



OPEN ACCESS

EDITED BY

Jeffrey Velotta,
Kaiser Permanente, United States

REVIEWED BY

Hailin Tang,
Sun Yat-sen University Cancer Center
(SYSUCC), China
Wei Wei,
Anhui Provincial Hospital, China

*CORRESPONDENCE

Hong Zhao
✉ 178331090@qq.com

†PRESENT ADDRESS

Peng Han,
Department of Radiology, The Fifth Affiliated
Hospital of Anhui
Medical University, Fuyang, China

RECEIVED 06 June 2025

ACCEPTED 29 August 2025

PUBLISHED 18 September 2025

CITATION

Han P, Zhang D, Yao W, Lv M, Qian Y and
Zhao H (2025) Noninvasive prediction of
EGFR 19Del and 21L858R subtypes in lung
adenocarcinoma: a comparative study of
logistic regression and decision tree models.
Front. Oncol. 15:1642253.
doi: 10.3389/fonc.2025.1642253

COPYRIGHT

© 2025 Han, Zhang, Yao, Lv, Qian and Zhao.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Noninvasive prediction of EGFR 19Del and 21L858R subtypes in lung adenocarcinoma: a comparative study of logistic regression and decision tree models

Peng Han^{1,2†}, Dai Zhang^{1,2}, Wenjun Yao^{1,2}, MengYu Lv^{1,2},
YunHong Qian^{1,2} and Hong Zhao^{1,2*}

¹Department of Radiology, The Second Affiliated Hospital of Anhui Medical University, Hefei, China,

²Medical Imaging Research Center, Anhui Medical University, Hefei, China

Objective: Despite the increasing interest in radiogenomic prediction, few studies have directly compared the performance of logistic regression and decision tree models in distinguishing epidermal growth factor receptor (EGFR) mutation subtypes. This study provides the first systematic comparison of the predictive performance of these two models in identifying exon 19 deletions (19Del) and exon 21 L858R point mutations (21L858R) in patients with lung adenocarcinoma. By leveraging imaging and clinical parameters, we aimed to address a critical gap in the literature by establishing an optimal prediction model and providing a noninvasive tool to support personalized treatment strategies for patients with unknown EGFR mutation status.

Materials and methods: We retrospectively collected clinical and radiological data from 193 patients with histologically confirmed lung adenocarcinoma who were admitted to the Second Affiliated Hospital of Anhui Medical University between May 2018 and June 2024. Based on EGFR genotyping results, patients were stratified into two groups: the EGFR 19Del mutation group and the EGFR 21L858R mutation group. Comparative statistical analyses—including Student's *t*-test, Mann–Whitney *U* test, chi-square test, or Fisher's exact test—were performed to evaluate differences in clinical and CT imaging characteristics between groups. Variables with $P < 0.05$ in the univariate analysis were subsequently included in both logistic regression and decision tree models to identify independent predictors of EGFR mutation subtype. Model performance was assessed using ROC curve analysis. The area under the curve (AUC) was calculated for each model, and their predictive accuracy was further compared using DeLong's test, net reclassification improvement (NRI), and integrated discrimination improvement (IDI).

Results: In the decision tree model, age and brain metastasis emerged as key decision nodes for differentiating 19Del and 21L858R mutations, with an AUC of 0.712 (95% CI: 0.639–0.785). In contrast, the logistic regression model identified age, pleural thickening, lymphadenopathy, and brain metastasis as independent predictors, achieving a higher AUC of 0.740 (95% CI: 0.671–0.810). The NRI and IDI values were 0.498 ($P < 0.001$, 95% CI: 0.238–0.758) and 0.043 ($P = 0.004$,

95% CI: 0.013–0.072), respectively, suggesting improved reclassification and discrimination by the logistic model. However, DeLong's test revealed no statistically significant difference between the AUCs of the two models ($Z = 1.314$, $P = 0.189$).

Conclusion: Both logistic regression and decision tree models demonstrated value in predicting EGFR 19Del and 21L858R mutations in lung adenocarcinoma, each offering distinct methodological advantages. The logistic regression model exhibited higher interpretability and statistical robustness, making it well-suited for clinical decision-making. Meanwhile, the decision tree model offered superior visual clarity and intuitive structure, which may enhance practical utility. A combined modeling approach that harnesses the strengths of both methods may provide a more accurate and comprehensive tool for early mutation identification and individualized treatment planning in patients with lung adenocarcinoma.

KEYWORDS

EGFR, lung adenocarcinoma, decision tree model, logistic regression model, NRI

1 Introduction

Lung adenocarcinoma (LUAD), a predominant histological subtype of non-small cell lung cancer (NSCLC), continues to pose a substantial global health burden, with 2024 estimates (Global Cancer Observatory preliminary data) indicating approximately 1.1 million new cases and 720,000 deaths annually worldwide, accounting for 45–50% of total NSCLC mortality (1, 2). Among its molecular drivers, mutations in the epidermal growth factor receptor (EGFR) gene are the most frequently observed, particularly in Asian populations, where their prevalence underscores the critical need for genotype-guided treatment strategies. The two most common EGFR mutation subtypes—exon 19 deletion (19Del) and exon 21 L858R point mutation (21L858R)—collectively account for approximately 85% of all EGFR mutations, and exhibit distinct therapeutic sensitivities (3–5).

Accumulating evidence suggests that patients harboring EGFR 19Del mutations generally derive greater clinical benefit from EGFR tyrosine kinase inhibitors (TKIs) compared to conventional chemotherapy (6, 7), while the response to TKIs in those with L858R mutations appears more variable, with some studies indicating comparable or even superior outcomes with cytotoxic chemotherapy (8, 9). These findings highlight the necessity of accurately identifying EGFR mutation subtypes to enable more precise and individualized therapeutic decisions (10, 11).

