Thoracic Cancer

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Progression-Free Survival Prediction Model Based on AI-Enhanced Dynamic Radiomics for Personalized EGFR-TKI Treatment Monitoring Patients With Lung Adenocarcinoma

Yan'e Liu[1,2] | Xiangfeng Luo[3] | Lu Yang[1,2] | Xueliang Cheng[3] | Xin Zhu[1,2] | Hua Zhang[4] | Bolin Hou[3] | Baoshan Cao[1,2]

[1]Department of Medical Oncology and Radiation Sickness, Peking University Third Hospital, Beijing, China | [2]Cancer Center, Peking University Third Hospital, Beijing, China | [3]Linkdoc AI Research (LAIR), Linkdoc Information Technology (Beijing) Co. Ltd., Beijing, China | [4]Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing, China

Correspondence: Baoshan Cao (caobaoshan0711@aliyun.com)

Received: 25 December 2024 | **Revised:** 21 January 2025 | **Accepted:** 24 January 2025

Funding: This work was supported by the National Key R&D Program of China (2020AAA0105203) and the Innovation & Transfer Fund of Peking University Third Hospital (BYSYCY2024088).

Keywords: deep learning | epidermal growth factor receptor (EGFR) | lung adenocarcinoma | prognosis prediction | progression-free survival (PFS) | radiomics

ABSTRACT

Background and Objective: Epidermal growth factor receptor tyrosine kinase inhibitors (EGFR-TKIs) are the standard first-line treatment for patients with advanced lung adenocarcinoma (LUAD) with EGFR mutations. However, treatment effectiveness varies widely among individuals, and effective models to predict treatment response are lacking. This study aims to establish a progression-free survival (PFS) prediction model based on dynamic changes in pre- and post-treatment CT scans combined with patients' clinical features.

Methods: A total of 183 patients with advanced LUAD who received first-line treatment at Peking University Third Hospital from January 2013 to December 2022 were enrolled. A 3D-UNet model was fine-tuned using data from 405 patients with non-small cell lung cancer for advanced lesion segmentation. Clinical and radiomic features extracted using 3D models from 80 EGFR-mutant LUAD patients were used to develop PFS prediction models with a deep-learning binary classification model. The accuracy, specificity, sensitivity, AUC, and F1 score of the models were validated in patients with mutant and wild-type EGFR.

Results: In the EGFR-mutant test set ($N = 53$), the AUC for the 9-month and 12-month progression prediction models were 0.858 (95% CI, 0.707–0.972) and 0.873 (95% CI, 0.747–0.974). Their accuracies were 81.1% (95% CI, 69.8%–90.6%) and 84.9% (95% CI, 73.6%–94.3%), specificities were 87.5% and 72.2%, sensitivities were 80.0% and 91.4%, and F1 scores were 0.878 and 0.889, respectively.

Conclusion: This study developed treatment response prediction models for EGFR-mutant LUAD patients. These models demonstrated strong predictive value for PFS in patients treated with EGFR-TKIs, potentially enabling a more efficient personalized CT scan schedule.

1 | Introduction

Lung cancer remains a leading cause of morbidity and mortality worldwide, representing a significant global health challenge. In China, it has the highest incidence and mortality rates among all cancer types, posing a substantial public health burden [1]. Non-small cell lung cancer (NSCLC) is the most common type, with lung adenocarcinoma (LUAD) being the most prevalent histological subtype, accounting for approximately 40% of all lung cancers [2]. Epidermal growth factor receptor (EGFR) is the most common driver gene in LUAD patients, with a prevalence as high as 51.4% in Asian populations [3]. EGFR-tyrosine kinase inhibitors (EGFR-TKIs) have become the first-line treatment for unresectable or advanced NSCLC patients harboring EGFR mutations. However, treatment response significantly varies among individuals, highlighting the urgent need for effective predictive models to optimize therapeutic outcomes.

Clinically, tumor effectiveness evaluation is primarily evaluated by the Response Evaluation Criteria in Solid Tumors (RECIST), where the diameter or long axis of lesions is measured to assess the change in tumor size using x-rays, MRI, or computed tomography (CT). In lung cancer, CT scans are commonly used in RECIST, but two significant limitations affect accuracy: bias or errors from manual measurements and variability in medical standards across regions. In addition, routine CT scans performed on a fixed schedule during treatment lead to unnecessary radiation exposure and increased medical costs for patients with favorable responses, while potentially delaying timely detection of progression in those with poor responses. Therefore, an effective predictive model for treatment response and tumor progression can greatly benefit patients.

In recent years, artificial intelligence (AI) facilitated radiomics technology has been increasingly applied in lung cancer management, including lung cancer risk prediction [4], early lung cancer detection [5], and pathological staging prediction [6]. Previously, several progression-free survival (PFS) prediction models have been reported in EGFR-mutant NSCLC patients using traditional handcrafted radiomics methods such as DeepSurv [7], logistic regression models [8], Cox proportional hazard models [9, 10], stepwise regression [11], and LASSO-Cox [12]. With the aid of AI, PFS prediction models based on deep learning models and radiomics have been explored, including EfficientNetV2 architecture [13], BigBiGAN [14], LASSO-Cox [15], and 3D convolutional neural networks [16]. Although the accuracy of these models has been improved with AI, these studies primarily rely on single baseline CT scans for model prediction. Since the dynamic features of lesions were not considered in those models, they are not informative for personalized CT schedules. Although Zhang et al.'s study [17] demonstrated that imaging features of continuous follow-up could provide valuable information for assessing disease progression, they overlooked clinical features critical for lung cancer patients, such as smoking history and gene mutation type, which may affect the prognosis.

In this study, we aimed to establish an accurate model for predicting response for LUAD patients with EGFR-TKI treatment using deep learning approaches to analyze their dynamic radiological changes along with clinical features. Previously, we established an efficient 3D-RPNUNet model to detect lung nodules on CT imaging for cancer early detection [18]. To detect advanced lesions in this study, we first fine-tuned the 3D-UNet model with CT images of patients with advanced NSCLC. We employed the 3D-RPNUNet model and the fine-tuned 3D-UNet lung segment model to extract various features, including lesion location, volume, shape, and density, to calculate the malignancy possibility of lesions and identify target lesions for further analysis. Then, we incorporated radiomic features from both baseline and dynamic post-treatment CT images, as well as clinical features, to build an EGFR-TKI treatment response prediction model facilitated by convolutional neural networks. In the EGFR-TKIs treatment validation cohort, the area under the curve (AUC) of the 9-month PFS model reached 0.858, and the AUC of the 12-month progression prediction model reached 0.873. Therefore, our study provided a better predictive model for treatment effectiveness, which could benefit clinicians and patients with a potentially more efficient personalized CT scan schedule.

2 | Methods

2.1 | Participants

This study included patients with locally advanced or metastatic LUAD who received first-line treatment at the Department of Medical Oncology and Radiation Sickness, Peking University Third Hospital, from January 1, 2013 to December 31, 2022. We retrospectively analyzed a total of 438 patients, including 183 patients with EGFR mutations ($N = 133$) and wild-type (WT) EGFR ($N = 50$) (Figure 1).

The inclusion criteria were as follows: (1) aged 18 years and older; (2) pathologically confirmed LUAD; (3) newly diagnosed locally advanced or metastatic LUAD with no prior systemic treatment, including curative surgery, radiation treatment, chemotherapy, biologic therapy, immunotherapy, or any targeted therapies; (4) underwent a baseline chest thin-layer CT scan within 4 weeks before systemic therapy and two follow-up thin-layer CT scans conducted approximately 6 weeks after the initial treatment with subsequent scans scheduled every 6–12 weeks before the disease progressed; (5) at least one measurable targeted lesion that has not been irradiated, with the longest diameter of ≥ 10 mm on baseline thin-layer CT images; and (6) complete clinical data available.

