who manually outlines the target area of the



**Figure 3** Coefficient convergence plot of LASSO screening features (A) and the relationship between the MSE and the regularization parameter lambda (B). LASSO, least absolute shrinkage and selection operator; MSE, mean square error.

tumour in the medical image. In our study, two radiologists with more than 10 years of experience and one radiation oncologist with 15 years of experience collaborated on the manual target sketching and cross-validation. ROI delineation criteria included: (I) the ROI boundary at the lesion edge; (II) layer-by-layer delineation of brain metastases; (III) avoiding surrounding edema or normal brain tissue as much as possible. As shown in *Figure 2*, it seems that lung cancer brain metastases and breast cancer brain metastases roughly showed round-like cystic cavity with low signal, accompanied by rim enhancement, and surrounded by irregular oedema bands. However, the size of the breast cancer brain metastases was smaller than that of the lung cancer brain metastases, and the surrounding oedema band was also smaller. At T1+C, the rim enhancement of lung cancer brain metastasis was more obvious, and the cystic cavity showed mixed signals, while the signals in the cystic cavity of breast cancer brain metastasis were more homogeneous.

### Feature extraction

MRI images were first preprocessed, including voxel normalization ( $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$ ) and z-score standardization of image intensities. The standardized images and delineated ROIs were imported into AK software (version 3.3.0, Artificial Intelligence Kit, GE Healthcare, China) for feature extraction, which included first-order features, shape features, gray level co-occurrence matrix (GLCM) features, gray level dependence matrix (GLDM) features, gray level run length matrix (GLRLM) features, gray level size zone

matrix (GLSZM) features, and neighboring gray tone difference matrix (NGTDM) features.

### Feature selection and modeling

To prevent overfitting, radiomic features were selected using the Z-score method for standardization. The least absolute shrinkage and selection operator (LASSO) was applied for further feature selection, eliminating irrelevant and ineffective features. Ten-fold cross-validation was used to determine the  $\lambda$  value corresponding to the minimum binomial deviation (*Figure 3*), identifying the most valuable radiomic features. Feature coefficient convergence plots displayed how feature coefficients converged during the training process, indicating their influence on the outcome variable (*Figure 3A*). A  $\lambda$  value of 0.0391, where the model had the minimum mean square error (*Figure 3B*), suggested optimal performance without overfitting or underfitting. Predictive models for differentiating lung cancer brain metastases from non-lung cancer brain metastases were established using logistic regression (LR), support vector machine (SVM), K-nearest neighbors (KNN), multilayer perceptron (MLP), and light gradient boosting machine (LightGBM) algorithms, with model efficacy evaluated using the test set data.

### Statistical analysis

All statistical analyses were performed using SPSS software (version 26) and R software. Count data were presented as cases (%), and continuous variables were compared using



**Table 1** Training set and validation set clinical data

Variable	Training set (n=126)	Validation set (n=54)	P value
Brain metastasis from different primary foci, n (%)			0.63 <sup>①</sup>
Brain metastases from non-lung cancer	42 (33.3)	20 (37)	
Brain metastases from lung cancer	84 (66.7)	34 (63)	
Gender, n (%)			0.49 <sup>①</sup>
Female	77 (61.1)	30 (55.6)	
Male	49 (38.9)	24 (44.4)	
Smoking history, n (%)			0.46 <sup>①</sup>
No	99 (78.6)	45 (83.3)	
Yes	27 (21.4)	9 (16.7)	
Drinking history, n (%)			0.73 <sup>①</sup>
No	118 (93.7)	52 (96.3)	
Yes	8 (6.3)	2 (3.7)	
Family history, n (%)			0.28 <sup>①</sup>
No	118 (93.7)	53 (98.1)	
Yes	8 (6.3)	1 (1.9)	
Age (years)	54.58±11.398	54.78±9.113	0.92 <sup>②</sup>
Height (cm)	161.75±6.8	162.59±8.227	0.51 <sup>②</sup>
Weight (kg)	58.312±9.1126	61.386±11.5647	0.11 <sup>②</sup>
BMI (kg/m <sup>2</sup> )	22.296±3.102	23.134±3.521	0.14 <sup>②</sup>
Body surface area (m <sup>2</sup> )	1.613±0.139	1.653±0.174	0.17 <sup>②</sup>

