© Pol J Radiol 2025; 90: e475-e483
DOI: https://doi.org/10.5114/pjr/207440

Received: 16.06.2025
Accepted: 23.06.2025
Published: 30.09.2025

Polish Journal of Radiology

http://www.polradiol.com

Original paper

Can radiomics in brain magnetic resonance imaging predict the mutational status of primary lung cancer based on brain metastasis?

Beyza Nur Kuzan[1,A,E,F], Can Ilgın[2,C,E], Gonca Gül Geçmen[3,B], Naciye Işık[4,B], Hediye Pınar Günbey[1,D], Murat Emeç[5,C,D,E]

[1]Department of Radiology, Kartal Dr. Lütfi Kırdar City Hospital, Istanbul, Turkey
[2]Department of Radiation Oncology, Istanbul University School of Medicine, Istanbul, Turkey
[3]Department of Pathology, Kartal Dr. Lütfi Kırdar City Hospital, Istanbul, Turkey
[4]Department of Radiation Oncology, Kartal Dr. Lütfi Kırdar City Hospital, Istanbul, Turkey
[5]Department of Computer Science Application and Research, Istanbul University, Istanbul, Turkey

Abstract

Purpose: Lung cancer is one of the most common types of cancer, and the presence of brain metastases has a significant impact on the clinical course and prognosis. EGFR, BRAF, ALK, and ROS1 mutations have previously been identified in lung cancer, and knowing the tumour mutation status is important for molecular therapy. In our study, we investigated the performance of radiomics in predicting the status of brain metastases detected by brain magnetic resonance imaging (MRI), a noninvasive method, in with brain metastases patients diagnosed with lung cancer.

Material and methods: Lung cancer cases with brain metastasis in our hospital between 2014 and 2024 were analysed retrospectively. Histopathological data were obtained from tissue biopsy results, and EGFR, BRAF, ALK, and ROS1 mutation status were recorded. A total of $N = 84$ patients were included in the study, and 107 original radiomics parameters were obtained from the segmentation files extracted from the patient images. Due to the class unbalance, the performance of the model was tested using the stratified folding method.

Results: Five (6.02%) of the patients had EGFR, 3 (4.17%) had ALK, and 2 (2.78%) had ROS1 mutations. Model 1 used for EGFR mutation prediction showed high performance with 93.82% accuracy, Model 2 used for ALK with 84.76% accuracy, and Model 3 used for ROS1 with 79.33% accuracy.

Conclusion: Our study showed that EGFR mutations, in particular, can be detected with high accuracy by radiomics in lung cancer patients with brain metastases without additional invasive procedures.

Key words: lung cancer, brain, metastasis, magnetic resonance imaging, radiomics.

Introduction

Lung cancer is the second most common type of cancer diagnosed, and non-small cell lung cancer (NSCLC) is the most common type, representing about 85% of cases [1]. Brain metastases develop in up to 90% of all cases of lung cancer and are associated with a shorter survival time and a poor prognosis [2-4]. Survival and prognosis are worse in cases with brain metastases. Prior to treatment in lung cancer, molecular testing for the epidermal growth factor receptor (EGFR), v-Raf murin sarcoma viral oncogene homologue B (BRAF), anaplastic lymphoma kinase (ALK), and ROS oncogene 1 (ROS1) genes are commonly used, and treatment options are based on the mutation status detected in the cancer. The treatment options for cases of non-small cell lung cancer have recently been expanded, and molecularly targeted therapies have been added to the protocol. Pharmacological agents that

Correspondence address:
Beyza Nur Kuzan, MD, Department of Radiology, Kartal Dr. Lütfi Kırdar City Hospital, D-100 Güney Yanyol No: 47 Cevizli Mevkii, 34865, Kartal, Istanbul, Turkey, e-mail: drbeyzauzun@hotmail.com

Authors' contribution:
A Study design · **B** Data collection · **C** Statistical analysis · **D** Data interpretation · **E** Manuscript preparation · **F** Literature search · **G** Funds collection

may enter the central nervous system in NSCLC patients have been reported to have a positive effect on survival in the treatment of brain metastases. Therefore, the availability of molecular information is essential for the planning of treatment and the prediction of survival [5].

However, invasive biopsy or surgical resection of metastases for molecular diagnosis is not practical because brain metastases are usually small and widely distributed. Consequently, most metastatic brain lesions are detected by magnetic resonance imaging (MRI) without histopathological confirmation. Thus, non-invasive imaging modalities are currently the preferred approach to evaluate the mutational status of lung cancer brain metastasis [6,7].

Radiomic analysis can be used to comprehensively study spatially and temporally heterogeneous tumours through the extraction of a large number of points from radiological images [8]. Conventional radiologic imaging evaluates brain lesions for size, location, signal characteristics, and peritumoral area. However, the human eye may fail to discriminate features related to tumour texture, shape, and image intensity in radiological analysis [9,10]. Since microstructural features of tumour tissue can be determined by radiomic analysis, it is a potentially useful tool for the identification of genetic mutations and personalised treatment protocols.

The focus of radiomics may be to obtain high-dimensional features in order to capture all of the features of the image under investigation [11]. In radiomics, first-order methods typically rely on histogram analysis based on analysis of individual voxels [12,13]. Second-order methods are typically based on texture analysis and reveal statistical relationships between voxels as a function of contrast values [14,15]. Higher-order methods can extract repetitive or non-repetitive patterns using various filters [16,17]. Examples of higher-order methods include Gaussian bandpass filtering and Minkowski filtering [18,19].