The exclusion criteria were as follows: (1) incomplete clinical data or thin-layer CT images; (2) prior treatment with any systemic anti-cancer therapy for LUAD, including curative surgery, radiation treatment, chemotherapy, biologic therapy, immunotherapy, or any targeted therapies; and (3) lost to follow up (patients failed to attend regular follow-up appointments).

2.2 | Study Design

This study aimed to develop a PFS model for predicting the response of EGFR-TKI treatment for patients with advanced LUAD harboring EGFR mutations. Patient-level features and lesion-level features collectively served as inputs for prediction models. The models were created using a deep learning method to analyze

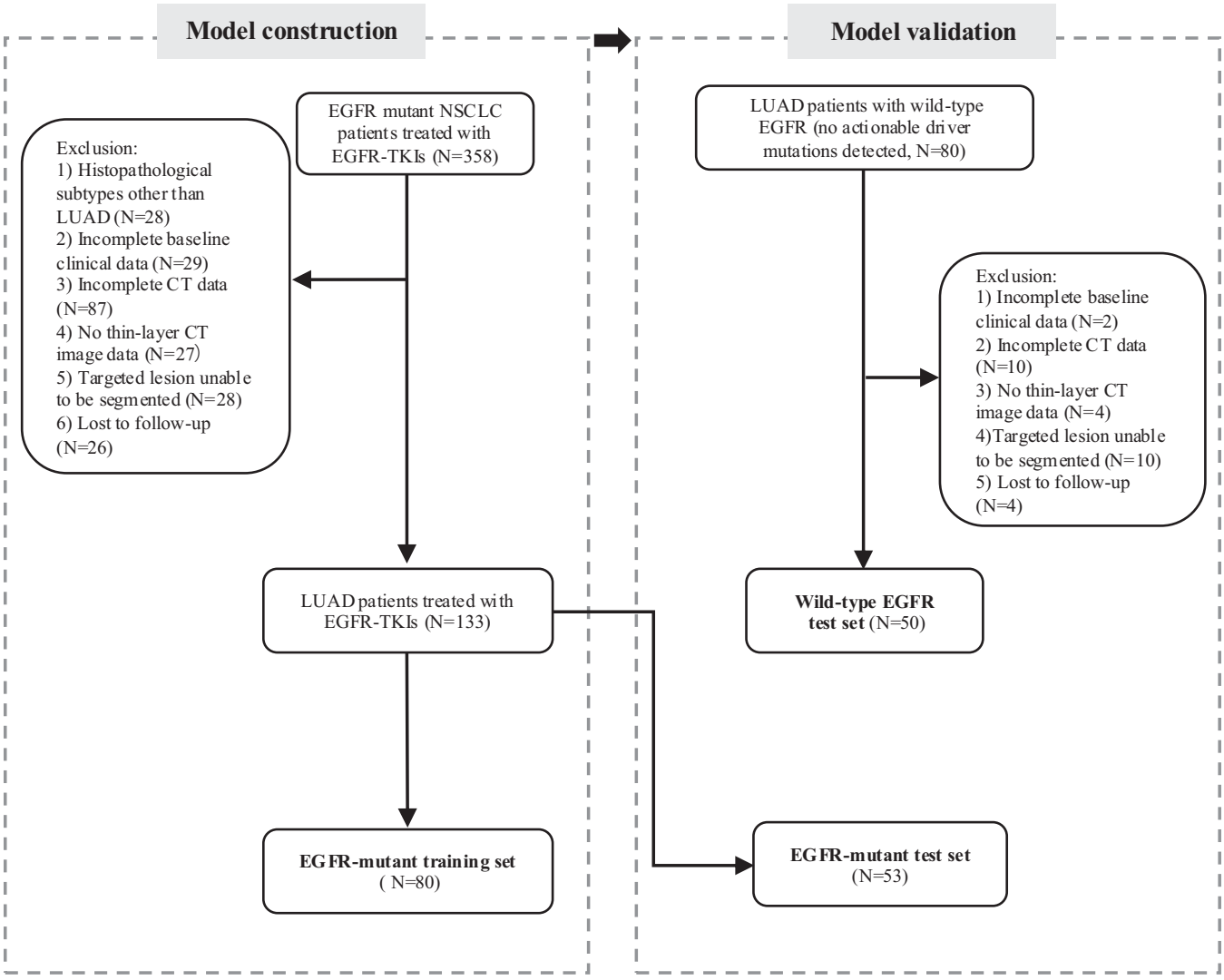

FIGURE 1 | Flow chart of the enrollment of patients with LUAD in the study. In total, 133 LUAD patients with mutant EGFR were treated with EGFR-TKIs. Among them, 80 patients were used to develop the model as a training set. The remaining 53 patients and another 50 LUAD patients with WT EGFR were used as two test sets to validate the performance of the model in different genetic backgrounds, respectively.

four pivotal features of target lesions of the patient, including lesion malignancy probability that developed in this study, patients' clinicopathologic characteristics at baseline, radiomic features detected with fine-tuned 3D convolutional models, and lesion location. The detailed procedures are illustrated in Figure 2.

Before lesion detection from the thin-layer CT images, a series of processes were applied: normalizing the CT pixel values (i.e., scale HU values proportionally to a range between 0 and 1), augmenting CT image data [19], and segmenting a bounding box of the entire lung. We then divided the segmented bounding box of the 3D convolutional model into multiple 3D patches with dimensions of $64 \times 128 \times 128$, significantly reducing the GPU requirements. These processed CT images could facilitate the accuracy of lesion identification, segmentation, and feature extraction.

To extract lesion-level features, four deep learning models were employed: the lesion detection model, the lesion segmentation model, the malignancy probability prediction model, and the lung lobe segmentation model. The lesion detection model was based on the 3D-RPNUNet framework for object detection. These processed CT images were input into the lesion detection model to determine the locations of lesions. The detected lesions were then used for feature extraction in the following aspects:

1. Malignancy probability prediction: A malignancy probability prediction model was built using the ResNet-18 framework. The processed CT images and the detected lesion location information were used in this model to predict the malignancy probability for each lesion, with the probability ranging from 0 to 1.
2. Lesion segmentation and radiomic feature extraction: The imaging and lesion locations were also fed into the lesion segmentation model based on the 3D-UNet framework to segment the contours for all detected lesions. Using the data from the lesion contours and bounding box, we applied Radiomics [20] to extract the radiomic