The data of age, height, weight, BMI, body surface area are presented as mean ± SD. <sup>①</sup> Chi-square test; <sup>②</sup> independent sample T test. BMI, body mass index; SD, standard deviation.

the Mann-Whitney *U* test or *t*-test, while group differences were assessed using the chi-square test. The efficacy of different radiomic models was evaluated and compared using receiver operating characteristic (ROC) curves. P value <0.05 was considered statistically significant.

## Results

### Patient demographics

A total of 180 patients with brain metastases were analyzed in this study, including 118 lung cancer brain metastasis patients and 62 patients with brain metastases from other sources (all breast cancer). The distribution of baseline clinical characteristics in the training and validation sets is detailed in *Table 1*. There were no statistical differences in gender, age, source of brain metastasis, smoking history, alcohol history, family history, height, weight, body mass index (BMI), or body surface area between the training and

validation sets.

Radiomic data included first-order, second-order, and higher-order statistics. Combining these data with other clinical features and using complex bioinformatics tools for data mining could help develop more precise predictive models. In this study, 1,197 radiomic features were extracted from the post-contrast T1WI phases of MRI images of 180 patients with brain metastases. Using LASSO (selection process as shown in *Figure 3*) and stepwise multivariate logistic regression for feature selection, 15 significant features were identified (*Table 2*). Using these 13 significant radiomic features, predictive models based on LR, SVM, KNN, MLP, and LightGBM algorithms were constructed. The models' efficacy in differentiating lung cancer brain metastases from non-lung cancer brain metastases was assessed by plotting ROC curves and calculating the corresponding AUC values. Results showed that the LightGBM radiomic model performed best in

**Table 2** The final preserved image characteristics and weights

Imaging characteristics	Pretreatment	Feature type	Specific eigenvalue	Characteristic
log_sigma_5_0_mm_3D_glszm_SizeZoneNonUniformityNormalized	Gauss-Laplace	GLSZM	Non-uniformity of large and small regions	0.097515324
wavelet_LLL_firstorder_RootMeanSquared	Small wave	First order	Root-mean-square error	0.087675246
log_sigma_4_0_mm_3D_glrlm_ShortRunEmphasis	Gauss-Laplace	GLRLM	Short-stroke low-gray emphasis	0.086062958
log_sigma_5_0_mm_3D_ngtgm_Strength	Gauss-Laplace	NGTDM	Strengthen	0.084395413
wavelet_HHH_firstorder_Mean	Small wave	First order	Average value	0.082015217
wavelet_HHH_firstorder_Kurtosis	Small wave	First order	Peak value	0.079964101
log_sigma_5_0_mm_3D_glrlm_ShortRunEmphasis	Gauss-Laplace	GLRLM	Short line emphasis	0.079282664
log_sigma_3_0_mm_3D_gldm_DependenceVariance	Gauss-Laplace	GLDM	Dependent variance	0.078448176
wavelet_LLL_firstorder_10Percentile	Small wave	First order	10% eigenvalue	0.076517276
wavelet_HLH_firstorder_Median	Small wave	First order	Median	0.066439943
wavelet_HHH_glszm_SmallAreaEmphasis	Small wave	GLSZM	Small-area emphasis	0.066200971
wavelet_HLL_firstorder_RootMeanSquared	Small wave	First order	Root-mean-square error	0.062199447
wavelet_HLL_gldm_ClusterShade	Small wave	GLCM	Cluster shadow	0.053283265