In our study, we hypothesise that radiomics of tumour segmentation on post-contrast T1-weighted (W) imaging in lung cancer patients with brain metastases can be used to predict EGFR, BRAF, ALK, and ROS1 mutation status on contrast-enhanced brain MRI. Therefore, decision tree machine learning models were developed to determine mutation status. The models were optimised considering the imbalanced dataset problem and evaluated using 5-fold cross-validation with the layered convolutional method.

Material and methods

Study design

A retrospective study was performed on lung cancer cases with brain metastasis in our hospital between January 2014 and April 2024. The study was approved by the institutional review board. The data of the study population were collected in a retrospective manner from the hospital information system. Histopathologic data were extracted from pathology reports after tissue biopsy, noting tumour subtype and EGFR, BRAF, ALK, and ROS1 mutation status. The presence of brain metastases was confirmed on post-contrast T1A series obtained after using gadolinium-based contrast. Patients were excluded from the study for the following reasons: (1) previous neurosurgery (n = 25), (2) previous brain radiotherapy (n = 21), (3) presence of another primary tumour (n = 7), (4) poor image quality (n = 5), (4) EGF, BRAF, ALK, and ROS1 mutation status not checked (n = 42), (5) failure to select detectable brain metastases (n = 3), and (6) cases without contrast-enhanced brain MRI (n = 3). In addition, due to the difficulty of assessment, metastases with a largest diameter of less than 5 mm were not included in the study. After all exclusions, 84 patients were included in the study. For cases with multiple metastases, the largest single metastasis was included in the study. Informed consent was waived due to the retrospective nature of the study. All data were made completely anonymous by masking personal information.

Pathological mutation analysis

The specimens were fixed in 10% formalin and cut into 5 to 10 mm thick sections. To evaluate tumour morphology and expression of IHC markers, haematoxylin and eosin (H&E) slides and IHC were examined by a thoracic pathologist. All cases were staged in accordance with the American Joint Committee on Cancer (AJCC) 8th edition.

The AmoxyDx EGFR 29 Mutation Detection Kit is a real-time PCR assay for the qualitative detection of 29 somatic mutations in exons 18, 19, 20, and 21 of the *EGFR* gene in human genomic DNA extracted from formalin-fixed paraffin-embedded tumour tissue, and it was used for mutation analysis.

The AmoyDx BRAF Mutation Detection Kit is a real-time PCR test for the qualitative detection of V600E, V600E2, V600K, V600D, V600D2, V600A, and V600R mutations in the BRAF gene, and it was used for mutation analysis.

It is designed for the qualitative detection of translocations involving the *ALK* and *ROS1* genes by fluorescence *in situ* hybridisation (FISH). Fluorescently labelled DNA fragments and complementary target DNA strands are denatured together and then allowed to bind to each other during the hybridisation process. This method has been used to investigate the presence of mutations.

Imaging process

A total of N = 84 patients were included in the study. A flow chart summarising the criteria for the study is shown in Figure 1. The radiomic parameters from patient images and segmentation files in .nrrd format were extracted with PyRadiomics (version 3.1.0, with Python

3.11.7) [20]. The textural features, including Gray Level Co-occurrence Matrix (GLCM, 24 features), Gray Level Run Length Matrix (GLRLM, 16 features), Gray Level Size Zone Matrix (GLSZM, 16 features), and Neighboring Gray Tone Difference Matrix (NGTDM, 5 features), with a total of n = 107 original radiomics parameters and n = 13 diagnostic parameters, were extracted. Data were recorded in .csv format, edited in Microsoft Excel software, and imported to Stata 15.1 software (StataCorp 4905 Lakeway Drive, College Station, Texas 77845 USA). The patient database and radiomics databases were combined by using Stata 15.1 via established key variables. All data visualisations, including biplots, ROC curves, and heat maps, were performed with Stata 15.1. Heat maps (including histograms and correlation matrices) were created with the heatplot module [21]. The scheme for visualising metastasis in postcontrast T1W sequences on MRI, segmentation, extraction radiomics features, and making predictions is shown in Figure 2.

In this study, decision tree-based machine learning models were developed to detect EGFR, ALK, and ROS mutations. The models were optimised by considering the unbalanced dataset and evaluated by using Stratified Fold method 5-fold cross validation [22]. For the preprocessing step, the missing data were imputated with appropriate methods, where the observations with missing target feature were dropped. Diagnostic features, including size and spacing features of image and mask, bounding box, centre of mass and centre of index features of mask, were used intensively in the data preprocessing step. In the feature engineering step, physical dimensions and volume were calculated using size and spacing diagnostic features of the original image. Centre of mass and distribution features were calculated using centre of mass and centre of mass index of mask. The features were normalised to improve model performance. For each model, true positive, true negative, false positive, and false negative observations were classified with confusion matrices. Accuracy, precision, recall, and F1-score were reported for each model. The dataset included diagnostic and original features, in addition of engineered features. According to the feature selection step, features with highest importance were Gray Level Co-occurrence Matrix Joint Average (0.125), Maximum 2D diameter (Slice) (0.075), Surface Area (0.067), Neighboring Gray Tone Difference Matrix Complexity (0.050), and Gray Level Co-occurrence Matrix Cluster Prominence (0.045) (Table 1). For the unbalanced dataset problem, the hyperparameters of maximum depth = 4, minimum samples split = 2, minimum samples leaf = 6, and balanced class weight were used [23]. The stratified fold method was used to stratify the dataset by keeping class unbalance for each stratum. This method was employed to evaluate the model performance in unbalanced datasets correctly.