Currently, tissue biopsy remains the gold standard for determining EGFR mutation status. However, this invasive procedure is associated with potential complications such as hemorrhage, infection, and pneumothorax. Moreover, due to intratumoral heterogeneity, small biopsy samples may not fully reflect the genomic landscape of the entire tumor (12). Consequently, there is a pressing need for

noninvasive, rapid, and reliable methods to predict EGFR mutation subtypes in clinical settings.

With the rapid evolution of medical imaging technologies, radiological assessment has emerged as an essential tool not only for diagnosis and treatment monitoring but also for exploring molecular correlates through imaging biomarkers (13–15). Although numerous studies have investigated the potential of imaging features for predicting EGFR mutation status in LUAD (16, 17), the comparative efficacy of logistic regression versus decision tree models in discriminating EGFR mutation subtypes (particularly 19Del vs. 21L858R) remains underexplored, representing a critical knowledge gap this study seeks to address. Therefore, in this study, we sought to develop and compare logistic regression and decision tree models based on clinical parameters and CT imaging features to noninvasively predict EGFR mutation subtype status in patients with lung adenocarcinoma.

2 Materials and methods

2.1 General information

This retrospective study was approved by the Institutional Review Board of the Second Affiliated Hospital of Anhui Medical University (Approval No. YX2023-212), with the requirement for written informed consent waived. A total of 1,200 patients with pathologically confirmed lung adenocarcinoma and documented epidermal growth factor receptor (EGFR) mutations were initially screened. These patients were diagnosed and treated at the Second Affiliated Hospital of Anhui Medical University between November 2018 and June 2024.

The inclusion criteria were as follows:

1. histopathological diagnosis of primary lung adenocarcinoma;
2. presence of EGFR mutations limited to exon 19 deletions (19Del) or exon 21 L858R point mutations (21L858R) as determined by gene testing;
3. availability of preoperative chest CT scan data;
4. no history of other malignancies;
5. no history of preoperative radiotherapy or chemotherapy.

Exclusion criteria included:

1. incomplete clinical information (e.g., gender, age, smoking history);
2. receipt of any treatment prior to surgery;
3. time interval exceeding one month between the preoperative CT scan and surgical resection;
4. missing EGFR mutation subtype results;
5. The CT image quality is poor, making it difficult to determine the tumor lesions. Ultimately, 193 patients with lung adenocarcinoma EGFR mutation subtypes were included in this study. The screening process is shown in the figure (Figure 1).

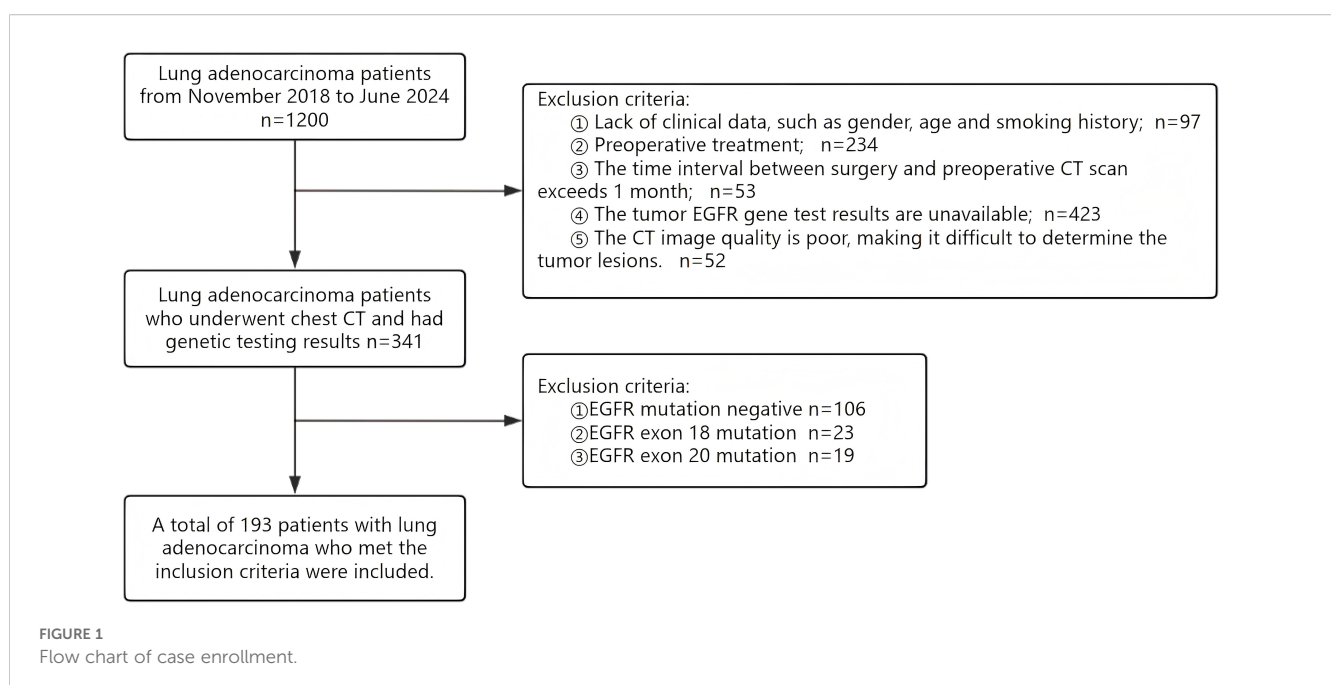
2.2 CT scan parameters

All patients underwent chest computed tomography (CT) scanning using a SOMATOM Force 128-slice scanner (Siemens Healthineers, Erlangen, Germany). The imaging protocol was standardized as follows: tube voltage was set at 120 kV, with a tube current ranging from 150 to 200 mA. The display field of view

(DFOV) was 350 mm. Axial images were acquired with a slice thickness and interslice gap of 5 mm. For image reconstruction, a slice thickness of 1.25 mm and a reconstruction interval of 1.25 mm were applied, with a matrix size of 512×512 . These parameters ensured optimal spatial resolution for the assessment of pulmonary and mediastinal structures.