GLSZM, gray level size zone matrix; GLRLM, gray level run length matrix; NGTDM, neighboring gray tone difference matrix; GLDM, gray level dependence matrix; GLCM, gray level co-occurrence matrix.

differentiating lung cancer brain metastases from non-lung cancer brain metastases, with AUC values in the training and validation sets of 0.875 (95% CI: 0.819–0.931) and 0.866 (95% CI: 0.740–0.993), respectively. In comparison, the LR, SVM, KNN, and MLP models showed slightly inferior performance. In the training set, the AUC values for the LR, SVM, KNN, and MLP models were 0.771 (95% CI: 0.689–0.853), 0.875 (95% CI: 0.814–0.937), 0.806 (95% CI: 0.737–0.876), and 0.796 (95% CI: 0.718–0.873), respectively. In the validation set, the AUC values were 0.871 (95% CI: 0.743–0.998), 0.806 (95% CI: 0.664–0.948), 0.735 (95% CI: 0.578–0.892), and 0.795 (95% CI: 0.652–0.939), respectively (Figure 4).

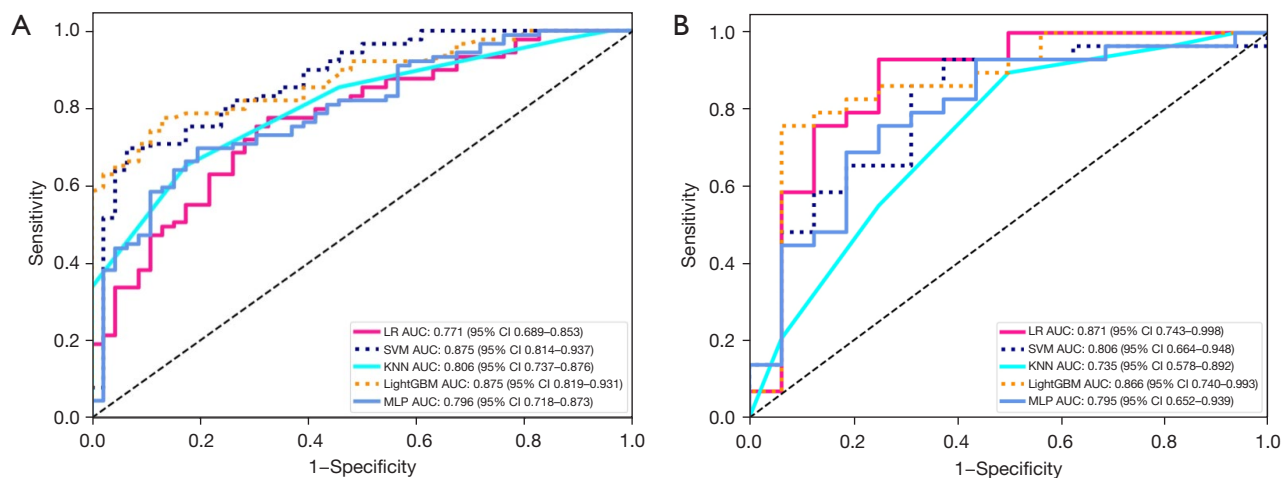
## Discussion

### Features of lung cancer brain metastases and breast cancer brain metastases

The most common primary tumour for brain metastases is lung cancer, accounting for about 50%. Common lung cancer gene mutations include *EGFR*, *ALK*, *ROS1*, *BRAF*, *NTRK*, *MET*, *RET*, *KRAS*, and *HER2* (9). These gene mutations may lead to different clinical features and