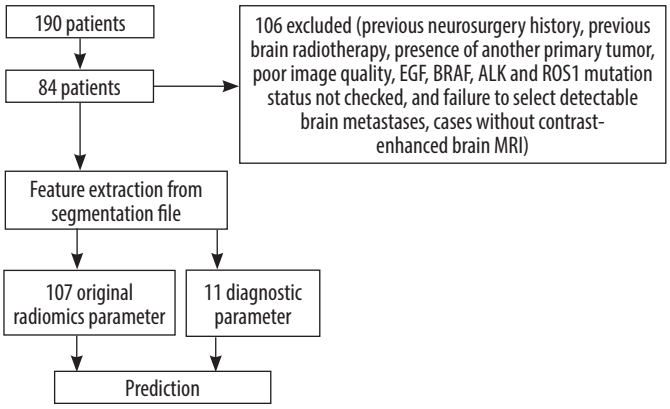

Figure 1. Flowchart of the study population and process

Statistical analysis

The normal distribution of variables was tested with histogram, normal quantile graph, Kolmogorov-Smirnov, and skewness kurtosis tests. The variables with normal distribution were presented with mean ± standard deviation, and the variables without normal distribution were

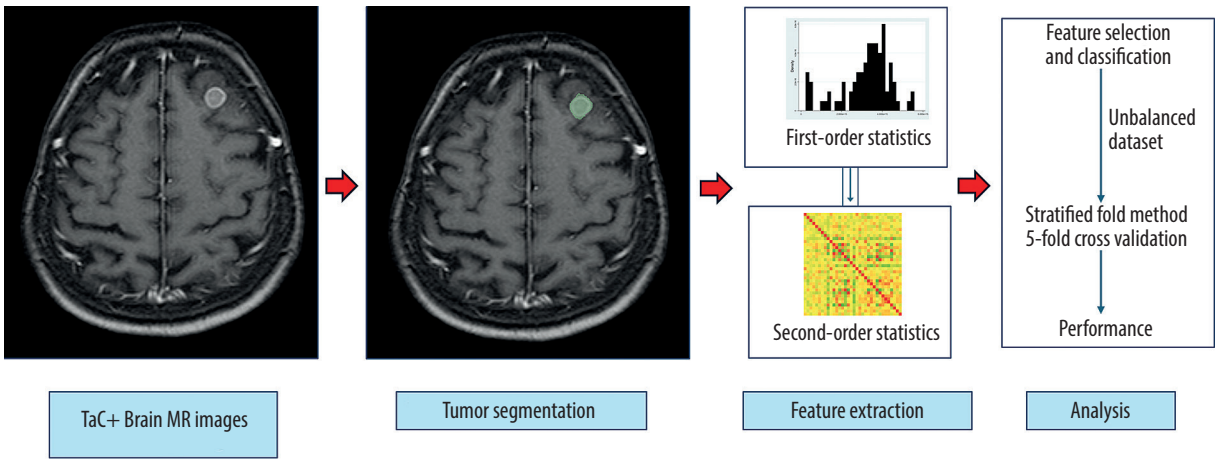

Figure 2. Schema for brain tumour segmentation and radiomic feature extraction. **A**) Contrast-enhanced T1W sequence to determine metastasis localisation. **B**) Segmented enhancing tumour on contrast-enhanced T1W sequence. **C**) Radiomic feature extraction with first- and second-order statistics. **D**) Analysis

Table 1. Selected features are shown

Feature	Importance
og_glcm_JointAverage	0.125303
og_shp_Max2DDiamSlice	0.075222
og_shp_SurfaceArea	0.066667
og_ngtdm_Complexity	0.049600
og_glcm_ClusterProminence	0.044888
og_glszm_LargeAreaE	0.040694
og_shp_VoxVol	0.036880
og_glszm_SmallAreaHighGylvlE	0.024892
og_shp_SurfaceVolRatio	0.024245
og_glrlm_sh_RunHighGylvlE	0.023736
og_1_order_Entropy	0.022761
og_ngtdm_Busyness	0.020985
og_gldm_DepEntropy	0.020121
og_ngtdm_Strength	0.018255
og_1_order_MeanAbsolutedev	0.016667
dx_Imageog_Max	0.015079
dx_Maskog_C_OfMassIndex_Mean	0.015079
og_glcm_Imc1	0.014844
og_glcm_Id	0.014541
dx_Maskog_VoxNum	0.014541
og_glszm_SmallAreaE	0.014137
og_1_order_Max	0.014137
og_glszm_GylvlNonUni	0.014137
og_glcm_MaxProbability	0.014137
og_glcm_Autocorrelation	0.013818
og_shp_Flatness	0.013571
og_glcm_ClusterShade	0.013212
og_gldm_HighGylvlE	0.013102
dx_Maskog_C_OfMass_Mean	0.012723
og_ngtdm_Contrast	0.012338
og_glcm_DifferenceEntropy	0.012315
og_gldm_DepNU	0.011548
og_glrlm_RunVar	0.011515
age	0.011515
og_1_order_IQR	0.011310
og_glszm_SizeZoneNUN	0.011310
og_glrlm_RunEntropy	0.010608
og_glszm_ZoneEntropy	0.010560
og_glcm_JointEntropy	0.009481
dx_Maskog_BoundingBox_Volume_mm3	0.008985
og_1_order_Var	0.008796
Image71	0.008636
og_1_order_Range	0.008064
og_glcm_Imc2	0.007499
og_shp_Sphericity	0.006781
og_1_order_Energy	0.006473
og_gldm_LargeDepE	0.005758
og_1_order_RootMeanSquared	0.004398
og_glcm_ClusterTendency	0.004353
gender	0.002848

