






SHORT COMMUNICATION

A deep-learning model using chest computed tomography images to predict epidermal growth factor receptor (EGFR) T790M mutation after first-line treatment with EGFR-tyrosine kinase inhibitor in patients with non-small cell lung cancer

Peng Min Liu¹, Jiang Feng Shi², Shan Wu³, Ye Hang Chen²,
Jun Ping Zhang¹, Bao Feng², and Hui Jing Feng^{1*}

¹Cancer Center, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Third Hospital of Shanxi Medical University, Taiyuan, China

²Laboratory of Artificial Intelligence of Biomedicine, Guilin University of Aerospace Technology, Guilin, Guangxi Province, People's Republic of China

³Department of Radiology, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Third Hospital of Shanxi Medical University, Taiyuan, China

***Corresponding author:**

Hui Jing Feng
(doctorfjh@sina.com)

Citation: Liu PM, Shi JF, Wu S, *et al.* A deep-learning model using chest computed tomography images to predict epidermal growth factor receptor (EGFR) T790M mutation after first-line treatment with EGFR-tyrosine kinase inhibitor in patients with non-small cell lung cancer. *Cancer Plus.* 2025;7(3):116-125. doi: 10.36922/cp.5587

Received: October 27, 2024

Revised: December 26, 2024

Accepted: February 17, 2025

Published online: March 11, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

To predict the epidermal growth factor receptor (EGFR) T790M status of patients with advanced non-small cell lung cancer (NSCLC) following the first-line first-/second-generation EGFR-tyrosine kinase inhibitor (EGFR-TKI) therapy, the related clinical features and chest computed tomography (CT) images of patients with advanced NSCLC in our hospital were retrospectively collected. All patients who met the criteria were randomly divided into training and validation cohorts. Then, a clinical model with the filtered clinical characteristics and a deep-learning model (DLM) were constructed. The area under the curve (AUC), specificity, sensitivity, accuracy, and decision curve analysis were used to evaluate model performance. In total, 66 patients met the inclusion criteria of the study (training cohort, $n = 40$; validation cohort, $n = 26$). EGFR19del and the use of gefitinib were significant ($P < 0.05$), and then, the clinical model was established using multivariate logistic regression analysis. The AUCs of the clinical model were 0.862 (95% confidence interval [CI], 0.570 – 0.966) and 0.755 (0.566 – 0.943) in the training and validation cohorts, respectively. The AUCs of the DLM from the chest CT image analysis were 0.839 (95% CI, 0.708 – 0.970) and 0.842 (0.680 – 1.000) in the training and validation cohorts, respectively. In the validation cohort, the DLM and clinical model exhibited an accuracy of 0.7308 and 0.5000, specificity of 0.6667 and 0.2000, positive probability values of 0.6429 and 0.4545, and negative probability values of 0.8333 and 0.7500, respectively. The DLM was developed using chest CT images to predict the *EGFR T790M* status following the first-line first- and second-generation EGFR-TKI treatment of advanced EGFR-positive NSCLC.

Keywords: Non-small cell lung cancer; Epidermal growth factor receptor; Epidermal growth factor receptor *T790M*; Chest computed tomography image; Deep learning

1. Introduction

Non-small cell lung cancer (NSCLC), as an aggressive malignant tumor, has high incidence and mortality rates,¹⁻³ representing approximately 80% of all lung cancer cases. They are typically in an advanced stage at the time of diagnosis.^{4,5} At present, the treatment of advanced NSCLC is mainly based on targeted therapy, immunotherapy, and chemotherapy.⁶⁻⁹ First-line molecular targeted therapy was approved for patients with advanced driver mutation-positive NSCLC, and the most common mutation of epidermal growth factor receptor (EGFR) gene accounts for approximately 50% of lung adenocarcinomas in Asian patients.¹⁰⁻¹² EGFR-tyrosine kinase inhibitor (TKI) is within the recommended guidelines for patients with EGFR-positive advanced NSCLC,^{13,14} with a median progression-free survival under 11 months following first-/second-generation EGFR-TKIs.¹⁵⁻¹⁸ However, the rapid development of therapeutic resistance and disease advancement in patients who received first-line treatment with either first- or second-generation EGFR-TKIs poses a significant challenge, particularly those accompanied with *EGFR T790M* mutation and activation of alternative signaling pathways.¹⁹ The *EGFR T790M* mutation is the most prevalent mechanism of resistance to first-/second-generation EGFR-TKIs,^{20,21} which can be effectively overcome by third-generation targeted drugs. Unfortunately, patients without *EGFR T790M* mutation can only receive chemotherapy and/or immunotherapy, which is less efficient, has more side effects, and is more costly than third-generation EGFR-TKIs. Therefore, predicting *EGFR T790M* status following first- and second-generation EGFR-TKI therapy is crucial for the long-term management of patients with advanced NSCLC with *EGFR* mutations.