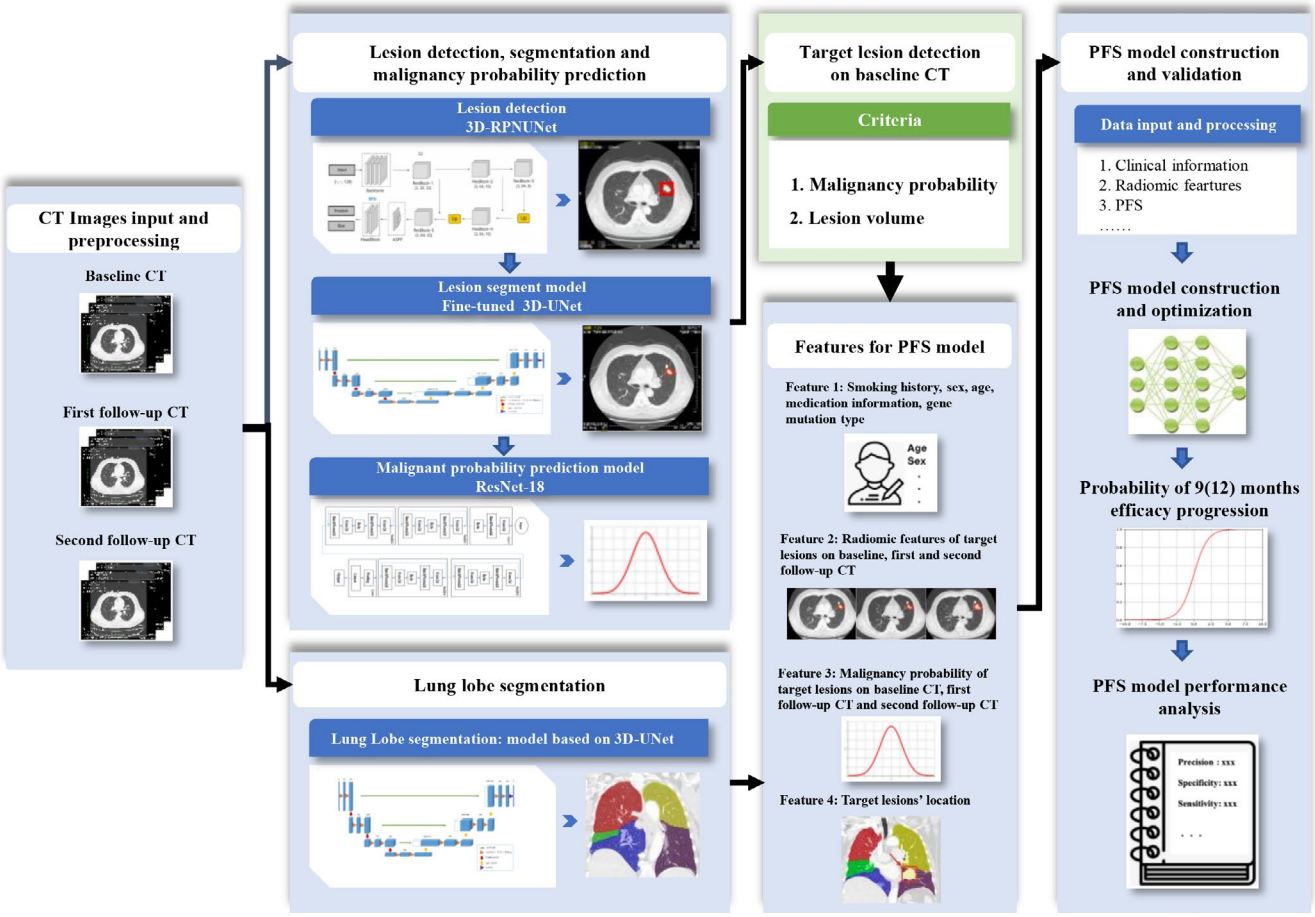

FIGURE 2 | Study design of EGFR-TKIs treatment response prediction model for patients with advanced LUAD. The technical process included three parts: data preparation, feature extraction, and PFS model construction and validation. Data preparation stage collected data from LUAD patients at the hospital, including patients' personal medical history information and CT imaging data. For radiomic feature extraction, four 3D models were utilized to obtain lesion-level features: the lesion identification model, malignancy probability prediction model, lesion segmentation model, and lobe segmentation model. To develop the PFS model, both patient- and lesion-level features were used as inputs.

features of the lesions, including the long diameter, short diameter, tumor volume, and the proportion of solid components.

3. Lesion lung lobe location: A lung lobe segmentation model based on the 3D-UNet framework was used to analyze the processed CT images and segment the lung lobes. Combined with the detected lesion location on CT images, every lesion in the lung lobe was located.

Based on the above lesion-level features of CT images at baseline and follow-up, the target lesions for developing the PFS predictive model were identified. First, target lesions identified in the baseline CT were the top two largest lesions with malignancy probability $\geq 50\%$. These lesions were also tracked as the target lesions in the follow-up CT images. Features of all identified target lesions were then extracted, completing feature extraction at the lesion dimension.

2.3 | Input Data Preparation

Patient clinicopathologic characteristics were collected by reviewing the Hospital Information System (HIS), including age, sex, pathological type, smoking status, EGFR mutation, tumor-node-metastasis (TNM) stage, types of EGFR-TKIs, medication start date, EGFR-TKI treatment response evaluation results, disease progression time, and progression site. Serial CT scans were obtained from Picture Archiving and Communication Systems (PACS), including baseline chest CT scans (within 4 weeks before initiation of therapy) and follow-up CT scans performed approximately 6 weeks after the initial treatment, with subsequent scans scheduled every 6–12 weeks thereafter.

2.4 | Lesion Segmentation Model Fine-Tuning for Patients With Advanced NSCLC

A lung lesion segmentation model based on 3D-UNet [21], enhanced by deep learning, was previously established to detect lesions in patients with lung cancer at early stages [22]. In this study, this model was fine-tuned to improve the accuracy of primary tumor segmentation in NSCLC patients, particularly for advanced stages. Specifically, 422 cases of NSCLC were screened from previously published data [23]. In total, CT images of 405 NSCLC cases were collected from patients with thin-layer CT scans of primary lesions whose format was compatible with

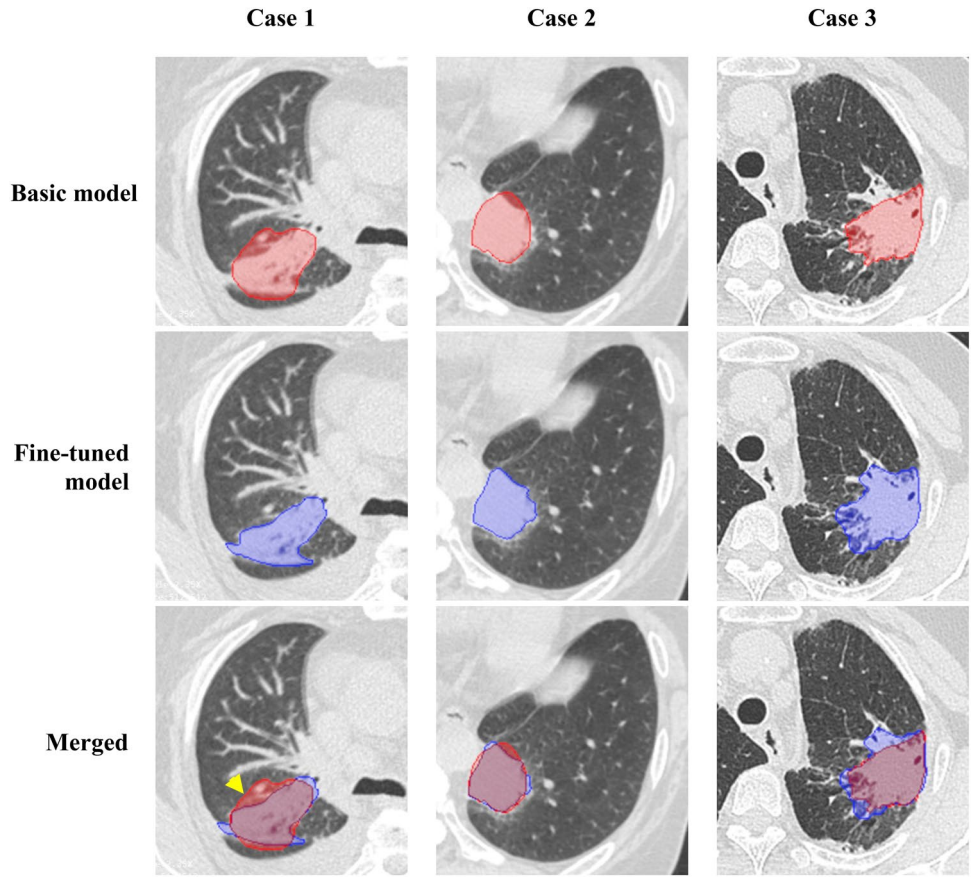

FIGURE 3 | Comparison of 3D-UNet CT segmentation models. The baseline CT images of three LUAD patients treated with EGFR-TKIs were analyzed using 3D-UNet models before and after fine-tuning. The targeted lesions were automatically marked by the AI model on the CT image, with red indicating the lesion identified by the model before fine-tuning and blue after fine-tuning. The arrow indicates a cross-section of a blood vessel, corrected in the fine-tuned model in Case 1.