2.3 CT image interpretation

Following communication and approval from the hospital's administrative and medical records departments, clinical case data were retrieved through the institution's electronic medical record system. Computed tomography (CT) images were independently reviewed by two board-certified radiologists, each with five years of diagnostic experience. During the image interpretation process, both radiologists were blinded to the patients' clinical data and mutation status to minimize bias. Interobserver agreement for qualitative CT features was assessed using Cohen's Kappa statistic, with detailed results provided in Supplementary Material [Supplementary Table S1](#). In instances of initial disagreement, a consensus was reached through discussion. The following clinical and radiological features were systematically documented. Clinical variables included sex, age, and smoking history. CT images were evaluated using standard lung window settings (window width: 1500 HU; window level: -700 HU) and mediastinal window settings (window width: 350 HU; window level: 40 HU). The assessed CT features encompassed tumor morphology (including lesion type, maximum diameter, lobulation, and spiculation), internal characteristics (such as calcification, necrosis, cavitation, and air bronchogram), pleural abnormalities (including thickening, retraction, and effusion), peritumoral changes (e.g., emphysema and vascular convergence), lymph node involvement, distant



metastases (brain, liver, bone, and contralateral lung), and CT attenuation values.

2.4 Detection of EGFR mutations in tumor specimens

EGFR mutation analysis was conducted in the Department of Pathology at the Second Affiliated Hospital of Anhui Medical University. Mutations in exons 19 and 21 of the EGFR gene were detected using the PCR-ARMS (amplification refractory mutation system) technique with a commercially available human EGFR mutation detection kit (Beijing SinoMD Gene Detection Technology Co., Ltd., Beijing, China).

Cases harboring exon 19 deletions were classified as the EGFR 19Del mutation group, while those exhibiting the exon 21 L858R point mutation were assigned to the EGFR 21L858R mutation group.

2.5 Statistical analysis methods

Statistical analyses were performed using SPSS software version 27.0 (IBM Corp., Armonk, NY, USA), with a two-tailed P -value < 0.05 considered indicative of statistical significance. For continuous variables, the independent samples t -test was applied to data with a normal distribution, while the Mann–Whitney U test was used for non-normally distributed data. Categorical variables were analyzed using the chi-square test or Fisher's exact test, as appropriate. Variables demonstrating statistical significance ($P < 0.05$) in univariate analyses were subsequently entered into a multivariate logistic regression model to identify independent predictors associated with 19Del and 21L858R mutations in patients with lung adenocarcinoma. A classification-based decision tree algorithm was employed to construct the decision tree model, with node splitting based on chi-square values. Internal validation was conducted via 10-fold cross-validation to assess model robustness. Receiver operating characteristic (ROC) curves were generated for both the logistic regression and decision tree models, and the areas under the curve (AUCs) were calculated to evaluate predictive performance. Comparative analyses between the two models were further conducted using the DeLong test, net reclassification improvement (NRI), and integrated discrimination improvement (IDI) metrics.

3 Results

3.1 Clinical and imaging characteristics of patients

Patients with the EGFR 21L858R mutation exhibited a significantly higher prevalence of pleural thickening and brain metastasis compared to those with the EGFR 19Del mutation ($P < 0.05$). Conversely, lymphadenopathy was more frequently

observed in the EGFR 19Del mutation group ($P < 0.05$). Additionally, age was found to be negatively associated with the presence of EGFR 19Del mutation. However, no statistically significant differences were identified between the two groups in terms of general clinical characteristics (e.g., sex, smoking history) or CT imaging features, including tumor type, location, maximum diameter, lobulation, and other morphological signs ($P > 0.05$), as detailed in [Table 1](#).

3.2 Establishment of multi-factor logistic regression model

Variables identified as statistically significant in the univariate analysis were subsequently included in the multivariate logistic regression model. The results revealed that age ($OR = 1.065$, $P = 0.001$; 95% CI: 1.031–1.101), brain metastasis ($OR = 1.975$, $P = 0.036$; 95% CI: 1.045–3.730), and pleural thickening ($OR = 2.124$, $P = 0.026$; 95% CI: 1.096–4.117) were positively associated with an increased likelihood of harboring the EGFR 21L858R mutation. In contrast, lymphadenopathy was more frequently associated with the EGFR 19Del mutation ($OR = 0.462$, $P = 0.019$; 95% CI: 0.242–0.881). These findings are summarized in [Table 2](#), with representative clinical imaging examples presented in [Figures 2 and 3](#).

3.3 Building a decision tree model

The results of the decision tree model, constructed using the classification decision tree algorithm, are illustrated in [Figure 4](#). This model comprises three levels and five nodes, three of which are terminal nodes. Age and brain metastasis emerged as the key factors in predicting EGFR 19Del and 21L858R mutation status in patients with lung adenocarcinoma. At the first level, the model was stratified by age, highlighting a strong association between age and EGFR mutation status. Specifically,

among patients aged ≤ 69 years, 56.1% exhibited the EGFR 19Del mutation and 43.9% the EGFR 21L858R mutation; among those aged > 69 years, 24.6% had the EGFR 19Del mutation, while 75.4% had the EGFR 21L858R mutation. Further analysis revealed that for patients aged ≤ 69 years without brain metastases (node 3), 37% presented with the EGFR 19Del mutation and 63% with the EGFR 21L858R mutation. Conversely, in patients with brain metastases (node 4), 45.1% had the 19Del mutation, and 54.9% had the 21L858R mutation.

3.4 Comparison of the predictive performance and diagnostic efficacy of the two models

The area under the curve (AUC) of the decision tree model and logistic regression model for identifying EGFR 19Del and 21L858R mutations were 0.712 (95% CI = 0.639–0.785) and 0.740 (95% CI =

TABLE 1 Association between clinical and CT imaging features and EGFR 19Del or 21L858R mutations.