imaging manifestations. For example, *EGFR* mutation has the highest incidence in lung adenocarcinoma, accounting for about 50% to 60%, mainly in females and non-smokers, which may be manifested as multiple metastases in both lungs, ground-glass shadows, pleural appendage, and air-bronchial sign on CT imaging. The mutation rate of *ALK* in lung cancer is not high, about 4.90–6.03%, and is often seen in young, non-smoking women. Imaging often shows solid nodules and sometimes ground glass nodules with pericardial effusion and local lymph node enlargement. Lung cancer with *HER2* gene mutation on imaging shows lobulated lesions, burr sign, necrosis sign, accompanied by pleural pulling sign. *ROS1* about 1–2% gene mutation occurs in lung cancer, mainly in young, non-smoking women, these patients are more likely to have distant lymph node metastasis, pericardial metastasis and bone metastasis (10). Lung cancer patients with *RET* mutations present with small solid nodules of the peripheral type, and pleural dissemination may be observed. Other types of mutated lung cancers are not yet specific on imaging. However, few studies have reported differences in imaging in patients with brain metastases from lung cancer with different mutations.

The pathological types of lung cancer mainly include



**Figure 4** ROC curves of LR, SVM, KNN, LightGBM and MLP radiomics models in the training sets (A) and validation sets (B). LR, logistic regression; AUC, area under the curve; CI, confidence interval; SVM, support vector machine; KNN, K-nearest neighbor; LightGBM, light gradient boosting machine; MLP, multilayer perceptron; ROC, receiver operating characteristic.

adenocarcinoma, squamous cell carcinoma and large cell carcinoma, of which lung adenocarcinoma is the most common, accounting for about 50 per cent of the cases, with a majority of females, and mainly manifesting as peripheral lung cancer, with haematogenous metastasis being common. Squamous cell carcinoma is the next most common type, accounting for about 25–40% of the cases, mostly in males, and manifesting mostly as central lung cancer. Its growth rate is slower, disease duration is longer, local metastasis through lymphatic vessels is more common, and haematogenous distant metastasis occurs later. Large cell carcinoma is less differentiated, more malignant, has a poorer prognosis and is rarer (11). In our study, among 118 patients with brain metastases from lung cancer, 72 patients with adenocarcinoma and 46 patients with squamous carcinoma were included. In the process of outlining ROI, we found that patients with brain metastases from lung adenocarcinoma often showed multiple or single nodular lesions on MRI, and some lesions may show ring-shaped enhancement with obvious oedema around the lesions, while patients with brain metastases from squamous carcinoma of the lung showed mostly single metastases morphologically and ring-shaped enhancement was common. In our study, among 118 patients with brain metastases from lung cancer, 72 patients with adenocarcinoma and 46 patients with squamous carcinoma were included. In the process of outlining ROI, we found that patients with brain metastases from lung adenocarcinoma often showed multiple or single nodular lesions on MRI, and some lesions may show ring-

shaped enhancement with obvious oedema around the lesions, while patients with brain metastases from squamous carcinoma of the lung showed mostly single metastases morphologically and ring-shaped enhancement was common.

Based on the different expressions of estrogen (ER), progesterone (PR), HER-2 and Ki67, breast cancer can be classified into the following four subtypes: Luminal A, Luminal B, HER-2-positive and triple-negative breast cancer (TNBC). With the advances in diagnostic techniques and surgical, radiological, chemotherapeutic, targeted and endocrine therapies, the incidence of brain metastases from breast cancer has been increasing year by year. Breast cancer patients have about 20–40% probability of developing brain metastases during the course of their disease. Metastatic breast cancer is first diagnosed in about 10% due to the development of neurological symptoms, with a median survival of about 1 month. Among these subtypes, HER2-positive breast cancer has a higher incidence of brain metastases, about 35–50%, followed by TNBC, about 25–46% (12). The data in our study were mainly from a single-centre institution and included 62 patients with breast cancer brain metastases, including 26 TNBCs, 13 HER-2-positive types, 17 Luminal A types, and 16 Luminal B types. In the process of outlining ROI, we found that breast cancer brain metastases have some unique features, multiple brain metastases are common, metastatic sites are mainly in the cerebellum and frontal lobe, meningeal metastases are prone to occur. On MRI, the tumour lesions often



showed solid enhancement and marginal enhancement, and irregular T1WI low-signal and T2WI high-signal oedema bands were seen around the lesions with unclear borders.