the lesion segmentation model, and their baseline clinicopathological features were summarized in Table S1. To fine-tune the 3D-UNet model, 364 cases were used for training. The model was adjusted with an attention mechanism and attention loss function (SAD loss) [24] to allow the model to focus more on the regions of lesions. The fine-tuned 3D-UNet model was tested on the remaining 41 cases. Upon fine-tuning, the dice similarity coefficient of the lesion segmentation model improved from the original 0.68 to 0.77 on the test set. More accurate nodule segmentation enabled the extraction of precise lesion features, such as nodule volume, long diameter, and short diameter.

Figure 3 demonstrates the performance of the model in lesion segmentation before and after fine-tuning. The fine-tuned segmentation model showed more accurate contours of the lesion in the lung. For instance, it correctly differentiated a blood vessel from a cancer lesion in Case 1, indicating that the fine-tuned model can provide more precise tumor features. This can help improve the prediction of treatment response and progression probability.

2.5 | Treatment Response Evaluation Criteria

The response to treatment was assessed according to RECIST version 1.1. Disease progression is determined when the diameter of targeted lesions increases by 20% or the appearance of new lesions. Extrathoracic progression is considered if these criteria are met for lesions located outside the thoracic area. For patients with EGFR-TKI treatment, disease progression might occur in various locations, including the lungs, brain, liver, and bone. These could be categorized into three types: intrapulmonary progression, extrapulmonary progression, and both intrapulmonary and extrapulmonary. PFS was defined as the duration from initiation of systemic therapy to disease progression.

2.6 | PFS Model

The true PFS for each patient was accurately obtained from the HIS in this study. A 9-month disease progression probability prediction model was built to predict whether PFS would be shorter than 9 months. If progression occurred within 9 months, the label value was set to 0 (indicating a 0% probability). If progression occurred at 9 months or later, the label value was set to 1 (indicating a probability of 100%). A similar model for 12-month disease progression probability prediction was developed using the same scoring system.

The 9-month and 12-month disease progression prediction models were developed using a multilayer neural network. The input layer matched the same dimension of the features, the hidden

layer contained 128 neurons with ReLU [25] functions as activation functions, and the output layer used the Softmax [26] function to obtain the progression prediction probability. During model training, Adaptive Moment Estimation [27] was used as the optimizer, and cross-entropy [28] was used as the loss function. The learning rate was set to 1e−4, and the programming tool used was PyTorch [29].

2.7 | Statistics

Clinical data were statistically analyzed using SPSS 27.0 software. Continuous variables with skewed or unknown distributions were expressed as median (range), and comparisons between two groups were performed using the Mann–Whitney *U*-test. Categorical data were described as counts (%) and compared between groups using Pearson's chi-square test or Fisher's exact test. A p value of <0.05 was considered statistically significant.

When testing model performance, accuracy, specificity, sensitivity (also known as recall), AUC, and F1 score were analyzed.

For the predicted results:

a. When the model's predicted probability exceeds the probability threshold, the prediction is considered a positive sample.
b. When the predicted probability is less than or equal to the probability threshold, the prediction is considered a negative sample.

The calculation methods for accuracy and F1 score are shown in Formulas (1) and (2).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

Where Precision is the ratio of true positive (TP) to the sum of true positive and false positive (FP), and Recall is the ratio of TP to the sum of TP and false negative (FN).

3 | Results

3.1 | Characteristics of Patients

Our study enrolled 183 LUAD cases, including 133 patients who received EGFR-TKI treatment and 50 WT-EGFR patients. Consequently, 80 EGFR-TKI-treated cases were used to develop a PFS prediction model as the training set, while the testing data comprised the remaining 53 cases. Additionally, the data of 50 patients with WT-EGFR were used as a comparative test set to further assess the prediction model performance.

The baseline clinicopathological features of participants in the EGFR-mutant training and test sets and the WT-EGFR test set are summarized in Table 1. The EGFR-mutant training set consisted of 80 patients with a median age of 67 years. Among these patients, 38.8% were male, and 15% were smokers. Disease progression occurred in 65% of the patients with a median PFS (mPFS) of 12.5 months. The EGFR-mutant test set included 53 patients with a median age of 64 years, 37.7% of males and 20.8% of smokers. About 69.8% of patients had disease progressed with mPFS of 12.5 months. As shown in Table 1, no significant discrepancies in baseline characteristics were observed between the training and test sets. A total of 50 WT-EGFR patients were used as the test set to examine the specificity of the model. Their median age was 66 years, with 76% of participants being male. Smokers were 30.0%, and 78.0% of participants had disease progressed whose mPFS was 8.4 months (Table 1).

3.2 | Construction of PFS Prediction Model

To develop a PFS prediction model, we determined the target lesions (see Method) and extracted their radiomic features from CT images at baseline and two follow-ups. We combined these radiomic features with patient baseline clinicopathological features, target lesion locations, and predicted malignancy probability that was used to determine the target lesions, which were further processed with a deep learning model. The developed PFS prediction model reached an AUC of 0.938 (95% CI, 0.879–0.984) for the 9-month progression prediction and 1.0 (95% CI, 1.0–1.0) for the 12-month progression prediction (Figure 4).

Prior to the development of the PFS prediction model, we speculated that the precision of lesion contouring could affect the accuracy of the model. Therefore, we fine-tuned the 3D-UNet model to improve the lesion contour for advanced LUAD patients (see Method). To validate the effectiveness of this process, we extracted the lesion features using the model before and after fine-tuning and PFS prediction models were constructed. The accuracy, specificity, sensitivity, AUC, and F1 score were compared between fine-tuned and non-fine-tuned models using the same training and test sets (Table 2). The AUC was improved in both models in the training set: 0.801 (95% C1, 0.665–0.906) vs. 0.938 (95% CI, 0.879–0.984) for 9-month PFS prediction and 0.701 (95% CI, 0.567–0.824) vs. 1.00 (95% CI, 1.00–1.00) for 12 months. These results indicate that a better segmentation model significantly improves the performance of the PFS prediction model.

3.3 | Validation of PFS Prediction Model

We validated the performance of the PFS prediction model in a test set of 53 cases of EGFR-TKI treatment. The AUC for the 9-month and 12-month progression prediction models were 0.858 (95% CI, 0.707–0.972) and 0.873 (95% CI, 0.747–0.974), respectively, in the test set (Figure 4). Except for the 12-month progression prediction specificity of 72.2%, the accuracy, specificity, sensitivity, and AUC on the EGFR-mutant test set were all $>80\%$ or 0.8 (Table 2). Considering that we developed this PFS prediction model focused on patients with EGFR mutations, we further investigated its specificity in WT-EGFR cases. Therefore, we applied the PFS model to a WT-EGFR set of 50 cases. Among these patients, the AUC for the 9-month progression prediction

TABLE 1 | The baseline clinicopathological features of 183 participants in this study.