presented with median and IQR values, in addition to minimum and maximum values. Categorical variables were presented as counts and percentages. The distribution of variables without normal distribution among independent groups were tested with Mann-Whitney *U* test. For the discriminatory analysis, radiomic variables showing significant difference were selected. Kth nearest neighbour method was used for discriminant analysis, where different *k*-values were used iteratively to minimise the error rate. Leave-one-out tables were used for cross validation in discriminant analysis. Sensitivity, specificity, and positive and negative predictive values were calculated for each model and presented with 95% confidence intervals. A *p*-value less than 0.05 was considered statistically significant.

Results

The mean age of the patients was 67.61 ± 9.74 years, and most of the patients were male (n = 70, 83.33%). The diagnosis of most of the patients was lung adenocarcinoma (n = 77, 91.67%); however, 2 cases were diagnosed as lung squamous cell carcinoma, 1 case was large cell neuroendocrine carcinoma, and 1 case was diagnosed as cystic mesothelioma. Additionally, 3 cases were diagnosed with NSCLC that could not be subtyped [24]. Other demographical and clinical characteristics of the patients are shown in Table 2.

The confusion matrices for each model are presented in Table 3. The model for EGFR had 32 true negative and 2 true positive observations out of 34 patients, where no patients were classified as false negative or false positive. Similarly, the model for ALK classified 28 patients as true negative, and one patient as true positive without any false negative or false positive classifications. The model for ROS1 correctly classified all true negative patients (n = 28), but one patient was misclassified as false positive.

The performance metrics of 3 models predicting EGFR, ALK, and ROS1 mutations is presented in Table 4. The decision tree model for EGFR has the highest diagnostic performance, including an accuracy of 93.82%, precision of 97.85%, recall of 93.82%, and F-1 score of 95.10%. While the model classified all 14 true negative patients correctly, only one patient has been classified as true posi-

tive out of 2 positive patients. The model for ALK has an accuracy of 84.76%, precision of 93.58%, recall of 84.76%, and F-1 score of 88.48%. Similarly to the EGFR model, the ALK model classified all negative patients (n = 11) correctly; however, 2 patients out of 3 positives were classified as false negative. Finally, the ROS1 model had accuracy of 79.33%, precision of 94.20%, recall of 79.33%, and F-1 score of 85.91%. Similarly to the previous models, the model predicted true negative patients (n = 12) perfectly, but the model could not detect true positive patients and misclassified them as negative (n = 2 false negative). Receiver operating characteristic (ROC) curves for classifying EGFR, ALK, and ROS1 mutation status are shown in Figure 3. Heatmaps of the EGFR, ALK, and ROS1 mutation status of the cases are shown in Figure 4.

Discussion

This study shows that radiomics analysis can be used to predict the tumour subtype and EGFR, ALK, and ROS1 mutation status in NSCLC cases with brain metastases. Although the EGFR model has the highest diagnostic performance among the models, all 3 models have high precision values, ranging from 93.58% to 97.85%, which reflects the low rate of false positive cases. The stratified fold method enabled us to evaluate the model performance by keeping the class unbalance [22]. Evidence shows that decision tree models are effective in unbalanced datasets with hyperparameter optimisation.

Radiomics is a computational method for the transformation of tumour images into a large number of quantitative features. Here, we developed machine learning models to classify the molecular mutation status of EGFR (more commonly reported) and ALK and ROS1 (less commonly reported) in lung cancer. Thus, we have demonstrated that MRI and radiomics analysis of brain metastases can be an alternative, non-invasive method for the classification of EGFR, ALK, and ROS1 mutations in lung cancer patients.

There is no study in the literature on the analysis of ROS1 mutations using artificial intelligence algorithms based on the radiological features of the tumour in cases of primary lung cancer. In this regard, our study can be considered as the first study to investigate the less common mutations such as ROS1. Mayer *et al.*'s study [25], conducted on pathological diagnosis prediction of ALK and ROS1 mutations using deep learning algorithms, reported specificities of 100% and 98.48%, respectively. In the study of Chen *et al.* [26], which used imaging-based radiomics and machine learning (ML) in the field of radiology, it focused on EGFR, ALK, and KRAS mutations