The *EGFR T790M* mutation entails a change from threonine to methionine at the 790th position within the EGFR-tyrosine kinase domain,²²⁻²⁴ leading to a change in the spatial structure of *EGFR* and resulting in resistance. For patients with advanced EGFR-positive NSCLC, the *EGFR T790M* mutation rate following first-line treatment with first-/second-generation EGFR-TKIs ranges from 57% to 60%.^{25,26} However, only few studies have reported the prediction of *T790M* mutation status before first-/second-generation EGFR-TKI therapy. Previous studies have reported that the presence of favorable EGFR 19 mutations is correlated with a higher incidence of *EGFR T790M* mutation.²⁷ TP53-negative mutations correlate with *EGFR T790M* mutation.²⁸⁻³⁰ In addition, a research team adopted whole-exome sequencing at initial diagnosis to demonstrate the association between extended molecular profiling and the emergence of *T790M* resistance mutation

following first-/second-generation TKI therapy.³¹ Clinically, the detection of *T790M* mutation after disease progression despite treatment with first- or second-generation EGFR-TKIs is currently the sole method for determining the use of third-generation TKIs, which are invasive, painful, and time-consuming.

Radiomics, which utilizes medical imaging, is a non-invasive method of extracting features from medical image data to construct suitable models that assist in medical diagnosis. Traditional radiomics methods require manual annotation of tumor contours, which not only reduces the accuracy of lesion boundaries but also requires considerable manual effort. Recently, deep-learning (DL) methods have been applied to radiological analysis because they can automatically extract task-relevant features from image data without precise tumor contour annotations, thereby aiding clinical diagnosis.^{32,33} Many radiomic studies have attempted to investigate the subtle diagnosis, molecular typing, and prognostic prediction of NSCLC.³⁴⁻³⁶ In our previous study, an excellent radiomics model was established to estimate the programmed death-ligand 1 (PD-L1) expression in patients with NSCLC, which lays the foundation for the present study.³⁷

We have innovatively constructed a DL model (DLM) by analyzing pretreatment chest computed tomography (CT) images of patients newly diagnosed with advanced NSCLC, thereby predicting the *EGFR T790M* status following first-line first-/second-generation EGFR-TKI therapy. Based on radiomics features derived from chest CT data, the DLM was characterized as non-invasive, fast, and user-friendly. In addition, the area under the curve (AUC) demonstrated outstanding predictive capability, the decision curve analysis (DCA) demonstrated good decision ability, and the calibration curve revealed strong predictive power. The DLM developed using chest CT image data to estimate the *EGFR T790M* status following first-line first-/second EGFR-TKI therapy in advanced NSCLC with EGFR mutation will assist clinicians in selecting the initial treatment plans and managing the overall treatment course for patients with first-line EGFR-positive status.

2. Materials and methods

2.1. Patients

Following the approval of our hospital's ethics committee, this retrospective study collected data on all patients with NSCLC treated in our hospital between January 2014 and December 2021. The inclusion criteria were as follows: (1) confirmation of NSCLC through histopathological examination, including squamous cell carcinoma, adenocarcinoma, large cell carcinoma, adenosquamous carcinoma, and sarcomatoid carcinoma; (2) availability of

chest CT images before first-line treatment; (3) documented *EGFR* mutations before first-line treatment, including common exon 19 deletions, some point mutations in exon 21, or other rare *EGFR* mutations; (4) receipt of first-/second-generation *EGFR*-TKIs as first-line treatment, such as erlotinib, gefitinib, and icotinib; (5) assessment of *EGFR* T790M status following disease progression after the initial *EGFR*-TKI therapy. The exclusion criteria were as follows: (1) patients with NSCLC in the early or intermediate stage and (2) patients with unclear chest CT images. In total, 66 patients with advanced lung cancer were included in the study: data from 40 patients were used for DLM training, whereas data from 26 patients were used for model validation. Patient data are presented in Table 1.

2.2. CT imaging equipment

Non-contrast CT images of the chest were obtained using SOMATOM equipment (Siemens, Forchheim, Germany). The CT scanning parameters were set as follows: 120 kV, automatic tube current modulation, slice thickness of 0.75 – 1.5 mm, collimation of 128 × 0.6 mm, scan field of 336 × 336 mm, and pixel matrix of 512 × 512. The CT scan

dose ranged from 50 to 300 mGy·cm, determined by the patient's body mass index. High-resolution or standard images (slice thickness of 1.0 mm and interslice spacing of 1.0 mm) were reconstructed simultaneously.

2.3. EGFR and EGFR T790M status assessment

Before the initial treatment of patients newly diagnosed with advanced NSCLC, *EGFR* testing was performed using polymerase chain reaction (PCR) or next-generation sequencing (NGS). If the disease progressed despite receiving first-line treatment with first/second-generation *EGFR*-TKI therapy, the *EGFR* T790M status was reassessed using PCR or NGS. Tissue and peripheral blood samples were used, and peripheral blood samples were primarily utilized for patients without accessible tissue samples or for whom tissue sampling was challenging.

2.4. Construction of the clinical model

The clinical model was constructed in two steps. First, statistical analysis was performed on clinical characteristics (age, sex, etc.) using the Wilcoxon rank-sum test and Pearson's Chi-square test. Clinical features that exhibited

Table 1. Clinical features of the training and validation cohorts

Classification metrics	Training cohort (n=40)		P-value	Validation cohort (n=26)		P-value
	T790M Unmutated	T790m Mutated		T790M Unmutated	T790m Mutated	
EGFR 19						
Expression	19	5	0.001	10	10	0.147
Non-expression	4	12		5	1	
Gefitinib						
Present	14	4	0.019	7	10	0.019
Absent	9	13		8	1	
Sex						
Male	10	4	0.331	5	7	0.257
Female	13	13		10	4	
Age (years)						
<65	10	10	0.337	12	4	0.064
≥65	13	7		3	7	
Brain metastases						
Present	6	5	0.816	6	0	0.055
Absent	17	12		9	11	
Liver metastasis						
Present	3	0	0.347	1	1	1.000
Absent	20	17		14	10	
Bone metastases						
Present	11	7	0.676	6	4	1.000
Absent	12	10		9	7	

Abbreviation: EGFR: Epidermal growth factor receptor.