Variable	EGFR mutation subtype		Statistic value	P value
	19Del n(%)	21L858R n(%)		
Gender			0.01	0.921
Female	49(55.06)	58(55.77)		
Male	40(44.94)	46(44.23)		
Age(year)	58.000(51.0,67.0)	69.000(58.0,73.0)	4.437	<0.001*
Smoking history			2.853	0.091
Absent	79(88.76)	83(79.81)		
Present	10(11.24)	21(20.19)		
Tumor location			0.306	0.580
Central	16(17.98)	22(21.15)		
Peripheral	73(82.02)	82(78.85)		
Tumor size	34.200(23.8,48.4)	35.000(25.9,53.3)	0.352	0.725
Lobular sign			0.432	0.511
Absent	10(11.24)	15(14.42)		
Present	79(88.76)	89(85.58)		
Spike Sign			0.016	0.900
Absent	36(40.45)	43(41.35)		
Present	53(59.55)	61(58.65)		
Edge			0.170	0.680
Clarity	48(53.93)	53(50.96)		
Unclear	41(46.07)	51(49.04)		
Air bronchogram			1.325	0.250
Absent	51(57.30)	68(65.38)		
Present	38(42.70)	36(34.62)		
Bubble sign			0.039	0.843
Absent	67(75.28)	77(74.04)		
Present	22(24.72)	27(25.96)		
Peripheralempysema			1.182	0.277
Absent	80(89.89)	88(84.62)		
Present	9(10.11)	16(15.38)		
Vascular clustering			0.534	0.465
Absent	12(13.48)	18(17.31)		
Present	77(86.52)	86(82.69)		
Necrosis in tumor			0.232	0.630
Absent	51(57.30)	56(53.85)		
Present	38(42.70)	48(46.15)		
Pleural traction			0.029	0.865
Absent	23(25.84)	28(26.92)		

(Continued)

TABLE 1 Continued

Variable	EGFR mutation subtype		Statistic value	P value
	19Del n(%)	21L858R n(%)		
Present	66(74.16)	76(73.08)		
Pleural thickening			10.418	<0.001*
Absent	64(71.91)	51(49.04)		
Present	25(28.09)	53(50.94)		
Calcificationintumor			2.407	0.121
Absent	77(86.52)	81(77.88)		
Present	12(13.48)	23(22.12)		
Pleural effusion			0.642	0.423
Absent	63(70.79)	68(65.38)		
Present	26(29.21)	36(34.62)		
Inanition			0.341	0.559
Absent	81(91.01)	97(93.27)		
Present	8(8.99)	7(6.73)		
Swollen lymph nodes			4.628	0.031*
Absent	30(33.71)	51(49.04)		
Present	59(66.29)	53(50.96)		
CT value	38.57 ± 10.76	39.72 ± 9.71	0.787	0.432
Bone metastasis			1.513	0.219
Absent	55(61.80)	73(70.19)		
Present	34(38.20)	31(29.81)		
Brain metastases			6.129	0.013*
Absent	61(68.54)	53(50.96)		
Present	28(31.46)	51(49.04)		
Liver metastasis			2.006	0.157
Absent	80(89.89)	99(95.19)		
Present	9(10.11)	5(4.81)		
Lung metastasis			0.264	0.607
Absent	83(93.26)	94(91.26)		
Present	6(6.74)	9(8.74)		

*P < 0.05.

0.671-0.810), respectively. The ROC curves of these two models are shown in [Figure 5](#). The NRI and IDI values for diagnostic efficacy evaluation are 0.498 ($P<0.001$, 95% CI: 0.238-0.758) and 0.043 ($P = 0.004$,95% CI: 0.013-0.072), respectively. The P values for both are<0.05, indicating a significant improvement in the diagnostic ability of the logistic regression model compared to the decision tree model. However, DeLong’s test results showed no statistically significant difference between the two models ($Z = 1.314$, P value=0.189), as shown in [Table 3](#).

4 Discussion

With the continued advancement of molecular biology, the treatment paradigm for lung adenocarcinoma has shifted from conventional approaches based on histopathological and clinical characteristics toward precision medicine guided by individual genetic alterations—most notably, mutations in the epidermal growth factor receptor (EGFR) gene. Among EGFR mutations, exon 19 deletions (19Del) and the exon 21 L858R point mutation

TABLE 2 Multivariate logistic regression model associated with EGFR19DeL and 21L858R mutations.

Variable	B	s.e	Wald	Sig.	Exp(B)		95% CI
Age	0.063	0.017	14.068	0.001	1.065	1.031	1.101
Pleural thickening	0.753	0.338	4.978	0.026	2.124	1.096	4.117
Lymphadenopathy	-0.772	0.329	5.497	0.019	0.462	0.242	0.881
Brain metastases	0.680	0.324	4.396	0.036	1.975	1.045	3.73

(21L858R) represent the two most prevalent subtypes, together accounting for the vast majority of EGFR-mutant lung adenocarcinomas (18, 19). Both subtypes have demonstrated substantial sensitivity to EGFR tyrosine kinase inhibitors (TKIs); however, emerging evidence suggests notable differences in their molecular profiles, drug resistance mechanisms, clinical behavior, and therapeutic responses (20, 21). Despite the frequent clinical practice of treating 19DeL- and 21L858R-mutant lung adenocarcinomas with similar TKI-based regimens, studies have increasingly shown that patients harboring 19DeL mutations tend to experience longer progression-free and overall survival compared to those with EGFR 21L858R mutation, particularly in the setting of advanced disease (22–24). In light of these findings, the present study aimed to develop logistic regression and decision tree models leveraging clinical and radiological features to distinguish between 19DeL and 21L858R subtypes. This approach seeks to enhance

individualized therapeutic planning, especially in patients with indeterminate or unavailable EGFR genotyping.