### ***Significance of identifying the primary site and limitations***

Approximately 10% of patients with brain metastases are first admitted to the hospital for the development of central nervous system symptoms, such as headache, vomiting, cognitive decline, behavioural changes, sensory disturbances, motor weakness, aphasia, visual disturbances, ataxia, and seizures. Patients presenting with these symptoms will be sent to the emergency department at the first time to improve the cranial CT or MRI examination to find the cranial brain occupation after the surgeon opened and confirmed the diagnosis in the intraoperative pathology. However, patients in this advanced stage of the disease may not benefit much from surgery, and may be able to relieve intracranial hypertension by dehydration or even by radiating the tumour to the brain metastases. However, after this series of invasive operations in a short period of time is difficult to obtain clear histological confirmation, according to the imaging results report further imaging histological analysis can be further extracted from the lesion characteristics, for the admitting physician in the diagnosis and treatment of more information, for the initial assessment of the nature of brain metastatic lesions to play a complementary role. In addition, for patients with dual-origin or even multiple-origin tumours and brain metastases, perfect chest and abdominal CT is not enough to assess the origin of the tumour, even though the incidence of this type of incidence is not high, the number of cases of dual-origin incidence has been found to be increasing in clinical practice year by year. In addition, with the refinement of examination equipment, more and more small lung and breast nodules can be detected, so it is sometimes not easy to accurately determine the source of the primary lesion.

In conclusion, timely and accurate identification and diagnosis of the primary tumour site in cases of brain metastases of unknown primary origin is crucial for prognostic assessment and clinical decision-making. For patients with neurological symptoms as the first diagnosis of multiple origins and secondary brain metastases, there is an urgent need for contrast-enhanced T1WI-based histology and machine learning models, which are expected to achieve non-invasive differential diagnosis of cerebral malignant tumours, and thus alleviate the mental, psychological,

and economic burdens of patients and the pain of invasive examinations, and help to speed up the diagnostic and therapeutic decision-making process.

However, there are several limitations in our research:

- (I) Data imbalance: this single-center retrospective study involves a limited number of cases, focused only on breast cancer for non-lung metastases, lacking a broad pathological and origin scope, which might introduce bias. The absence of reliable external validation and the prevalence disparity between lung and non-lung metastases can cause sample imbalance, affecting model training and minority class performance.
- (II) Reliability and standardization: variability in post-contrast T1-weighted images of lung and non-lung brain metastases might arise from differences in equipment, parameters, operators, image noise, artifacts, and reconstruction algorithms, creating variability and inconsistency. Manual ROI delineation used here may introduce errors, affecting reproducibility and posing a challenge for image reliability and standardization.
- (III) Feature selection and extraction: identifying distinctive features from extensive imaging data is complex. The outcomes can vary significantly based on the chosen feature extraction methods and selection algorithms, necessitating careful selection and evaluation.
- (IV) Curse of dimensionality: with high-dimensional imaging data, the rapid increase in feature numbers can lead to the curse of dimensionality, complicating model training and demanding more computational resources.

### ***Comparison of enhanced T1WI models based on different algorithms***

MRI, as the preferred imaging diagnostic method for intracranial tumors, has gradually been applied in precision individualized diagnosis and treatment of brain metastases. Post-enhancement T1WI is a routine and optimal sequence for diagnosing brain metastases, which can clearly display the lesions and provide information on intra-tumoral vascular distribution, reflecting the heterogeneous nature of tumor blood supply (13). This imaging technique plays a crucial role in the diagnosis and treatment of brain metastases. Research indicates that, compared to non-enhanced T1W sequence, the post-enhancement T1W

sequence in MRI radiomics of brain metastases exhibits advantages in lesion enhancement display and dynamic monitoring, vascular enhancement display, clear lesion boundary display, and histological feature analysis, resulting in higher accuracy and reliability in the diagnosis and evaluation of brain metastases (14).