$P < 0.05$ were selected. Subsequently, a multivariate logistic regression model was constructed by employing stepwise backward variable selection based on these significant clinical features.

2.5. Construction of the DLM

The construction of the DLM involved feature extraction and classifier construction. In this study, BotNet18 was used as the model for extracting DL features, with its specific structures shown in Figure 1. The feature extraction process was as follows: first, a pre-trained BotNet18 model, trained on the ImageNet dataset, was imported to accelerate model convergence. Then, the pre-trained model was retrained using local T790 data until it reached the pre-defined training epoch number (120 epochs in the study) or the model converged, at which point the training was stopped. Finally, the outputs of all convolutional layers' kernels in the BotNet18 model were extracted as DL features for each CT image. The DL features of multiple CT image slices from the same patient were averaged to obtain the patient's DL feature set. A total of 14,400 DL features were extracted using BotNet18, as illustrated in Figure 2. The Mann–Whitney U test was used to identify significant DL characteristics. To prevent overfitting in the classifier model due to redundant features, the maximum relevance minimum redundancy method was utilized for dimensionality reduction of the DL features. Finally, to construct the classifier model, the top 23% of the most valuable and least redundant DL characteristics were selected.

The construction of the classifier involved several key steps. Initially, DL features with non-zero coefficients were identified in the training set using least absolute shrinkage and selection operator (Lasso) logistic regression combined with threefold cross-validation. These features were deemed valuable for distinguishing between *EGFR T790M*-positive and *EGFR T790M*-negative cases. Subsequently, the selected features and their corresponding coefficients were integrated to generate the classifier's predictive values.

2.6. Model evaluation

To calibrate and evaluate the DLM, DCA and calibration curve analysis were employed. For a comprehensive validation of the model's outcomes, the study utilized the receiver operating characteristic (ROC) curve. In addition, a range of metrics was utilized to comprehensively assess the model's performance, including the AUC, negative predictive value, positive predictive value, sensitivity, specificity, and accuracy.

2.7. Statistics

All data were analyzed using R language and Python version 3.7. Categorical variables (such as sex and smoking status) were investigated by the Chi-square test and continuous variables (such as age) by the Wilcoxon rank-sum test. All statistical tests were performed on both tails, and significance was determined using a p-value threshold of < 0.05 . Based on the Akaike information criterion, factors/variables showing significant differences ($k > 0.60$) were employed. A multivariate logistic regression clinical model was then constructed utilizing stepwise backward variable selection.

3. Results

3.1. Results and analysis of the clinical model

In constructing the clinical model, 40 patients were randomly assigned to the training cohort (male, $n = 14$; female, $n = 26$). Statistical tests on all clinical factors showed that *EGFR T790M* mutation status was significantly related to two clinical factors: *EGFR19del* and the use of gefitinib ($P < 0.05$). However, *EGFR T790M* mutation status was not significantly related to any other factors, such as age, sex, and brain, bone, and liver metastases ($P > 0.05$). Specific details are provided in Table 1.

Considering the above factors, the clinical model was built by multivariate logistic regression analysis, which included *EGFR19* factor (odds ratio [OR], 3.279; 95% confidence intervals [CI], 1.127 – 8.462; $P = 0.004$), use of

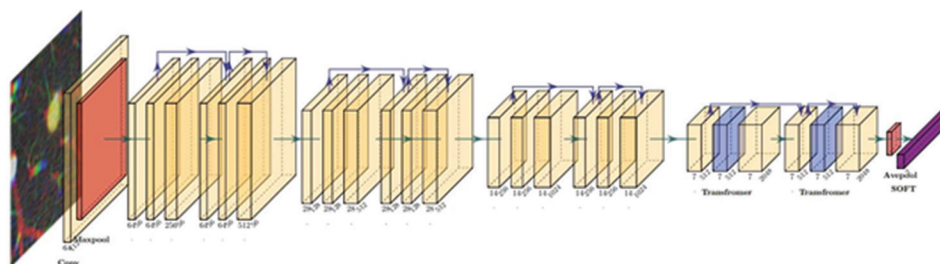


Figure 1. Structure of the deep-learning model. This model comprises 23 convolutional layers (yellow blocks), two multihead self-attention layers (blue blocks), two max-pooling layers (red blocks), and one fully connected layer (purple block).