Both modeling approaches identified age and presence of brain metastasis as key predictors of EGFR mutation subtype. Consistent with prior literature, our results indicate that patients with EGFR 19DeL mutation tend to be younger than those with EGFR 21L858R mutation (25). The EGFR 19DeL mutation rate declined with increasing age, whereas EGFR 21L858R mutations were infrequent in patients aged ≤ 40 years, highlighting a clear difference in age-related mutation distribution. In our cohort, the median age of patients with EGFR 19DeL and 21L858R mutations was 58.0 (IQR: 51.0–67.0) and 69.0 years (IQR: 58.0–73.0), respectively ($z = 4.437$, $P < 0.01$), corroborating previously reported trends. Brain metastasis also emerged as a distinguishing clinical feature. Among patients with EGFR 19DeL mutation, 28 of 89 (31.5%) had brain metastases, compared to 51 of 104 (49.0%) in

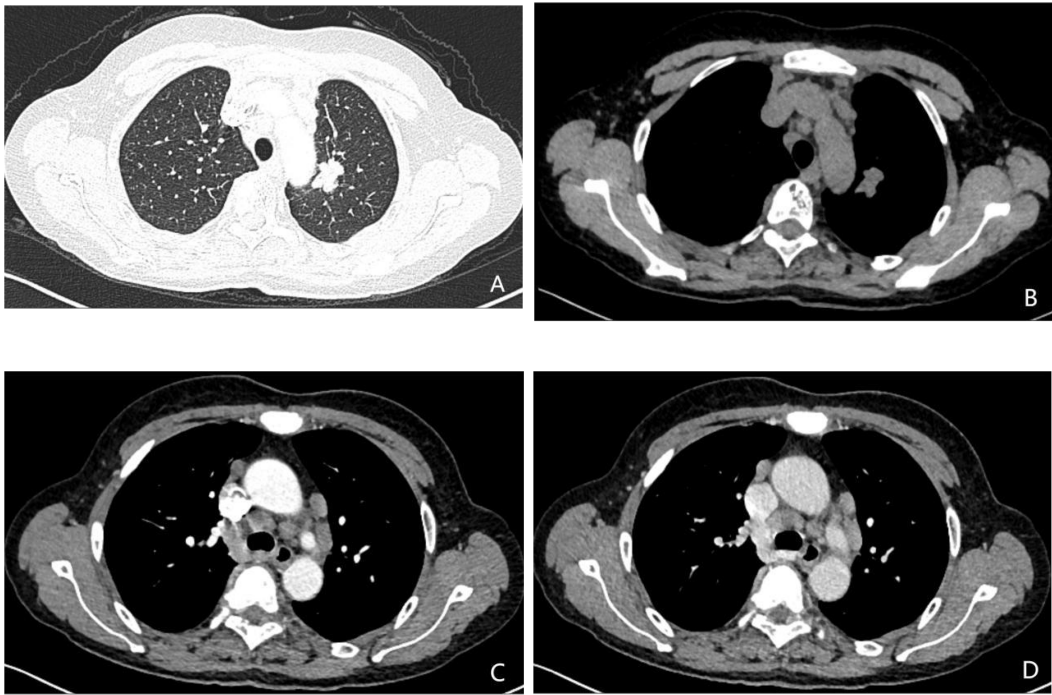


FIGURE 2 (A–D) A 58-year-old female patient with lung adenocarcinoma, the genetic test result was EGFR 19DeL mutation (+). (A–B) Unenhanced CT demonstrates an irregular soft-tissue density shadow in the left upper lobe, measuring 25.3 mm in maximum diameter with lobulated margins. (C, D) Contrast-enhanced CT shows multiple enlarged lymph nodes adjacent to the trachea and aorta during arterial and venous phases.

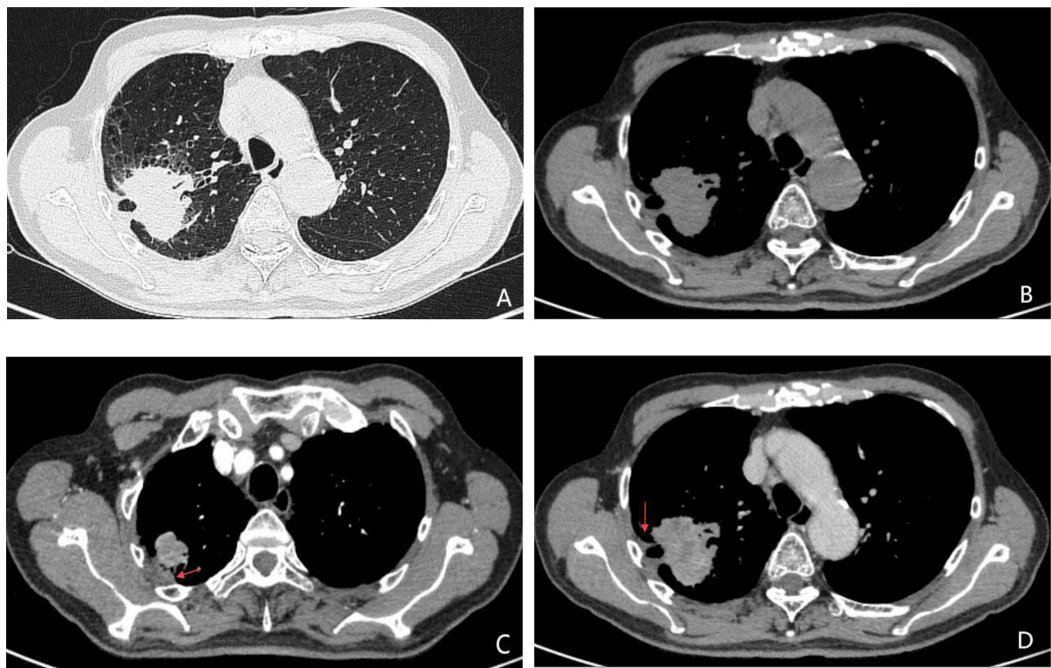


FIGURE 3
(A–D) A 74-year-old male patient with lung adenocarcinoma, the genetic testing result was EGFR 21L858R mutation (+). Unenhanced CT images (A, B) demonstrate an irregular soft-tissue nodule with indistinct margins and adjacent pleural thickening in the right upper lobe. Contrast-enhanced CT during arterial (C) and venous (D) phases reveals heterogeneous marked enhancement of the lesion, with no significant lymphadenopathy in the mediastinum.

the 21L858R group—a statistically significant difference ($\chi^2 = 6.129$, $P = 0.013$). This aligns with prior studies suggesting a higher propensity for brain dissemination in tumors harboring EGFR 21L858R mutation (26, 27). The underlying mechanism may involve increased tumor aggressiveness, enhanced permeability of the blood-brain barrier, and immune evasion capabilities associated with EGFR 21L858R-driven malignancies.