Characteristic	EGFR-mutant training set ($N=80$)	EGFR-mutant test set ($N=53$)	p	WT-EGFR test set ($N=50$)
Age, yr			0.192	
Median	67	64		66
Range	35–90	31–82		47–84
Sex, no. (%)			0.906	
Female	49 (61.2)	33 (62.3)		12 (24)
Male	31 (38.8)	20 (37.7)		38 (76)
Smoking status, no. (%)			0.626	
Never smoked	37 (46.2)	21 (39.6)		9 (18)
Smoker	12 (15)	11 (20.8)		15 (30)
Unknown	31 (38.8)	21 (39.6)		26 (52)
Type of gene mutation, no. (%)			0.287	
EGFR 19del	38 (47.5)	20 (37.7)		0 (0)
EGFR 21L858R	36 (45)	25 (47.2)		0 (0)
Rare mutation of EGFR	6 (7.5)	8 (15.1)		0 (0)
EGFR wild type	0 (0)	0 (0)		50 (100)
EGFR-TKI treatment, no. (%)			0.724	
Erlotinib	4 (5)	1 (1.9)		0 (0)
Icotinib	18 (22.5)	9 (17)		0 (0)
Gefitinib	32 (40)	21 (39.6)		0 (0)
Dacomitinib	2 (2.5)	3 (5.7)		0 (0)
Afatinib	16 (20)	11 (20.7)		0 (0)
Osimertinib	8 (10)	8 (15.1)		0 (0)
Disease progression, no. (%)			0.564	
Progressed	52 (65)	37 (69.8)		39 (78)
Long-term no progress	28 (35)	16 (30.2)		11 (22)
Progress location, no. (%)			0.906	
Only intrapulmonary	30 (37.5)	23 (43.4)		25 (50)
Only extrapulmonary	15 (18.8)	10 (18.9)		9 (18)
Intrapulmonary and extrapulmonary	7 (8.7)	4 (7.5)		5 (10)
No progress	28 (35)	16 (30.2)		11 (22)
Actual PFS, no. (%)			0.173	
<9 months	16 (20)	7 (13.2)		21 (42)
9–12 months	8 (10)	11 (20.8)		7 (14)
≥ 12 months	56 (70)	35 (66)		22 (44)

Note: Since the actual progression time could not be obtained for patients with long-term progression-free status, the mean and range of PFS only included patients with recorded progression times.
Abbreviations: mth, month; PFS, progression-free survival.

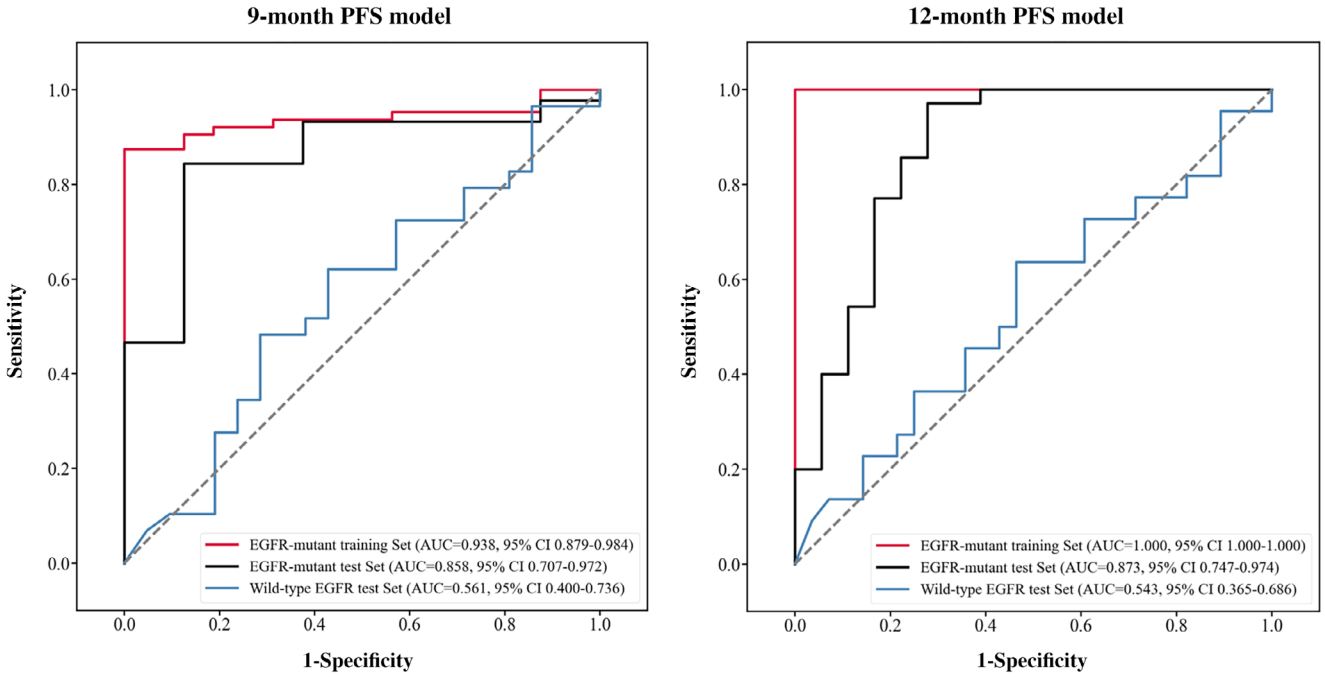

FIGURE 4 | ROC curve of disease progression prediction by 9-month and 12-month PFS prediction models in the EGFR-mutant training and test sets and the WT-EGFR test set.

TABLE 2 | Predicted possibility of disease progression for 9 and 12 months using PFS models based on 3D-UNet before and after fine-tuning.

	EGFR-mutant training set ($N=80$)		EGFR-mutant test set ($N=53$)	
	Basic model	**Fine-tuned model**	**Basic model**	**Fine-tuned model**
9-month PFS model				
Accuracy, % (95% CI)	70 (58.8–78.8)	88.8 (81.2–95)	71.7 (58.5–83)	81.1 (69.8–90.6)
Specificity, % (95% CI)	81.2 (60–100)	100 (100–100)	75.0 (40–100)	87.5 (62.5–100)
Sensitivity, % (95% CI)	67.2 (55–78)	85.9 (77.3–93.8)	71.1 (57.1–83.7)	80 (68.1–91.1)
AUC (95% CI)	0.801 (0.665–0.906)	0.938 (0.879–0.984)	0.742 (0.503–0.95)	0.858 (0.707–0.972)
F1 score (95% CI)	0.782 (0.686–0.857)	0.924 (0.872–0.968)	0.81 (0.706–0.899)	0.878 (0.80–0.944)
12-month PFS model				
Accuracy, % (95% CI)	66.2 (56.2–76.2)	100 (100–100)	60.4 (47.2–71.7)	84.9 (73.6–94.3)
Specificity, % (95% CI)	64.0 (44.4–81.8)	100 (100–100)	44.4 (22.2–70)	72.2 (47.6–91.7)
Sensitivity, % (95% CI)	67.3 (54.8–80)	100 (100–100)	68.6 (53.3–83.3)	91.4 (81.6–100)
AUC (95% CI)	0.701 (0.567–0.824)	1 (1–1)	0.524 (0.364–0.697)	0.873 (0.747–0.974)
F1 score (95% CI)	0.733 (0.632–0.828)	1 (1–1)	0.696 (0.567–0.811)	0.889 (0.806–0.951)

was 0.561 (95% CI, 0.4–0.736), with a sensitivity of 51.7% and a specificity of 61.9% (Figure 4). Similarly, the AUC for the 12-month progression prediction was 0.543 (95% CI, 0.365–0.686) with an accuracy of 54%, a specificity of 57.1%, a sensitivity of 50.0%, and an F1 score of 0.489 (Figure 4). Further, the confusion matrix analysis demonstrated relatively strong predictive performance for PFS <9 and <12 months, with true positive rates of 88% and 72%, respectively, and 80% and 0.91% for PFS ≥ 9 and ≥ 12 months, reflecting the good performance of PFS prediction models favorable for LUAD patients with EGFR-TKI treatment (Figure 5).