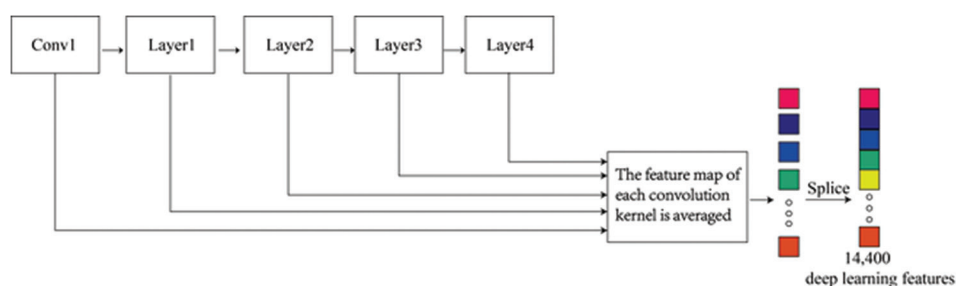


Figure 2. Feature extraction structure diagram. The mean value of each convolutional kernel feature map corresponds to a depth feature in the model, and all convolutional kernel features are extracted, and by column splicing, a total of 14,400 computed tomography image depth features are obtained.

gefitinib (OR, 2.624; 95% CI, 1.140 – 5.297; $P = 0.021$), and constant (OR, 2.671; 95% CI, 1.092 – 5.986; $P = 0.014$). The AUCs of the clinical model were 0.862 (95% CI, 0.570 – 0.966) and 0.755 (95% CI, 0.566 – 0.943) in the training and validation cohorts, respectively.

3.2. DLM diagnostic performance

In this study, 43 features exhibited significant differences in *EGFR T790M* expression, and the top 10 most valuable and least redundant features were selected to construct a Lasso classifier. The Lasso feature selection process is presented in Figure 3. Figure 3A shows the tuning parameter λ selection, where three-fold cross-validation based on the minimum criterion was used. The Log λ values are plotted on the X-axis, and classification errors are plotted on the Y-axis. The dashed lines represent the optimal values determined by the one standard error of the minimum criterion, with the optimal $\lambda = 0.000315$, Log $\lambda = -8.0629$. Figure 3B illustrates the Lasso coefficients for the selected features. According to the three-fold cross-validation, 11 non-zero coefficients of the best DL features were identified. The AUCs of the DLM were 0.839 (95% CI, 0.708 – 0.970) and 0.842 (95% CI, 0.680 – 1.000) in the training and validation cohorts, respectively. In Figure 4, the DCA results indicate that the DLM provides a higher net benefit in distinguishing *EGFR T790M* expression compared with the clinical model. The calibration curves presented in Figure 5 reveal that the predictions from the DLM are closer to the ideal probability curve than those from the clinical model, indicating better predictive performance.

3.3. Comparison of diagnostic performance of models

Figure 6 displays the ROC curves for both models. Compared with the clinical model (0.755), the DLM exhibits a significantly higher AUC (0.842) in the validation cohort, indicating that the diagnostic efficiency of the DLM is significantly better than that of the clinical model. The specific diagnostic performance of the clinical model and DLM is summarized in Table 2.

4. Discussion

Molecular targeted therapy is an important component of the treatment for patients with lung cancer, including adjuvant therapy, consolidation therapy, and maintenance therapy. First-/second-generation EGFR-TKIs are the standard initial treatment for patients with advanced EGFR-positive NSCLC. However, only patients harboring the *EGFR T790M* mutation can derive benefit from third-generation EGFR-TKIs if the first-/second-generation EGFR-TKIs are ineffective.^{38,39} Compared with chemotherapy and immunotherapy, third-generation EGFR-TKIs (osimertinib, almonertinib, etc.) are characterized by oral delivery, high efficacy, low risk, and short hospital stays. Moreover, predicting the *EGFR T790M* mutation status before starting first-line treatment can greatly optimize the management of patients with advanced EGFR-positive NSCLC. In this study, by analyzing pretreatment chest CT images, a clinical factor model and a DLM were constructed to predict *EGFR T790M* mutation status following the initial treatment (first-/second-generation EGFR-TKIs). Compared with our previous DLM based on PD-L1 using ResNet18, the present model adopted BotNet18 method, and natural images (ImageNet) were used to enhance its ability to extract features relevant to the present task, thereby improving its diagnostic performance. As expected, the DLM showed great diagnostic performance compared with the clinical model.

The AUC of the clinical model was 0.755, which represented fair predictive performance. The clinical model identified EGFR19del and gefitinib use as associated with *EGFR T790M* mutation status following first- or second-generation EGFR-TKI therapy, similar to previously reported results, which may be due to the longer progression-free survival that promoted the continuous activation of the EGFR pathway to induce *EGFR T790M* mutation.⁴⁰⁻⁴⁴ The molecular mechanism underlying the close correlation between EGFR19del and T790M is linked to intrinsic structural characteristics of EGFR19, in

Table 2. Diagnostic performance tables for the models

Classification metrics	Training cohort (n=40)		Validation cohort (n=26)	
	Clinical model	Deep-learning model	Clinical model	Deep-learning model
AUC (95%CI)	0.862 (0.757 – 0.966)	0.839 (0.708 – 0.970)	0.755 (0.566 – 0.943)	0.842 (0.680 – 1.000)
Sensitivity	0.4706 (8/17)	0.2273 (5/22)	0.9091 (10/11)	0.8182 (9/11)
Specificity	0.9565 (22/23)	1.0000 (18/18)	0.2000 (3/15)	0.6667 (10/15)
Accuracy	0.7500 (30/40)	0.5750 (23/40)	0.5000 (13/26)	0.7308 (19/26)
PPV	0.8889 (8/9)	1.0000 (5/5)	0.4545 (10/2)	0.6429 (9/14)
NPV	0.7097 (22/31)	0.5143 (18/35)	0.7500 (3/4)	0.8333 (10/12)

Abbreviations: AUC: Area under the curve; CI: Confidence interval; PPV: Positive predictive value; NPV: Negative predictive value.