In addition, the logistic regression model identified pleural thickening and lymphadenopathy as significant predictors. Our findings concur with those of Zhang et al. (28), who reported a positive association between lymphadenopathy and the EGFR 19Del mutation. However, while Kong et al. (29) observed a stronger link between pleural thickening and EGFR 19Del

mutation, our results suggest a contrary trend, warranting further investigation. Notably, these features were not selected as key decision points in the decision tree model—potentially due to differences in model architecture. Logistic regression, a parametric linear model (30), excels in quantifying direct associations between predictors and outcomes, even with relatively small or high-dimensional datasets. In contrast, decision trees operate via hierarchical segmentation and may prioritize variables with stronger or earlier splitting power, potentially overlooking subtler associations (31).

Despite these methodological differences, both models demonstrated utility in predicting EGFR mutation subtype. The logistic regression model identified four explanatory variables,

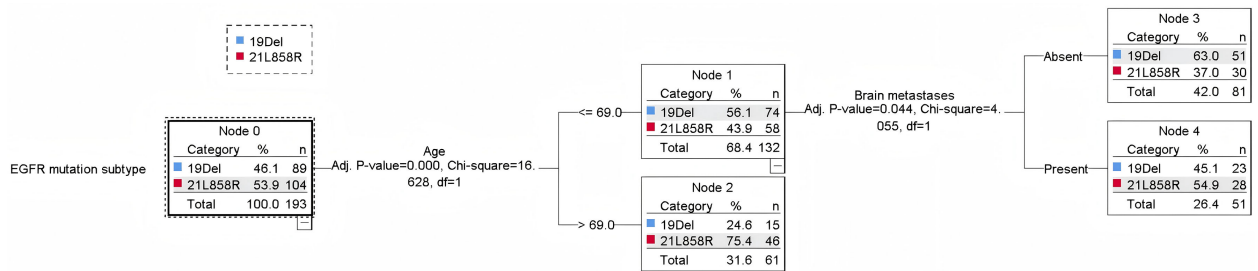


FIGURE 4
Decision tree model.

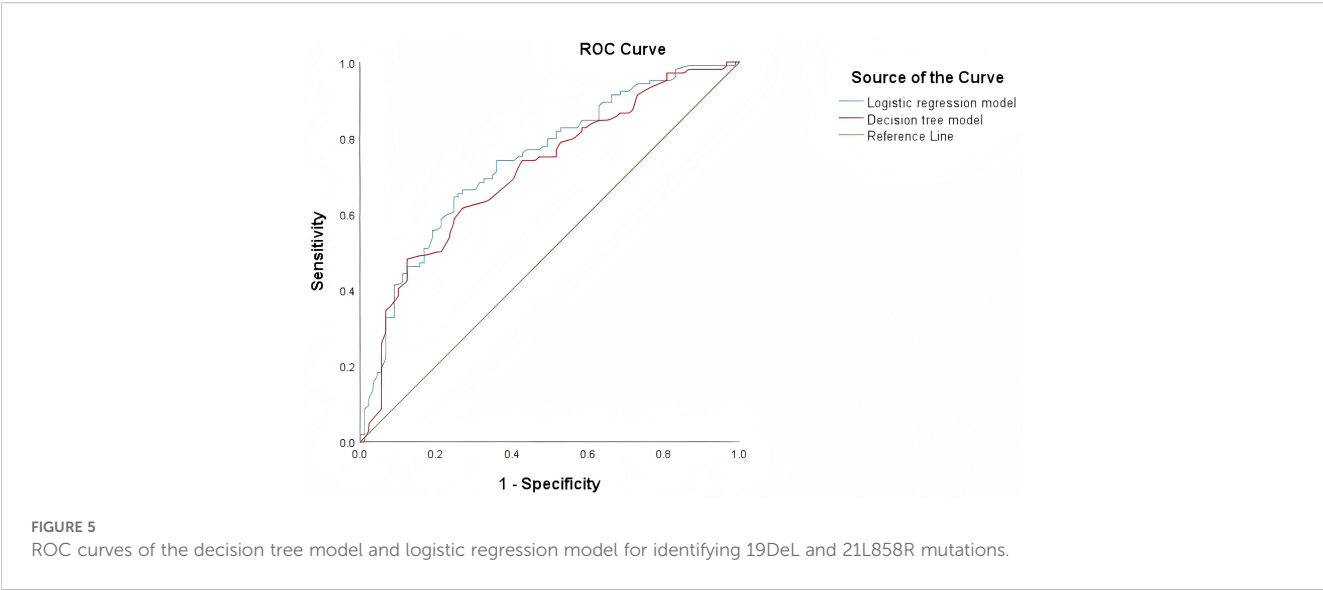


TABLE 3 Comparison of prediction performance of DeLong test decision tree model and logistic regression model.