Table 2. Demographic and clinical characteristics of the patients

Demographic and clinical characteristics	
Age (years), mean ± SD; range	67.61 ± 9.74; 36-94
Gender (male), *n* (%)	70 (83.33)
Histology, *n* (%)	
Lung adenocarcinoma	77 (91.67)
NSCLC	3 (3.57)
Lung squamous cell carcinoma	2 (2.38)
Large cell neuroendocrine carcinoma	1 (1.19)
Cystic mesothelioma	1 (1.19)
Number of metastases, *n* (%)	
1	31 (59.62)
2	12 (23.08)
3	5 (9.62)
4	1 (1.92)
5	1 (1.92)
7	1 (1.92)
> 10	1 (1.92)
Largest metastasis (in mm), median (IQR); range	18 (12.5); 3-60
Location of metastases, *n* (%)	
Frontal	25 (48.08)
Cerebellar	15 (28.85)
Occipital	5 (9.62)
Parietal	2 (3.85)
Temporal	2 (3.85)
Brain stem	1 (1.92)
Corpus callosum	1 (1.92)
Thalamic	1 (1.92)
Mutations, *n* (%)	
EGFR	5 (6.02)
BRAF	0 (0)
ALK	3 (4.17)
ROS1	2 (2.78)

NSCLC – non-small cell lung cancer, EGFR – epidermal growth factor receptor, BRAF –v-Raf murin sarcoma viral oncogene homologue B, ALK – anaplastic lymphoma kinase, ROS1 – ROS oncogene 1, mm – millimetre

Table 3. Metrics for model performance (5-fold cross validation)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Test counts
EGFR (DT-HPT)	93.82	97.85	93.82	95.10	0 : 15, 1 : 1
ALK (DT-HPT)	84.76	93.58	84.76	88.48	0 : 13, 1 : 1
ROS1(DT-HPT)	79.33	94.20	79.33	85.91	0 : 13, 1 : 1

Table 4. Model performance metrics (5-fold cross validation)

Model		True negative	True positive
EGFR	Negative prediction	14 (TN)	1 (FN)
	Positive prediction	0 (FP)	1 (TP)
ALK	Negative prediction	11 (TN)	2 (FN)
	Positive prediction	0 (FP)	1 (TP)
ROS1	Negative prediction	12 (TN)	2 (FN)
	Positive prediction	0 (FP)	0 (TP)

EGFR – epidermal growth factor receptor, BRAF – v-Raf murin sarcoma viral oncogene homologue B, ALK –anaplastic lymphoma kinase, ROS1 – ROS oncogene 1, DT-HPT – Decision Tree-Hyperparameter Tuning

in brain metastatic NSCLC. The specificities for EGFR, ALK, and KRAS mutations were 90.6%, 88.7%, and 97.7%, respectively [26]. Another study by Haim *et al.* [27] used deep learning (DL) to predict EGFR mutant status in NSCLC brain metastasis with a specificity of 97.7%.

Studies on the determination of primary lung cancer subtypes by non-invasive techniques based on radiomic features of tumour tissue have been reported in the literature [28,29]. Previously, Li *et al.* [30] used quantitative radiomic features from MR images of brain metastases to predict pathologic subtypes of primary lung cancers, reporting a 97.8% specificity for differentiating small-cell and non-small-cell lung cancers in their back-propagation artificial neural network (BP-ANN) model. The literature also reported the prediction of primary cancers based on radiomic analysis of brain metastases. Discrimination of primary tumour diagnoses such as lung cancer-breast cancer, lung cancer-melanoma have been reported with successful results [31,32].

Identification of common oncogenic mutations has implications for prognostic and therapeutic strategies in cancer treatment. Moreover, lung cancer patients have been reported to be more prone to brain metastasis with EGFR mutations and ALK translocations [33]. Therefore, pre-treatment mutation detection is helpful in predicting metastatic potential and prognosis. Tyrosine kinase inhibitors and ALK inhibitors, which target mutations during the treatment process, are known to positively affect overall survival in patients with EGFR and ALK alterations, respectively [7]. Another molecular subset is represented by the newly identified ROS1 mutation in lung adenocarcinoma, which has almost no overlap with the known driver mutations [34,35]. The MET/ALK/ROS1 inhibitor crizotinib has reportedly shown impressive clinical activity in patients with advanced ROS1-positive lung cancer [36]. In light of the above, the assessment of the metastatic status of the primary tumours using radiomics has become one of the most important steps in patient management and treatment. Radiomics plays a key role in this process by providing information through extraction and processing of visual features that are not discriminable by human eye [37,38].