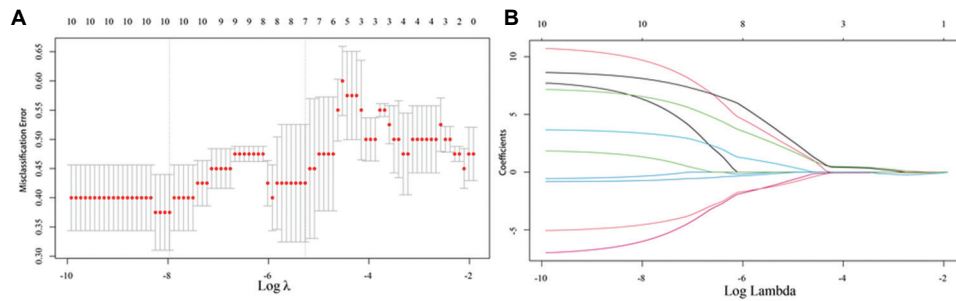


Figure 3. Lasso feature selection diagram. (A) The Lasso tuning parameters λ are filtered based on the 3-fold cross-validation of the minimum criteria. The dashed line indicates the optimal determined by the minimum criterion $\lambda = 0.000315$, and $\text{Loge } \lambda = -8.0629$; (B) Lasso coefficient diagram of all features. The horizontal coordinate is the tuning parameter λ , the vertical coordinate represents the corresponding coefficient of the depth feature, and the vertical dashed line λ is the value of the optimal parameter. When selecting the optimal parameters λ , a total of 11 depth features with nonzero coefficients are screened to build the regression model.

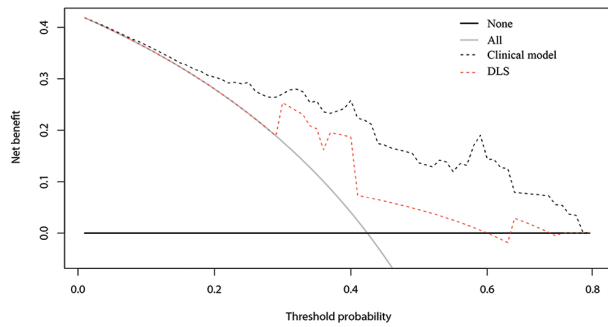


Figure 4. Decision curves of different models. The Y-axis denotes the net advantage of the model, whereas the X-axis is the probability of the decision threshold, representing the truncation threshold for patients with a positive model prediction under that threshold. The gray solid line indicates positive epidermal growth factor receptor *T790M* expression in all patients. The solid black line represents the benefit of negative epidermal growth factor receptor *T790M* expression in all patients. Abbreviations: DLS: Deep learning signature.

which 3 – 8 residues are shifted away from the loop leading into the α C-helix, whereas the *EGFR21 L858R* mutation happened in the activation loop.⁴⁵ Gefitinib, which has a 4-anilinoquinazoline core, could form a hydrogen bond with the *EGFR T790* residue. *T790M* mutation confers resistance to gefitinib, thereby allowing EGFR to maintain

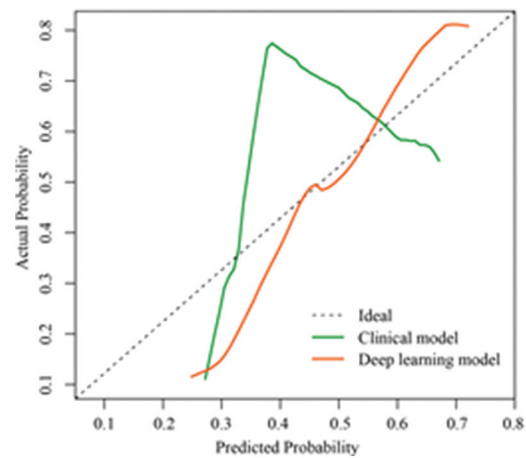


Figure 5. Correction curves of the two models. The abscissa and ordinate represent the predicted and actual probability values, respectively. The dotted line indicates the ideal model curve, and the green and red solid lines signify the performance curves of the clinical and deep-learning models, respectively.

its signaling activity and stimulate tumor cell proliferation.⁴⁶ The accuracy and specific underlying mechanisms need further validation. The construction of the clinical model provides us with a reference for predicting *EGFR T790M*

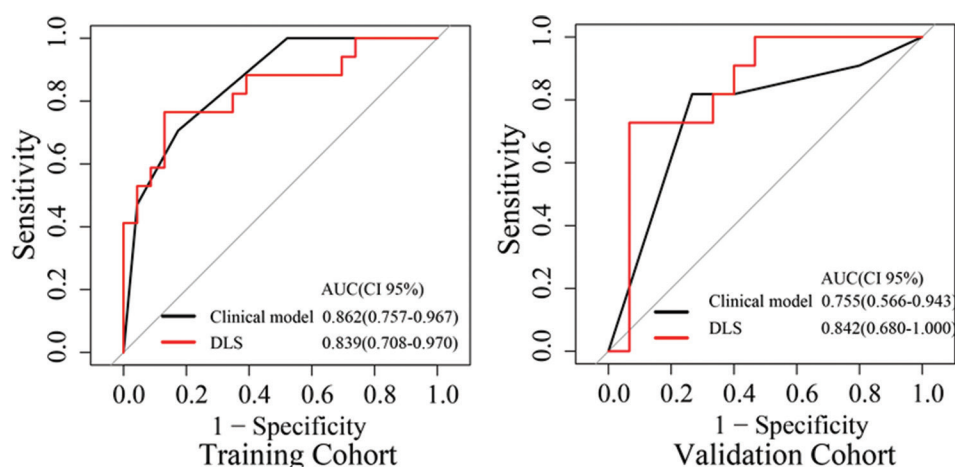


Figure 6. ROC curves for the training (left) and validation (right) sets of different models, including the clinical (black line) and deep-learning (red line) models

Abbreviations: DLS: Deep learning signature; AUC: Area under ROC curve; CI: Confidence interval; ROC: Receiving operating characteristics.

mutation following the initial EGFR-TKI therapy (first or second generation).