Test result variable(s)	AUC (95%CI)	P Value	NRI (95%CI)	P Value for NRI	IDI (95%CI)	P Value for IDI
Decision tree mode	0.712 (0.639-0.785)	<0.001	Reference	Reference	Reference	Reference
Logistic regression model	0.740 (0.671-0.810)	<0.001	0.498 (0.238-0.758)	0.001	0.043 (0.013-0.072)	0.004
DeLong test	0.028 (0.070-0.014)	0.189	NA	NA	NA	NA

whereas the decision tree model selected two. Both consistently highlighted age and brain metastasis as the most influential predictors. In terms of performance, the logistic regression model yielded an AUC of 0.740 (95% CI: 0.671–0.810), slightly outperforming the decision tree model (AUC = 0.712; 95% CI: 0.639–0.785). Net reclassification improvement (NRI) and integrated discrimination improvement (IDI) analyses further supported the superior classification performance of the logistic regression model (NRI = 0.498, $P = 0.001$; IDI = 0.043, $P = 0.004$).

However, DeLong’s test showed no statistically significant difference between AUCs ($Z = 1.314$, $P = 0.189$), suggesting that while logistic regression may offer nuanced advantages in certain contexts, both models are comparably effective overall. The discrepancy in statistical significance across performance metrics reflects their differing emphases: while NRI and IDI capture shifts in classification performance, particularly for individuals near decision thresholds, the DeLong test evaluates overall discriminative capacity. Therefore, integrating both modeling strategies may offer complementary strengths, enhancing predictive accuracy and clinical utility.

Notwithstanding these promising findings, several limitations warrant acknowledgment. First, this retrospective analysis focused exclusively on CT imaging and select clinical parameters, potentially omitting other relevant variables. Second, despite employing a double-blind approach to radiological interpretation, selection bias may have influenced the observed associations. Most importantly, while robust internal validation was performed through cross-validation techniques (applied to the decision tree model), the lack of external validation represents a key limitation that may affect the generalizability of our results. This constraint, common in radiomics studies, highlights the critical need for future prospective multicenter validation to confirm the clinical applicability of our models.

In summary, logistic regression and decision tree models constructed using age, lymphadenopathy, pleural thickening, and brain metastasis show promise for noninvasive prediction of EGFR 19DeL and 21L858R mutation subtypes in lung adenocarcinoma. These models may inform genotype-tailored treatment decisions, facilitate early identification of high-risk patients, and ultimately improve prognostic assessment in clinical practice.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author/s.

Ethics statement

This retrospective study was approved by the Medical Ethics Committee of the Second Affiliated Hospital of Anhui Medical University (No. YX2023-212), and the requirement for written informed consent was waived. The article contains only de-identified radiological images; written informed consent for publication would be obtained if any identifiable images or data were included.

Author contributions

PH: Investigation, Writing – original draft, Data curation, Visualization, Software, Validation, Methodology. DZ: Methodology, Software, Writing – original draft. WY: Writing – review & editing, Methodology, Conceptualization. ML: Software, Writing – original draft. YQ: Methodology, Software, Writing – original draft. HZ: Data curation, Methodology, Conceptualization, Validation, Visualization, Software, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This work was supported by the Health Research Program of Anhui (AHWJ2023BAb20009); the Key Natural Science Research Project of the Anhui Provincial Education Department (2024AH050796); and the Key Project of the Clinical Research Cultivation Program of The Second Affiliated Hospital of Anhui Medical University (2020LCZD12). The funders

had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2025.1642253/full#supplementary-material>

References

- Miao Z, Yu W. Significance of novel PANoptosis genes to predict prognosis and therapy effect in the lung adenocarcinoma. *Sci Rep.* (2024) 14:20934. doi: 10.1038/s41598-024-71954-7
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2024) 71:209–49. doi: 10.3322/caac.21660
- Laface C, Maselli FM, Santoro AN, Iaia ML, Ambrogio F, Laterza M, et al. The resistance to EGFR-TKIs in non-small cell lung cancer: from molecular mechanisms to clinical application of new therapeutic strategies. *Pharmaceutics.* (2023) 15:1604. doi: 10.3390/pharmaceutics15061604
- Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med.* (2004) 350:2129–39. doi: 10.1056/NEJMoa040938
- Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science.* (2004) 304:1497–500. doi: 10.1126/science.1099314
- Kaneda T, Yoshioka H, Tamiya M, Tamiya A, Hata A, Okada A, et al. Differential efficacy of cisplatin plus pemetrexed between L858R and Del-19 in advanced EGFR-mutant non-squamous non-small cell lung cancer. *BMC Cancer.* (2018) 18:6. doi: 10.1186/s12885-017-3952-7
- Kato T, Yoshioka H, Okamoto I, Yokoyama A, Hida T, Seto T, et al. Afatinib versus cisplatin plus pemetrexed in Japanese patients with advanced non-small cell lung cancer harboring activating EGFR mutations: Subgroup analysis of LUX-Lung 3. *Cancer Sci.* (2015) 106:1202–11. doi: 10.1111/cas.12723
- Lin J, Li M, Chen S, Weng L, He Z. Efficacy and safety of first-generation EGFR-TKIs combined with chemotherapy for treatment-naïve advanced non-small-cell lung cancer patients harboring sensitive EGFR mutations: A single-center, open-label, single-arm, phase II clinical trial. *J Inflammation Res.* (2021) 14:2557–67. doi: 10.2147/JIR.S313056
- Yang JC-H, Wu Y-L, Schuler M, Sebastian M, Popat S, Yamamoto N, et al. Afatinib versus cisplatin-based chemotherapy for EGFR mutation-positive lung adenocarcinoma (LUX-Lung 3 and LUX-Lung 6): analysis of overall survival data from two randomised, phase 3 trials. *Lancet Oncol.* (2015) 16:141–51. doi: 10.1016/S1470-2045(14)71173-8