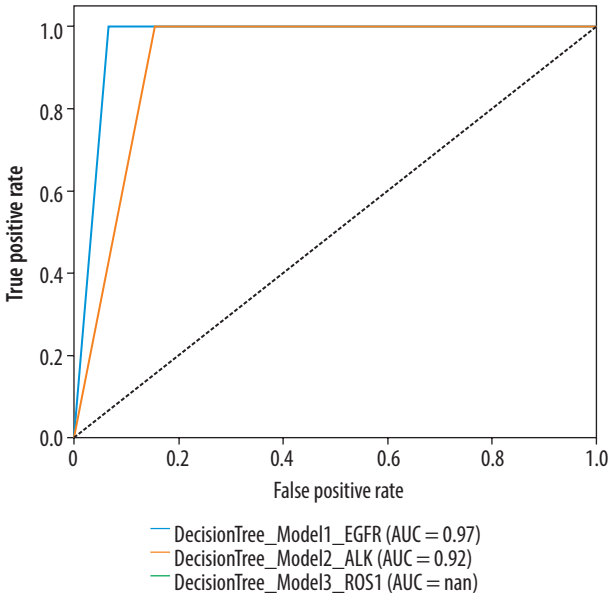

Figure 3. ROC-AUC curves for EGFR, ALK, and ROS1 models

Recently, an increasing number of studies have been performed to determine the mutation status in the primary tumour using radiomics [39-41]. Some of these studies used image features of the primary tumour in chest computed tomography and considered AUC values to estimate mutation status. Most of the studies in this field have used radiomic features extracted from chest CT images, and the highest AUC value reported among them was 0.89 in the study by Gevaert *et al.* [42]. More meaningful results can be obtained by estimating the radiomic features of the primary tumour and metastasis using the multimodality approach. Moreover, new results may be obtained by evaluating radiologic and pathologic features together, using different artificial intelligence algorithms [39,42].

The geratest limitation of this study is the limited number of positive observations for each mutation. Since the number of mutation-positive patients for *EGFR*, *ALK*, and *ROS1* genes was 5 or less, the positive predictive values of the discriminatory models were very low. Because the cases included in the study did not have BRAF mutation, it was not possible to perform radiomics for this mutation. Furthermore, the number of patients was limited with listwise deletion in case of missing target variable, 12 patients for ALK and 12 patients for ROS1 mutation. It is clear that further studies with larger case groups are needed to predict mutations associated with brain metastases in NSCLC with the help of radiomics.

In our study, the mutation analysis was performed on a sample taken from the primary tissue of lung cancer, and there is no histopathologic definition of brain metastasis. Recent studies have reported no overlap in EGFR mutation status between primary and metastatic tissues [43,44]. According to a meta-analysis, the EGFR incompatibility rate between central nervous system metastasis and primary tumour is 17.26% [45]. This situation is based on the different genetic profiles of cancer cells in the

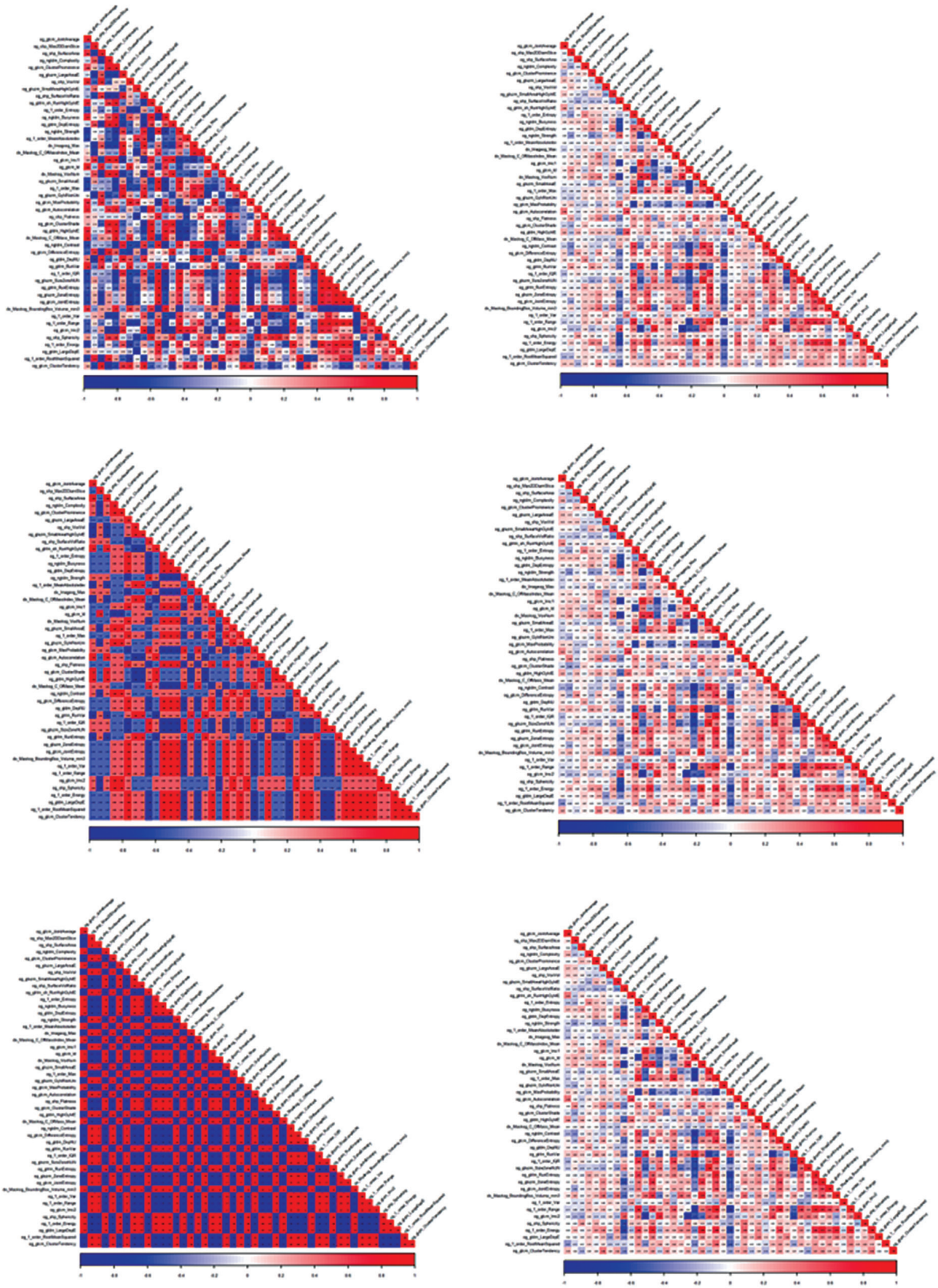

Figure 4. Heat maps for correlation analysis for feature selection of EGFR, ALK, and ROS1 mutations are shown (red colour refers to positive correlations, and blue colour refers to negative correlations; different colour depth indicates different values of correlation coefficients)