In addition, radiomic analysis of chest CT images was performed to predict *EGFR T790M* mutation status. The DLM demonstrated superior performance to the clinical model in the validation dataset (AUC = 0.842; 95% CI, 0.680 – 1.000 vs. AUC = 0.755; 95% CI, 0.566 – 0.943).

The proposed DL method offers the following advantages:

- (1) Detailed focus on lesion boundaries is no longer needed: DLMs can automatically learn relevant features without requiring precise delineation of lesion boundaries.
- (2) Extraction of multilayer abstract features: The method can extract high-level abstract features from the images, which are crucial for distinguishing *EGFR T790M* mutation status.
- (3) Convenient use without complex preprocessing: The model can validate raw CT images without needing intricate preprocessing steps, making it user-friendly and efficient. The superior diagnostic performance of the DLM highlights its potential utility in guiding initial treatment decisions and comprehensive patient management.

This study has several limitations: first, its retrospective design introduces potential selection bias. Second, the sample size is small and consists of data from a single center, with validation performed only internally. Therefore, larger prospective studies are necessary to validate the model's generalizability. Fortunately, the imaging reconstruction model used a slice thickness of 1.0 mm. Third, some samples relied on blood specimens for detecting the

EGFR T790M mutation due to difficulties in obtaining tissue samples, with the advantage that patients are spared the pain of secondary puncture. Thus, to enhance the discriminative capability of our model for *EGFR T790M* expression, future research should involve the collection of additional clinical data.

5. Conclusion

In this study, a non-invasive clinical model and a DLM derived from pre-treatment chest CT image data were constructed to predict the *EGFR T790M* status in patients with advanced NSCLC after failure of first-line EGFR-TKI therapy (first or second generation). The DLM exhibited superior diagnostic performance to the clinical model, providing strong support for oncologists in formulating initial treatment strategies and comprehensive management for patients newly diagnosed with advanced EGFR-positive NSCLC.

Acknowledgment

None.

Funding

None.

Conflict of interest

Huijing Feng is an Editorial Board Member of this journal but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. Separately, other authors declared that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Author contributions

Conceptualization: Peng Min Liu, Bao Feng, Jun Ping Zhang, Hui Jing Feng

Formal analysis: Peng Min Liu, Jiang Feng Shi, Shan Wu, Ye Hang Chen, Bao Feng, Hui Jing Feng

Investigation: Peng Min Liu, Shan Wu, Hui Jing Feng, Bao Feng

Methodology: Jiang Feng Shi, Shan Wu, Ye Hang Chen, Bao Feng

Writing – original draft: Peng Min Liu, Jiang Feng Shi

Writing – review & drafting: Hui Jing Feng, Bao Feng

Ethics approval and consent to participate

This work was approved by the ethics committee of Shanxi Bethune Hospital (approval ID: YXLL-2023-004).

Consent for publication

Not applicable.

Availability of data

Data will be accessible upon making a reasonable request to the corresponding author.