This study constructed and compared five radiomic models based on post-contrast T1WI to differentiate between lung cancer brain metastases and non-lung cancer brain metastases using 180 brain metastasis cases. The models were built using LR, SVM, KNN, MLP, and LightGBM algorithms. The results indicated that the LightGBM model demonstrated more robust performance, achieving the best metrics in both the training and validation sets with AUC values of 0.875 and 0.866, respectively. LightGBM is a high-performance machine learning framework that is particularly suited for dealing with large-scale and high-dimensional data sets, providing fast and accurate predictions. This study utilized a variety of machine learning models to establish radiomics models based on enhanced T1WI of brain metastases, facilitating further validation and exploration of the potential of MRI radiomics in accurately identifying the origins of brain metastases. This could enhance diagnostic accuracy and streamline the diagnostic and treatment workflow. The LightGBM model, which exhibited the best performance, achieved an AUC of 0.875 (95% CI: 0.819–0.931) in the training set and 0.866 (95% CI: 0.740–0.993) in the validation set. LightGBM adopts a specific decision tree structure and histogram-based approach, providing fast training speed and accuracy. It typically performs well in handling multi-class classification and regression problems, which has been widely used in different fields in recent years due to its efficient processing speed, accurate capture of nonlinear relationships between features, and high interpretability. In the medical field, big data in medicine including structured digital medical records, imaging test images, genetic testing, and pathology image results also provide opportunities for the construction and training of AI models. In oncology research, there have been studies applying LightGBM to clinical decision-making processes such as tumour prediction and differential diagnosis, prognostic risk assessment, and treatment plan recommendation. By mining and analysing medical data, LightGBM may help clinicians achieve more accurate medical decisions. A study has been conducted to construct a prognostic model based on immune-related gene (IRG) expression profile data and clinical data to predict the prognosis of lung

adenocarcinoma patients based on the integrated learning method of LightGBM. Lumbar spine MRI radiomics based on T2WI sequences for osteoporosis detection. Construction of a stacked integrated regression model for quantitative prediction of ER $\alpha$  (estrogen receptor  $\alpha$ ) activity of breast cancer drug candidates to improve the activity of anti-breast cancer drug candidates.

The LightGBM algorithm, a machine learning algorithm based on gradient boosting trees, which can efficiently process large-scale data. The LR and MLP algorithms are linear models and may not perform as well as SVM and LightGBM in handling nonlinear problems. The KNN algorithm is simple but inefficient in processing high-dimensional data. The SVM algorithm has good generalization ability for nonlinear problems, while the LightGBM algorithm can efficiently process large-scale data. Therefore, the selection of appropriate algorithms should be considered comprehensively based on specific problems and data characteristics.

### *The application and development of radiomics in differential diagnosis of diseases*

Radiomics, a technology combining medical imaging and machine learning, enables high-throughput extraction and analysis of numerous advanced quantitative imaging features from medical images obtained through computed tomography (CT), positron emission tomography (PET), or MRI. Its potential contribution to tumor-related decision support in oncology is also growing (15).

Radiomic features can encompass morphological characteristics (e.g., tumor size, shape, contour), textural features (e.g., tumor texture statistics, gray-level co-occurrence matrix), statistical features (e.g., mean, variance, standard deviation), and density features (e.g., pixel density distribution of tumors). These features can reflect the differences between lesion areas and normal tissues, assisting in cancer detection, early diagnosis, differential diagnosis, prognostic assessment, treatment response prediction, and disease status monitoring. Beig *et al.* utilized convolutional neural networks to analyze intratumoral and peritumoral radiomic features of pulmonary nodules (16). They found that the area under the curve (AUC) for distinguishing NSCLC from granulomatous lesions using intratumoral and peritumoral radiomic features was 0.8, with a sensitivity of 72% and a specificity of 76%. This diagnostic performance surpassed the traditional reading mode of radiologists (17). Xie *et al.* employed 1,018 instances