literature [46]. In daily practice, the second interventional procedure for tissue diagnosis of brain metastasis in cancer cases with primary tissue diagnosis may increase the morbidity and workload in the case. Therefore, we believe our study provides results that will influence routine clinical outcomes and facilitate treatment.

This study demonstrates the potential of machine learning models in predicting genetic mutations using radiomic features. In particular, the prediction of EGFR mutations was highly successful. Future studies aim to further improve the performance by expanding the dataset and using deep learning methods.

Disclosures

1. Institutional review board statement: The study was approved by the Kartal Dr. Lütfi Kirdar City Hospital Scientific Research Ethics Committee. Decision Number: 2024/010.99/J/8/ Date: 29.04.2024.
2. Assistance with the article: None.
3. Financial support and sponsorship: None.
4. Conflicts of interest: None.

References

1. Nolan C, Deangelis LM. Overview of metastatic disease of the central nervous system. Handb Clin Neurol 2018; 149: 3-23.
2. Ali A, Goffin JR, Arnold A, Ellis PM. Survival of patients with non-small-cell lung cancer after a diagnosis of brain metastases. Curr Oncol 2013; 20: e300-e306. DOI: 10.3747/co.20.1481.
3. Soria JC, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, et al. Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. N Engl J Med 2018; 378: 113-125.
4. AACR Project GENIE Consortium. AACR Project GENIE: powering precision medicine through an international consortium. Cancer Discov 2017; 7: 818-831.
5. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009; 361: 947-957.
6. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 2012; 13: 239-246.
7. Di Lorenzo R, Ahluwalia MS. Targeted therapy of brain metastases: latest evidence and clinical implications. Ther Adv Med Oncol 2017; 9: 781. DOI: 10.1177/1758834017736252.
8. Bozzetti C, Tiseo M, Lagrasta C, Nizzoli R, Guazzi A, Leonardi F, et al. Comparison between epidermal growth factor receptor (EGFR) gene expression in primary non-small cell lung cancer (NSCLC) and in fine-needle aspirates from distant metastatic sites. J Thorac Oncol 2008; 3: 18-22.
9. Kuo MD, Jamshidi N. Behind the numbers: decoding molecular phenotypes with radiogenomics – guiding principles and technical considerations. Radiology 2014; 270: 320-325.
10. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014; 5: 4006. DOI: 10.1038/NCOMMS5006.
11. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012; 48: 441-446.
12. Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. Radiother Oncol 2016; 119: 480-486.
13. Thawani R, McLane M, Beig N, Ghose S, Prasanna P, Velcheti V, Madabhushi A. Radiomics and radiogenomics in lung cancer: a review for the clinician. Lung Cancer 2018; 115: 34-41.
14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016; 278: 563-577.
15. Bhargava R, Madabhushi A. Emerging themes in image informatics and molecular analysis for digital pathology. Annu Rev Biomed Eng 2016; 18: 387-412.
16. Galloway MM. Texture analysis using gray level run lengths. Computer Graphics and Image Processing 1975; 4: 172-179.
17. Wu H, Sun T, Wang J, Li X, Wang W, Huo D, et al. Combination of radiological and gray level co-occurrence matrix textural features used to distinguish solitary pulmonary nodules by computed tomography. J Digit Imaging 2013; 26: 797-802.
18. Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, et al. Defining the biological basis of radiomic phenotypes in lung cancer. Elife 2017; 6: e23421. DOI: 10.7554/ELIFE.23421.
19. Larkin TJ, Canuto HC, Kettunen MI, Booth TC, Hu DE, Krishnan AS, et al. Analysis of image heterogeneity using 2D Minkowski functionals detects tumor responses to treatment. Magn Reson Med 2014; 71: 402-410.
20. Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017; 77: e104-e107. DOI: 10.1158/0008-5472.CAN-17-0339.
21. Jann B. Relative distribution analysis in Stata. Available at: http://ideas.repec.org/p/bss/wpaper/37.htmlhttp://econpapers.repec.org/paper/bsswpaper/37.htmTel.+41 (Accessed: 06.10.2024).
22. Szeghalmy S, Fazekas A. A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning. Sensors 2023; 23: 2333. DOI: 10.3390/S23042333.
23. Gomes Mantovani R, Horváth T, Rossi ALD, Cerri R, Barbon S Jr, Vanschoren J, de Carvalho ACPLF. Better trees: an empirical study