References

- Bray F, Laversanne M, Sung H, *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74(3):229-263.
doi: 10.3322/caac.21834
- Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin.* 2024;74(1):12-49.
doi: 10.3322/caac.21820
- Sung H, Ferlay J, Siegel RL, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209-249.
doi: 10.3322/caac.21660
- Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. *Ann Glob Health.* 2019;85(1):8.
doi: 10.5334/aogh.2419
- Gillespie CS, Mustafa MA, Richardson GE, *et al.* Genomic alterations and the incidence of brain metastases in advanced and metastatic NSCLC: A systematic review and meta-analysis. *J Thorac Oncol.* 2023;18(12):1703-1713.
doi: 10.1016/j.jtho.2023.06.017
- Chen P, Liu Y, Wen Y, Zhou C. Non-small cell lung cancer in China. *Cancer Commun (Lond).* 2022;42(10):937-970.
doi: 10.1002/cac2.12359
- Lahiri A, Maji A, Potdar PD, *et al.* Lung cancer immunotherapy: Progress, pitfalls, and promises. *Mol Cancer.* 2023;22(1):40.
doi: 10.1186/s12943-023-01740-y
- Miller M, Hanna N. Advances in systemic therapy for non-small cell lung cancer. *BMJ.* 2021;375:n2363.
doi: 10.1136/bmj.n2363
- Wu J, Lin Z. Non-small cell lung cancer targeted therapy: Drugs and mechanisms of drug resistance. *Int J Mol Sci.* 2022;23(23):15056.
doi: 10.3390/ijms232315056
- Kim ES, Melosky B, Park K, Yamamoto N, Yang JC. EGFR tyrosine kinase inhibitors for EGFR mutation-positive non-small-cell lung cancer: Outcomes in Asian populations. *Future Oncol.* 2021;17(18):2395-2408.
doi: 10.2217/fon-2021-0195
- Batra U, Biswas B, Prabhaskar K, Krishna MV. Differential clinicopathological features, treatments and outcomes in patients with exon 19 deletion and exon 21 L858R EGFR mutation-positive adenocarcinoma non-small-cell lung cancer. *BMJ Open Respir Res.* 2023;10(1):e001492.
doi: 10.1136/bmjresp-2022-001492
- Chen J, Yang H, Teo ASM, *et al.* Genomic landscape of lung adenocarcinoma in East Asians. *Nat Genet.* 2020;52(2):177-186.
doi: 10.1038/s41588-019-0569-6
- Riely GJ, Wood DE, Ettinger DS, *et al.* Non-small cell lung cancer, version 4.2024, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw.* 2024;22(4):249-274.
doi: 10.6004/jnccn.2204.0023
- Hendriks LE, Kerr KM, Menis J, *et al.* Oncogene-addicted metastatic non-small-cell lung cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol.* 2023;34(4):339-357.
doi: 10.1016/j.annonc.2022.12.009
- Hosomi Y, Morita S, Sugawara S, *et al.* Gefitinib alone versus gefitinib plus chemotherapy for non-small-cell lung cancer with mutated epidermal growth factor receptor: NEJ009 study. *J Clin Oncol.* 2020;38(2):115-123.
doi: 10.1200/jco.19.01488
- Jiang Z, Zhang J, Sun H, *et al.* Icotinib alone or with bevacizumab as first-line therapy in Chinese patients with advanced nonsquamous non-small cell lung cancer and activating EGFR mutations: A retrospective study. *Thorac Cancer.* 2021;12(17):2369-2374.
doi: 10.1111/1759-7714.14079
- Goldberg SB, Redman MW, Lilenbaum R, *et al.* Randomized trial of afatinib plus cetuximab versus afatinib alone for

- first-line treatment of EGFR-mutant non-small-cell lung cancer: Final results from SWOG S1403. *J Clin Oncol*. 2020;38(34):4076-4085.
doi: 10.1200/jco.20.01149
18. Wu YL, Cheng Y, Zhou X, *et al*. Dacomitinib versus gefitinib as first-line treatment for patients with EGFR-mutation-positive non-small-cell lung cancer (ARCHER 1050): A randomised, open-label, phase 3 trial. *Lancet Oncol*. 2017;18(11):1454-1466.
doi: 10.1016/s1470-2045(17)30608-3
 19. Kanaji N, Yokohira M, Nakano-Narusawa Y, *et al*. Hepatocyte growth factor produced in lung fibroblasts enhances non-small cell lung cancer cell survival and tumor progression. *Respir Res*. 2017;18:118.
doi: 10.1186/s12931-017-0604-z
 20. Dong RF, Zhu ML, Liu MM, *et al*. EGFR mutation mediates resistance to EGFR tyrosine kinase inhibitors in NSCLC: From molecular mechanisms to clinical research. *Pharmacol Res*. 2021;167:105583.
doi: 10.1016/j.phrs.2021.105583
 21. He J, Huang Z, Han L, Gong Y, Xie C. Mechanisms and management of 3rd-generation EGFR-TKI resistance in advanced non-small cell lung cancer (review). *Int J Oncol*. 2021;59(5):90.
doi: 10.3892/ijo.2021.5270
 22. Wu SG, Shih JY. Management of acquired resistance to EGFR TKI-targeted therapy in advanced non-small cell lung cancer. *Mol Cancer*. 2018;17(1):38.
doi: 10.1186/s12943-018-0777-1
 23. Borgeaud M, Parikh K, Banna GL, *et al*. Unveiling the landscape of uncommon EGFR mutations in NSCLC-a systematic review. *J Thorac Oncol*. 2024;19(7):973-983.
doi: 10.1016/j.jtho.2024.03.016
 24. Patil BR, Bhadane KV, Ahmad I, *et al*. Exploring the structural activity relationship of the Osimertinib: A covalent inhibitor of double mutant EGFR^{L858R/T790M} tyrosine kinase for the treatment of Non-small Cell Lung Cancer (NSCLC). *Bioorg Med Chem*. 2024;109:117796.
doi: 10.1016/j.bmc.2024.117796
 25. Nagano T, Tachihara M, Nishimura Y. Mechanism of resistance to epidermal growth factor receptor-tyrosine kinase inhibitors and a potential treatment strategy. *Cells*. 2018;7(11):212.
doi: 10.3390/cells7110212
 26. Johnson M, Garassino MC, Mok T, Mitsudomi T. Treatment strategies and outcomes for patients with EGFR-mutant non-small cell lung cancer resistant to EGFR tyrosine kinase inhibitors: Focus on novel therapies. *Lung Cancer*. 2022;170:41-51.
doi: 10.1016/j.lungcan.2022.05.011
 27. Christopoulos P, Kirchner M, Roeper J, *et al*. Risk stratification of EGFR⁺ lung cancer diagnosed with panel-based next-generation sequencing. *Lung Cancer*. 2020;148:105-112.
doi: 10.1016/j.lungcan.2020.08.007
 28. Canale M, Petracci E, Delmonte A, *et al*. Impact of TP53 mutations on outcome in EGFR-mutated patients treated with first-line tyrosine kinase inhibitors. *Clin Cancer Res*. 2017;23(9):2195-2202.
doi: 10.1158/1078-0432.CCR-16-0966
 29. Christopoulos P, Budczies J, Kirchner M, *et al*. Defining molecular risk in ALK⁺ NSCLC. *Oncotarget*. 2019;10(33):3093-3103.
doi: 10.18632/oncotarget.26886
 30. Offin M, Rizvi H, Tenet M, *et al*. Tumor mutation burden and efficacy of EGFR-tyrosine kinase inhibitors in patients with EGFR-mutant lung cancers. *Clin Cancer Res*. 2019;25(3):1063-1069.
doi: 10.1158/1078-0432.CCR-18-1102
 31. Menzel M, Kirchner M, Kluck K, *et al*. Genomic heterogeneity at baseline is associated with T790M resistance mutations in EGFR-mutated lung cancer treated with the first-/second-generation tyrosine kinase inhibitors. *J Pathol Clin Res*. 2024;10(2):e354.
doi: 10.1002/cjp.2.354
 32. Min Kim H, Ko T, Young Choi I, Myong JP. Asbestosis diagnosis algorithm combining the lung segmentation method and deep learning model in computed tomography image. *Int J Med Inform*. 2021;158:104667.
doi: 10.1016/j.ijmedinf.2021.104667
 33. Wang S, Liu Z, Chen X, *et al*. Unsupervised deep learning features for lung cancer overall survival analysis. *Annu Int Conf IEEE Eng Med Biol Soc*. 2018;2018:2583-2586.
doi: 10.1109/embs.2018.8512833
 34. Chen X, Feng B, Chen Y, *et al*. A CT-based deep learning model for subsolid pulmonary nodules to distinguish minimally invasive adenocarcinoma and invasive adenocarcinoma. *Eur J Radiol*. 2021;145:110041.
doi: 10.1016/j.ejrad.2021.110041
 35. Chen X, Feng B, Xu K, *et al*. Development and validation of a deep learning radiomics nomogram for preoperatively differentiating thymic epithelial tumor histologic subtypes. *Eur Radiol*. 2023;33(10):6804-6816.
doi: 10.1007/s00330-023-09690-1
 36. Feng B, Shi J, Huang L, *et al*. Robustly federated learning model for identifying high-risk patients with postoperative gastric cancer recurrence. *Nat Commun*. 2024;15(1):742.