from the LIDC-IDRI public dataset as experimental data, fusing textural, shape, and deep learning features to construct a classifier for diagnosing the benign and malignant nature of pulmonary nodules. The AUC of their operational characteristics reached 0.97 (18). Gibbs *et al.* identified 45 cases of breast cancer by extracting gray-level textural features from dynamic contrast-enhanced MRI (DCE-MRI) images of 79 patients. Combined with logistic regression analysis, the AUC for distinguishing benign and malignant lesions reached 0.80 (19). Lymph node metastasis, a crucial component of the TNM staging system, directly affects the survival prognosis of lung cancer patients. Accurate identification of metastatic lymph nodes in clinical practice is paramount. Radiomics methods may emerge as an effective approach for predicting lung cancer lymph node metastasis. Botta *et al.* constructed a radiomics model using the LASSO algorithm, achieving an AUC of 0.709 for predicting metastatic lymph nodes. This model successfully classified the enrolled population into low- and high-risk groups based on differences in overall survival (20). Zheng *et al.* employed ultrasound-based deep learning radiomics methods to preoperatively predict axillary lymph node status in early breast cancer patients, achieving an AUC of 0.90. This provides a non-invasive imaging technique for predicting the degree of axillary lymph node metastasis in early breast cancer patients (21). With the widespread application of computer technology and artificial intelligence in neuroimaging and clinical practice, it is believed that radiomics will have even more remarkable development prospects.

## Conclusions

The enhanced MRI radiomics model, especially the LightGBM model, can accurately predict the primary lesion types of brain metastases from lung cancer and breast cancer origins. The LightGBM algorithm holds significant clinical utility in differentiating the primary sources of brain metastases. By leveraging advanced machine learning techniques, it efficiently processes intricate medical imaging data alongside clinical records, enabling precise identification of the original tumor sites. This capability empowers clinicians to devise targeted therapies earlier in the treatment journey, enhancing patient outcomes and quality of life. In the future, the impact of LightGBM on brain metastasis research is poised to be transformative. As data volumes continue to grow, the algorithm's performance will further refine, fostering even more accurate predictions

and insights. Additionally, its integration with emerging medical technologies, such as advanced imaging modalities and precision medicine approaches, will pave the way for more personalized and effective treatment strategies. Ultimately, LightGBM's application in this field will foster interdisciplinary collaborations, driving fundamental breakthroughs in our understanding of brain metastasis mechanisms and treatment options.

## Acknowledgments

We sincerely thank the 2024 ESMO ASIA International Conference for providing us a valuable opportunity to display abstract in the form of poster, and we have received positive feedback and suggestions from our peers. In addition, the work was supported by The Fifth Affiliated Hospital of Sun Yat-sen University, IIT-GCP project: a randomised, parallel-controlled multicentre clinical trial on the protective role of calcium folinate in radiotherapy for brain metastases of lung cancer (ChiCTR1900025490).

## Footnote

**Reporting Checklist:** The authors have completed the TRIPOD reporting checklist. Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-1355/rc>

**Data Sharing Statement:** Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-1355/dss>

**Peer Review File:** Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-1355/prf>

**Funding:** The work was supported by a grant from the National Natural Science Foundation of China (No. 82172881) awarded by Z.L.

**Conflicts of Interest:** All authors have completed the ICMJE uniform disclosure form (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-1355/coif>). The authors have no conflicts of interest to declare.

**Ethical Statement:** The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study followed the Declaration of Helsinki (as revised in 2013) and was approved by the ethics committee of The Fifth Affiliated

Hospital of Sun Yat-sen University {Approval No. [2019] Ethical Letter (K03-1)}, and informed consent for this retrospective analysis was waived.