on hyperparameter tuning of classification decision tree induction algorithms. Data Min Knowl Discov 2018; 38: 1364-1416.
24. Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. J Thorac Oncol 2022; 17: 362-387.
25. Mayer C, Ofek E, Fridrich DE, Molchanov Y, Yacobi R, Gazy I, et al. Direct identification of ALK and ROS1 fusions in non-small cell lung cancer from hematoxylin and eosin-stained slides using deep learning algorithms. Mod Pathol 2022; 35: 1882-1887.
26. Chen BT, Jin T, Ye N, Mambetsariev I, Daniel E, Wang T, et al. Radiomic prediction of mutation status based on MR imaging of lung cancer brain metastases. Magn Reson Imaging 2020; 69: 49. DOI: 10.1016/J.MRI.2020.03.002.
27. Haim O, Abramov S, Shofty B, Fanizzi C, DiMeco F, Avisdris N, et al. Predicting EGFR mutation status by a deep learning approach in patients with non-small cell lung cancer brain metastases. J Neurooncol 2022; 157: 63-69.
28. Zhao H, Su Y, Wang M, Lyu Z, Xu P, Jiao Y, et al. The machine learning model for distinguishing pathological subtypes of non-small cell lung cancer. Front Oncol 2022; 12. DOI: 10.3389/fonc.2022.875761.
29. Fang W, Zhang G, Yu Y, Chen H, Liu H. Identification of pathological subtypes of early lung adenocarcinoma based on artificial intelligence parameters and CT signs. Biosci Rep 2022; 42: BSR20212416. DOI: 10.1042/BSR20212416.
30. Li Z, Mao Y, Li H, Yu G, Wan H, Li B. Differentiating brain metastases from different pathological types of lung cancers using texture analysis of T1 postcontrast MR. Magn Reson Med 2016; 76: 1410-1419.
31. Béresová M, Larroza A, Arana E, Varga J, Balkay L, Moratal D. 2D and 3D texture analysis to differentiate brain metastases on MR images: proceed with caution. MAGMA 2018; 31: 285-294.
32. Ortiz-Ramon R, Larroza A, Arana E, Moratal D. A radiomics evaluation of 2D and 3D MRI texture features to classify brain metastases from lung cancer and melanoma. Annu Int Conf IEEE Eng Med Biol Soc 2017; 2017: 493-496.
33. Rangachari D, Yamaguchi N, VanderLaan PA, Folch E, Mahadevan A, Floyd SR, et al. Brain metastases in patients with EGFR-mutated or ALK-rearranged non-small-cell lung cancers. Lung Cancer 2015; 88: 108-111.
34. Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell 2007; 131: 1190-1203.
35. Bergethon K, Shaw AT, Ou SHI, et al. ROS1 rearrangements define a unique molecular class of lung cancers. J Clin Oncol 2012; 30: 863-870.
36. Shaw AT, Camidge DR, Engelman JA, Clark J, Tye L, Wilner K, et al. Clinical activity of crizotinib in advanced non-small cell lung cancer (NSCLC) harboring ROS1 gene rearrangement. J Clin Oncol 2012; 10: 15. DOI: 10.1200/JCO.2012.30.15_SUPPL.7508
37. Avanzo M, Stancanello J, Pirrone G, Sartor G. Radiomics and deep learning in lung cancer. Strahlenther Onkol 2020; 196: 879-887.
38. Binczyk F, Prazuch W, Bozek P, Polanska J. Radiomics and artificial intelligence in lung cancer screening. Transl Lung Cancer Res 2021; 10: 1186-1199.
39. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. Clin Lung Cancer 2016; 17: 441-448.e6. DOI: 10.1016/J.CLLC.2016.02.001.
40. Zhang L, Chen B, Liu X, Song J, Fang M, Hu C, et al. Quantitative biomarkers for prediction of epidermal growth factor receptor mutation in non-small cell lung cancer. Transl Oncol 2018; 11: 94-101.
41. Rizzo S, Petrella F, Buscarino V, De Maria F, Raimondi S, Barberis M, et al. CT radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer. Eur Radiol 2016; 26: 32-42.
42. Gevaert O, Echegaray S, Khuong A, Hoang CD, Shrager JB, Jensen KC, et al. Predictive radiogenomics modeling of EGFR mutation status in lung cancer. Sci Rep 2017; 7: 41674. DOI: 10.1038/SREP41674.
43. Eichler AF, Kahle KT, Wang DL, Joshi VA, Willers H, Engelman JA, et al. EGFR mutation status and survival after diagnosis of brain metastasis in nonsmall cell lung cancer. Neuro Oncol 2010; 12: 1193-1199.
44. Italiano A, Vandenbos FB, Otto J, Mouroux J, Fontaine D, Marcy PY, et al. Comparison of the epidermal growth factor receptor gene and protein in primary non-small-cell-lung cancer and metastatic sites: implications for treatment with EGFR-inhibitors. Ann Oncol 2006; 17: 981-985.
45. Lee CC, Soon YY, Tan CL, Koh WY, Leong CN, Tey JCS, Tham IWK. Discordance of epidermal growth factor receptor mutation between primary lung tumor and paired distant metastases in non-small cell lung cancer: a systematic review and meta-analysis. PLoS One 2019; 14: e0218414. DOI: 10.1371/JOURNAL.PONE.0218414.
46. Ahn SJ, Kwon H, Yang JJ, Park M, Cha YJ, Suh SH, Lee JM, et al. Contrast-enhanced T1-weighted image radiomics of brain metastases may predict EGFR mutation status in primary lung cancer. Sci Rep 2020; 10. DOI: 10.1038/S41598-020-65470-7.