Given that our PFS model was developed based on the tumor dynamic changes during the EGFR-TKI treatment, we further validate the predicted PFS results with actual tumor changes captured in baseline and follow-up CT images. We categorized patients into four groups based on their PFS: PFS <9 months, 9 months $\leq$ PFS <12 months, PFS ≥ 12 months, and long-term no progression, and selected a representative case in each group (Figure 6). The lesion changes were demonstrated in the CT images across the four stages: baseline, first follow-up, and second follow-up, and progression or most recent if no progress was detected. In these cases, the results of 9-month and 12-month

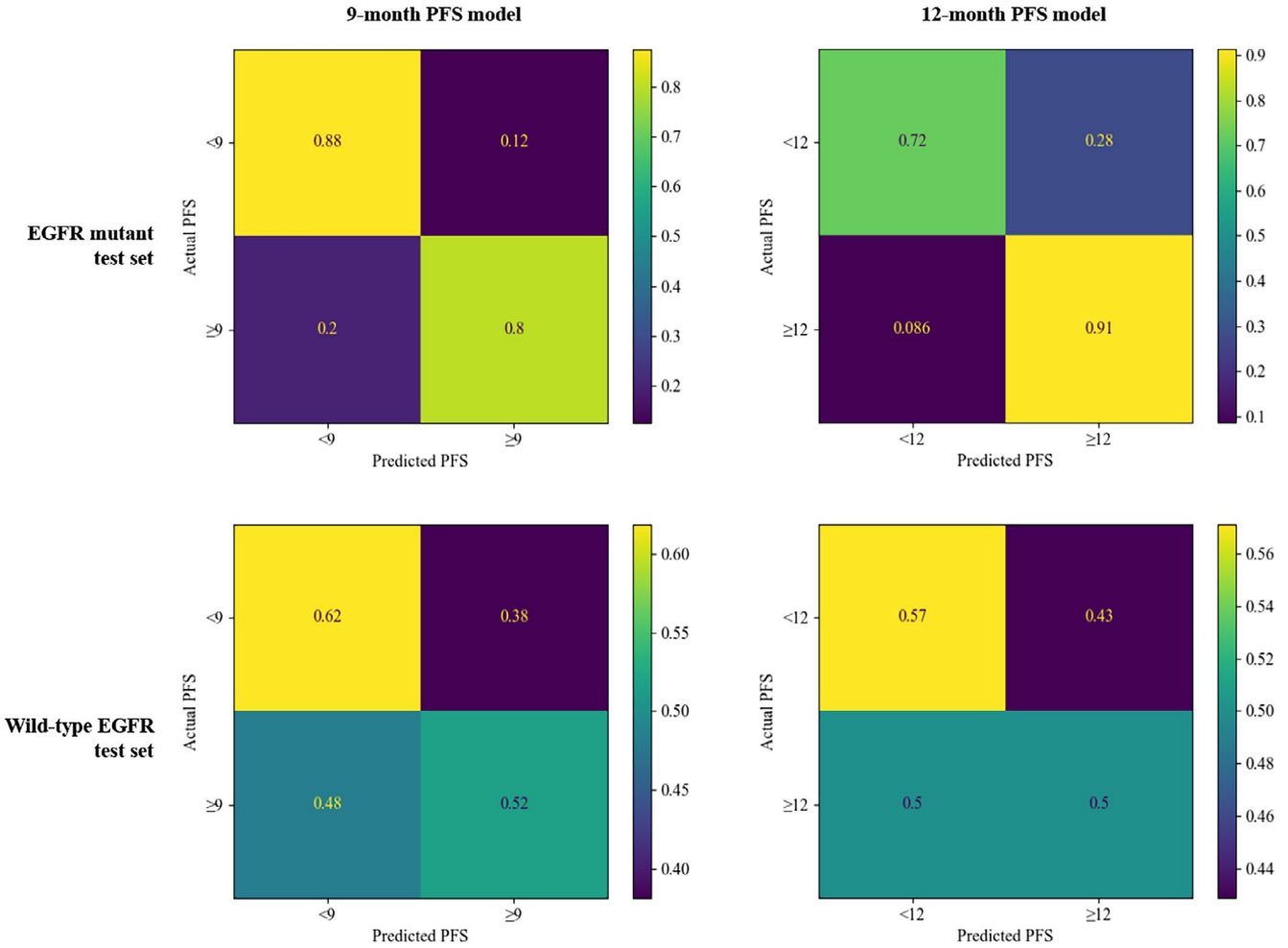

FIGURE 5 | Confusion matrix of disease progression prediction by 9-month and 12-month PFS prediction models in EGFR-mutant test set and WT-EGFR test set.

PFS prediction models were consistent with their actual PFS. In Case 2, our model correctly predicted the disease progression of the patient with brain metastases, although the target lesion on the CT scans did not show progression. The combination of two models for each patient provided a more accurate predictive result, which could be used for a better schedule of CT scans to monitor disease progression.

In our model, the location of the target lesion has been one of four key features used to construct the PFS prediction model. However, the disease progression of LUAD could occur extrathoracically, and lesion features of extrathoracic progression were not collected for model development. In the EGFR-mutant test set, 10 cases exhibited solely extrapulmonary disease progression. For the 9-month progression prediction, the AUC for these cases with extrapulmonary-only progression was 0.889 (95% CI, 0.653–1), but the AUC for 12-month PFS prediction was relatively low, 0.583 (95% CI, 0.125–1) (Table 3). Therefore, we applied the PFS models to each of the 10 cases to evaluate their accuracy for this specific progression type individually. Overall, both PFS models accurately predicted progression probabilities in 7 out of 10 cases. Among these cases, one patient had progression within 9 months, three patients between 9 and 12 months, and six patients beyond 12 months. The predicted and actual PFS values are summarized in Table 4.

3.4 | Comparison of Different PFS Prediction Models for EGFR-TKIs Response

We further compared the performance of our PFS model with those models reported in other studies [7, 8, 17]. As shown in Table 5, the AUC of our model for 9- and 12-month PFS predictions were 0.858 and 0.873, respectively. In comparison, the model by Zhu et al. achieved an AUC of 0.797 for a 10-month PFS prediction. The model by Lu et al. reported the AUC of 0.76, 0.77, 0.76, and 0.86 for 3-month, 12-month, 18-month, and 24-month PFS predictions, respectively. In addition, the model by Zhang et al. achieved AUC of 0.96, 0.76, and 0.80 for 6-month, 9-month, and 12-month progression predictions, respectively. These findings imply a consistent and better prediction of our models compared with other models.