**Open Access Statement:** This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023. *CA Cancer J Clin* 2023;73:17-48.
2. Bae S, An C, Ahn SS, et al. Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features: model development and validation. *Sci Rep* 2020;10:12110.
3. Araujo LH, Horn L, Merritt RE, et al. Cancer of the Lung: Non-Small Cell Lung Cancer and Small Cell Lung Cancer, Editor(s): Niederhuber JE, Armitage JO, Kastan MB, et al., editors. *Abeloff's Clinical Oncology* (Sixth Edition), Elsevier, 2020:1108-1158.e16.
4. Zhu Y, Cui Y, Zheng X, et al. Small-cell lung cancer brain metastasis: From molecular mechanisms to diagnosis and treatment. *Biochim Biophys Acta Mol Basis Dis* 2022;1868:166557.
5. Hollon TC, Pandian B, Adapa AR, et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat Med* 2020;26:52-8.
6. Taphoorn MJ, Heimans JJ, Kaiser MC, et al. Imaging of brain metastases. Comparison of computerized tomography (CT) and magnetic resonance imaging (MRI). *Neuroradiology* 1989;31:391-5.
7. Bera K, Brame N, Gupta A, et al. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol* 2022;19:132-46.
8. Jung WS, Park CH, Hong CK, et al. Diffusion-Weighted Imaging of Brain Metastasis from Lung Cancer: Correlation of MRI Parameters with the Histologic Type and Gene Mutation Status. *AJNR Am J Neuroradiol* 2018;39:273-9.
9. Benusiglio PR, Fallet V, Sanchis-Borja M, et al. Lung cancer is also a hereditary disease. *Eur Respir Rev* 2021;30:210045.
10. Zhu D, Shao Y, Yang Z, et al. Magnetic resonance imaging characteristics of brain metastases in small cell lung cancer. *Cancer Med* 2023;12:15199-206.
11. Sereno M, Hernandez de Córdoba I, Gutiérrez-Gutiérrez G, et al. Brain metastases and lung cancer: molecular biology, natural history, prediction of response and efficacy of immunotherapy. *Front Immunol* 2024;14:1297988.
12. Raghavendra AS, Ibrahim NK. Breast Cancer Brain Metastasis: A Comprehensive Review. *JCO Oncol Pract* 2024;20:1348-59.
13. Zhang M, Young GS, Chen H, et al. Deep-Learning Detection of Cancer Metastases to the Brain on MRI. *J Magn Reson Imaging* 2020;52:1227-36.
14. Batouli A, Kanal E, Gholamrezanezhad A, et al. T1-weighted parenchyma attenuated inversion recovery: A novel sequence that improves contrast ratio of enhancing brain lesions. *Diagn Interv Imaging* 2018;99:29-35.
15. Guiot J, Vaidyanathan A, Deprez L, et al. A review in radiomics: Making personalized medicine a reality via routine imaging. *Med Res Rev* 2022;42:426-40.
16. Beig N, Khorrami M, Alilou M, et al. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas. *Radiology* 2019;290:783-92.
17. Liu Y, Hou Z, Li X, et al. Pulmonary nodule detection method based on convolutional neural network. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi* 2019;36:969-77.
18. Xie Y, Zhang J, Xia Y, et al. Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. *Information Fusion* 2018;42:102-10.
19. Gibbs P, Turnbull LW. Textural analysis of contrast-enhanced MR images of the breast. *Magn Reson Med* 2003;50:92-8.
20. Botta F, Raimondi S, Rinaldi L, et al. Association of a CT-Based Clinical and Radiomics Score of Non-Small Cell Lung Cancer (NSCLC) with Lymph Node Status and Overall Survival. *Cancers (Basel)* 2020;12:1432.
21. Zheng X, Yao Z, Huang Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun* 2020;11:1236.

**Cite this article as:** Xie Y, Li X, Yang S, Jia F, Han Y, Huang M, Chen L, Zou W, Deng C, Liang Z. Radiomics models using machine learning algorithms to differentiate the primary focus of brain metastasis. *Transl Cancer Res* 2025;14(2):731-742. doi: 10.21037/tcr-24-1355