- doi: 10.1038/s41467-024-44946-4
37. Liu PM, Feng B, Shi JF, *et al.* A deep-learning model using enhanced chest CT images to predict PD-L1 expression in non-small-cell lung cancer patients. *Clin Radiol.* 2023;78(10):e689-e697.
doi: 10.1016/j.crad.2023.05.010
 38. Mok TS, Wu YL, Ahn MJ, *et al.* Osimertinib or platinum-pemetrexed in EGFR T790M-positive lung cancer. *N Engl J Med.* 2017;376(7):629-640.
doi: 10.1056/NEJMoa1612674
 39. Blair HA. Befotertinib: First approval. *Drugs.* 2023;83(15):1433-1437.
doi: 10.1007/s40265-023-01946-w
 40. Cheng WC, Hsia TC, Tu CY, Chen HJ. The impact of acquired EGFR T790M mutation and EGFR circulating cell-free DNA on survival in patients with lung adenocarcinoma following EGFR-TKI therapy. *Onco Targets Ther.* 2020;13:13425-13435.
doi: 10.2147/ott.S279540
 41. Lin YT, Chen JS, Liao WY, *et al.* Clinical outcomes and secondary epidermal growth factor receptor (EGFR) T790M mutation among first-line gefitinib, erlotinib and afatinib-treated non-small cell lung cancer patients with activating EGFR mutations. *Int J Cancer.* 2019;144(11):2887-2896.
doi: 10.1002/ijc.32025
 42. Nam Y, Kim HC, Kim YC, *et al.* Clinical impact of rebiopsy among patients with epidermal growth factor receptor-mutant lung adenocarcinoma in a real-world clinical setting. *Thorac Cancer.* 2021;12(6):890-898.
doi: 10.1111/1759-7714.13857
 43. Wang Z, Chen R, Wang S, *et al.* Quantification and dynamic monitoring of EGFR T790M in plasma cell-free DNA by digital PCR for prognosis of EGFR-TKI treatment in advanced NSCLC. *PLoS One.* 2014;9(11):e110780.
doi: 10.1371/journal.pone.0110780
 44. Lu S, Dong X, Jian H, *et al.* AENEAS: A randomized phase III trial of aumolertinib versus gefitinib as first-line therapy for locally advanced or metastatic non-small-cell lung cancer with EGFR exon 19 deletion or L858R mutations. *J Clin Oncol.* 2022;40(27):3162-3171.
doi: 10.1200/jco.21.02641
 45. Ke EE, Zhou Q, Zhang QY, *et al.* A higher proportion of the EGFR T790M mutation may contribute to the better survival of patients with exon 19 deletions compared with those with L858R. *J Thorac Oncol.* 2017;12(9):1368-1375.
doi: 10.1016/j.jtho.2017.05.018
 46. Dickerson H, Diab A, Al Musaimi O. Epidermal growth factor receptor tyrosine kinase inhibitors in cancer: Current use and future prospects. *Int J Mol Sci.* 2024;25(18):10008.
doi: 10.3390/ijms251810008