10. Li D, Ding L, Ran W, Huang Y, Li G, Wang C, et al. Status of 10 targeted genes of non-small cell lung cancer in eastern China: A study of 884 patients based on NGS in a single institution. *Thorac Cancer*. (2020) 11:2580–9. doi: 10.1111/1759-7714.13577
11. Castellanos E, Feld E, Horn L. Driven by mutations: the predictive value of mutation subtype in EGFR-mutated non-small cell lung cancer. *J Thorac Oncol*. (2017) 12:612–23. doi: 10.1016/j.jtho.2016.12.014
12. Huo JW, Luo TY, Diao L, Jiang B, Zhang P, Yang J, et al. Using combined CT-clinical radiomics models to identify epidermal growth factor receptor mutation subtypes in lung adenocarcinoma. *Front Oncol*. (2022) 12:846589. doi: 10.3389/fonc.2022.846589
13. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169
14. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J*. (2019) 53:1800986. doi: 10.1183/13993003.00986-2018
15. Yoon HJ, Choi J, Kim E, Kim H, Ahn YC, Lee HY, et al. Deep learning analysis to predict EGFR mutation status in lung adenocarcinoma manifesting as pure ground-glass opacity nodules on CT. *Front Oncol*. (2022) 12:951575. doi: 10.3389/fonc.2022.951575
16. Zhang B, Qi S, Pan X, Li C, Fan J, Liu S, et al. Deep CNN model using CT radiomics feature mapping recognizes EGFR gene mutation status of lung adenocarcinoma. *Front Oncol*. (2020) 10:598721. doi: 10.3389/fonc.2020.598721
17. Li Q, He XQ, Fan X, Zhu CN, Lv JW, Luo TY. Development and validation of a combined model for preoperative prediction of lymph node metastasis in peripheral lung adenocarcinoma. *Front Oncol*. (2021) 11:675877. doi: 10.3389/fonc.2021.675877
18. Shi A, Wang J, Wang Y, Guo G, Fan C, Liu J. Predictive value of multiple metabolic and heterogeneity parameters of (18)F-FDG PET/CT for EGFR mutations in non-small cell lung cancer. *Ann Nucl Med*. (2022) 36:393–400. doi: 10.1007/s12149-022-01718-8
19. Wu Y, Zhao Y, Wu Y, Chen H, Ma S, Wang Q. A retrospective real-world study of prognostic factors associated with EGFR mutated lung cancer with leptomeningeal metastasis. *Clin Lung Cancer*. (2024) 25:347–353.e1. doi: 10.1016/j.clcc.2024.02.001
20. Liang S, Wang H, Zhang Y, Tian H, Li C, Hua D. Prognostic implications of combining EGFR-TKIs and radiotherapy in Stage IV lung adenocarcinoma with 19-Del or 21-L858R mutations: A real-world study. *Cancer Med*. (2024) 13:e7208. doi: 10.1002/cam4.7208
21. Kobayashi Y, Mitsudomi T. Not all epidermal growth factor receptor mutations in lung cancer are created equal: Perspectives for individualized treatment strategy. *Cancer Sci*. (2016) 107:1179–86. doi: 10.1111/cas.12996
22. Wang Y, Yuan X, Yang M, Shen Z, Chen H, He X, et al. Efficacy of icotinib, an EGFR tyrosine kinase inhibitor in non-small cell lung cancer patients with exon 19 deletion and exon 21 L858R: A retrospective analysis in China. *Pharmacology*. (2021) 106:658–66. doi: 10.1159/000519847
23. Liu W, Duan Z, Wu Y, Ma R. Silencing of lncRNA LOC105376794 promotes migration, invasion, and Gefitinib resistance of lung adenocarcinoma cells with EGFR 19Del mutation by ATF4/CHOP axis and ERK phosphorylation. *Neoplasma*. (2024) 71:219–30. doi: 10.4149/neo_2024_230616N316
24. Qian J, Ye X, Huang A, Qin R, Cai Y, Xue Y, et al. Afatinib 30 mg in the treatment of common and uncommon EGFR-mutated advanced lung adenocarcinomas: a retrospective, single-center, longitudinal study. *J Thorac Dis*. (2022) 14:2169–77. doi: 10.21037/jtd-22-507
25. Hu M, Zhang T, Yang Y, Li S, Yang W, Wang H, et al. Driver gene alterations in lung adenocarcinoma: Demographic features of 2544 Chinese cases. *Int J Biol Markers*. (2020) 35:44–50. doi: 10.1177/1724600820967015
26. Xu J, Yang Y, Gao Z, Liu H, Zhang J, Li L, et al. Distinguishing EGFR mutation molecular subtypes based on MRI radiomics features of lung adenocarcinoma brain metastases. *Clin Neurol Neurosurg*. (2024) 240:108258. doi: 10.1016/j.clineuro.2024.108258
27. Li Y, Lv X, Chen C, Xu W, Zhang L, Wang F, et al. A deep learning model integrating multisequence MRI to predict EGFR mutation subtype in brain metastases from non-small cell lung cancer. *Eur Radiol Exp*. (2024) 8:2. doi: 10.1186/s41747-023-00396-z
28. Zhang Y, He D, Fang W, Lu H, Zhang Q, Zhou L, et al. The difference of clinical characteristics between patients with exon 19 deletion and those with L858R mutation in nonsmall cell lung cancer. *Medicine*. (2015) 94:e1949. doi: 10.1097/MD.0000000000001949
29. Kong Q, Wang W, Wang Q, Yang Y, Chen G, Jiang T. Clinical characteristics and establishment of a 2-year-OS predictive model of EGFR mutation-positive patients with pleural invasion of lung adenocarcinoma. *Medicine*. (2023) 102:e34184. doi: 10.1097/MD.00000000000034184
30. Wang X, Zhang D. Choices of medical institutions and associated factors in older patients with multimorbidity in stabilization period in China: A study based on logistic regression and decision tree model. *Health Care Sci*. (2023) 2:359–69. doi: 10.1002/hcs2.73
31. Zhang Q, Zhang Z, Huang X, Zhou C, Xu J. Application of logistic regression and decision tree models in the prediction of activities of daily living in patients with stroke. *Neural Plast*. (2022) 2022:9662630. doi: 10.1155/2022/9662630