4 | Discussion

EGFR-TKIs have been used as the first-line treatment for patients with EGFR-mutant NSCLC, whose effectiveness and resistance have been well demonstrated in the clinic. However, treatment response varies among patients, and disease progression is typically monitored according to routine schedules of CT scans. This approach can lead to delayed detection of rapid

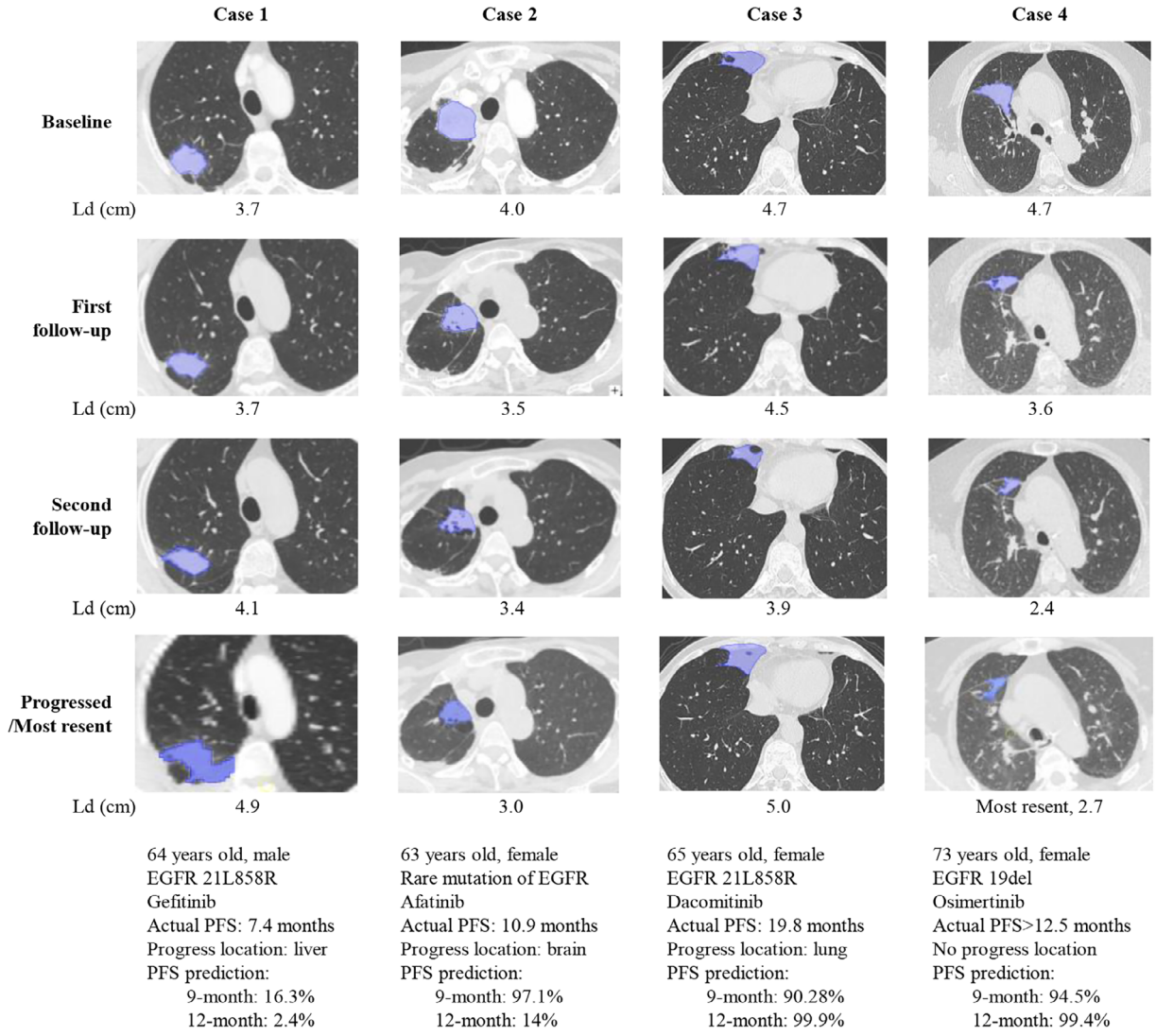

FIGURE 6 | Target lesions on the CT image of four typical cases of LUAD with EGFR mutation at the baseline, first follow-up, and second follow-up, and progress or most recent scans. The target lesion was detected and automatically marked with a blue color by the model. The LD of each lesion was reported by the model. Patient clinical characteristics, EGFR mutation types, EGFR-TKIs, actual PFS, and PFS model predictions were summarized for each case. Ld, long diameter.

TABLE 3 | Disease progression possibility predicted by 9-month and 12-month PFS models for patients with sole extrapulmonary progression.

	Accuracy	Specificity	Sensitivity	AUC	F1 score
Extrapulmonary ($N=10$)	**% (95% CI)**	**% (95% CI)**	**% (95% CI)**	**(95% CI)**	**(95% CI)**
9-month PFS model prediction	90 (70–100)	100 (100–100)	88.9 (65.3–100)	0.889 (0.653–1)	0.941 (0.79–1)
12-month PFS model prediction	80 (50–100)	50 (0–100)	100 (100–100)	0.583 (0.125–1)	0.857 (0.615–1)

development lesions and unnecessary CT exposure for patients without progression. Therefore, accurately predicting disease progression could allow for more timely adjustments tailored to individual cases. This study aimed to establish an AI model to predict PFS at 9 and 12 months and estimate the probability of no disease progression. They reached the accuracy of 81.1% (95% CI, 69.8%–90.6%) and 84.9% (95% CI, 73.6%–94.3%) in the EGFR-mutant test set, respectively. We independently trained both models. The predictions of the two models were uncorrelated. Both models demonstrated their great potential to assist clinicians in developing more appropriate treatment plans. In addition, given that GPU power is often a limiting factor for the application of AI models, we enhanced the compatibility of these AI-augmented models by dividing the segmented bounding box into multiple 3D patches, thereby significantly reducing GPU requirements.

Currently, prediction models of the disease progression after EGFR-TKI treatment are primarily based on lesion CT images at baseline, before treatment [7, 8]. Given that the efficacy progression times vary across studies, directly comparing AUC at the same progression times is not feasible. Therefore,

TABLE 4 | 9-month and 12-month PFS models predicting disease progression possibility for LUAD patients with solely extrapulmonary progression.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10
9-month PFS model probability, %	16.3	99.1	97.1	99.0	99.6	99.0	62.0	99.8	12.0	99.9
12-month PFS model probability, %	2.4	99.9	14	99.9	99	95.5	99.9	99.9	98.3	97.9
Actual PFS, mth	7.4	10.7	10.9	11.1	13.6	19.6	20.2	22.5	39.2	50.5

TABLE 5 | AUC comparison among PFS prediction models for patients with EGFR-TKI treatments.

	Our model (*N* = 133)	Zhu et al. [8] (*N* = 100)	Lu et al. [7] (*N* = 270)	Zhang et al. [17] (*N* = 172)
Age, yr				
Median	66	59	67.5	NA
Range	31–90	NA	60–75	NA
Sex, no. (%)				
Male	51 (38.3)	36 (36)	112 (41.5)	74 (43)
Female	82 (61.7)	64 (64)	158 (58.5)	98 (57)
Smoker, no. (%)	23 (17.3)	23 (23)	69 (25.6)	42 (24.4)
Number of CT scans, no.	3	1	1	2
Median PFS, mth	12.5	10	11.5	11.6
Model predicted PFS, mth	9, 12	10	3, 12, 18, 24	6, 9, 12
PFS model AUC (95% CI)	0.858 (0.707–0.972) 0.873 (0.747–0.974)	0.797 (0.657–0.937)	0.76, 0.77, 0.76, 0.86	0.96, 0.76, 0.80

we summarized recently reported PFS prediction models in Table 5. While these models predicted PFS at various times, the AUC (ROC curve) remains a reliable indicator of model robustness. Our models demonstrated strong performances at 9-month and 12-month PFS predictions, which could be partially due to the inputs of dynamic CT image information.

CT scans have been routinely scheduled during treatment to monitor the changes in diseases. However, this valuable information, which reflects treatment response, has often been overlooked in predicting disease progression. In this study, we developed the predictive model using chest CT images from patients both before and after treatment, including baseline and two follow-up CT scans taken within the subsequent 4–5 months (the first follow-up CT scan scheduled 6 weeks after the initial of the treatment with the second scan scheduled 6–12 weeks later). This time series of CT scans enabled us to track continuous lesion changes and gather an expanded set of lesion features. We also observed the consistency of the changes in the tumor volume and the accuracy of the PFS prediction. As shown in Figure 6, Case 1, with disease progression within 7.4 months, had a longer diameter of the target lesion (3.7–4.9 cm) during treatment, whose predicted PFS for 9 and 12 months were only 16.3% and 2.4%, respectively. Case 2, with 10.9 months of actual PFS, had the predicted probability of no disease progression within 9 or 12 months of 97.1% and 14%, respectively. These findings suggest the utilization of both models might provide a more accurate prediction, thereby leading to an optimized follow-up schedule for CT scans.

Model testing results demonstrated that our model was well-suited for EGFR-TKI-treated patients. We further assessed their performance with a subtype of extrapulmonary progression. In the EGFR-TKI treatment test, 3 out of 10 patients with only extrapulmonary progression were not correctly reported (Table 3). The actual PFS for Case 9 was 39.2 months, while the predicted probability of progression within 9 months was only 12%, but it increased to 98.3% for 12 months. Similarly, Cases 2 and 4 had incorrectly predicted probabilities of progression at 12 months; their 12-month PFS predictions were both above 90%, while the actual PFS was 10.7 months for Case 2 and 11.1 months for Case 4. We did not identify any obvious differences in the data characteristics of these cases compared to those of other cases. We suspected that intrapulmonary progression might have occurred later than 12 months in Cases 2 and 4. According to our HIS, intrapulmonary progression in Case 2 was detected approximately 7 months later, while no intrapulmonary progression was reported for Case 4. Although the intrapulmonary progression timeline for Case 2 aligns with our PFS model prediction, the accuracy of predicting 7 out of 10 extrapulmonary progressions further supports the utility of our models in these patients. Further investigation of the difference between actual

and predicted PFS in extrapulmonary progression might provide valuable information for patient care. In addition, the extrapulmonary progression prediction in our model performed comparably for 9-month PFS (0.889 vs. 0.841) but less effectively for 12 months (0.583 vs. 0.953). This discrepancy may be due to the limited number of cases used for model training. Nonetheless, this model shows predictive potential for EGFR-TKI-treated patients with extrapulmonary progression, which could be further investigated with more cases.

This study had some limitations. Although our findings indicate that models effectively predicted disease progression in patients treated with EGFR-TKIs, the feature dimensionality involved in these models was restricted by the available training data. With an increase in sample size and feature dimensionality, models' performance could be further improved. Moreover, more cases of extrapulmonary progression are needed to enhance the current model. Third, the current model was not designed to independently analyze first-generation, second-generation, and third-generation EGFR-TKIs. Future improvements to the model could potentially yield regression models for predicting therapeutic progression for different generations of EGFR-TKIs, providing a more accurate prediction, and guiding the use of EGFR-TKIs to benefit patients. This study focused on EGFR-mutant patients, but a significant number of WT-EGFR patients face the same dilemma of CT scan scheduling and treatment response assessment. With a sufficiently large dataset, we could separately develop models for WT-EGFR cases.

5 | Conclusions

This study utilized deep learning techniques to automatically segment lesions, extract lesion feature information, including shape, size, volume, and density features, and track and compare target lesions. By integrating radiomics features and clinical information, a deep learning efficacy prediction model was constructed for patients treated with EGFR-TKIs. Models demonstrated good sensitivity and specificity, accurately predicting disease progression. Its potential to reduce CT scan frequency and medical costs underscores its value in providing efficient, personalized medical care for cancer patients.

Author Contributions

Yan'e Liu: data curation, data analysis, investigation, establishing model, writing and editing the manuscript, funding acquisition. **Xiangfeng Luo:** data curation, establishing model, writing and editing the manuscript. **Lu Yang:** data curation, investigation, writing the original draft. **Xueliang Cheng:** data curation, establishing model, writing and editing the manuscript. **Xin Zhu:** data analysis, writing and editing the manuscript. **Hua Zhang:** statistical analysis, editing the manuscript. **Bolin Hou:** conceptualization, editing manuscript. **Baoshan Cao:** conceptualization, investigation, reviewing, and editing the manuscript, funding acquisition.

Acknowledgments

This study was supported by the National Key R&D Program of China (2020AAA0105203) and the Innovation & Transfer Fund of Peking University Third Hospital (BYSYCY2024088).

Ethics Statement

This study was approved by the Peking University Third Hospital Medical Science Research Ethics Committee (Ethics approval number: M2023746).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data analyzed in this study are available from the corresponding author upon reasonable request.

References

1. A. A. Thai, B. J. Solomon, L. V. Sequist, J. F. Gainor, and R. S. Heist, "Lung Cancer," *Lancet* 398 (2021): 535–554, https://doi.org/10.1016/S0140-6736(21)00312-3.

2. Y. Li, X. Wu, P. Yang, G. Jiang, and Y. Luo, "Machine Learning for Lung Cancer Diagnosis, Treatment, and Prognosis," *Genomics, Proteomics & Bioinformatics* 20 (2022): 850–866, https://doi.org/10.1016/j.gpb.2022.11.003.

3. Y. Shi, J. S. Au, S. Thongprasert, et al., "A Prospective, Molecular Epidemiology Study of EGFR Mutations in Asian Patients With Advanced Non-Small-Cell Lung Cancer of Adenocarcinoma Histology (PIONEER)," *Journal of Thoracic Oncology* 9 (2014): 154–162, https://doi.org/10.1097/JTO.0000000000000033.

4. P. G. Mikhael, J. Wohlwend, A. Yala, et al., "Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk From a Single Low-Dose Chest Computed Tomography," *Journal of Clinical Oncology* 41 (2023): 2191–2200, https://doi.org/10.1200/JCO.22.01345.

5. M. K. Gould, B. Z. Huang, M. C. Tammemagi, Y. Kinar, and R. Shiff, "Machine Learning for Early Lung Cancer Identification Using Routine Clinical and Laboratory Data," *American Journal of Respiratory and Critical Care Medicine* 204 (2021): 445–453, https://doi.org/10.1164/rccm.202007-2791OC.

6. L. Yu, G. Tao, L. Zhu, et al., "Prediction of Pathologic Stage in Non-Small Cell Lung Cancer Using Machine Learning Algorithm Based on CT Image Feature Analysis," *BioMed Research International* 19 (2019): 464, https://doi.org/10.1186/s12885-019-5646-9.

7. C. F. Lu, C. Y. Liao, H. S. Chao, et al., "A Radiomics-Based Deep Learning Approach to Predict Progression Free-Survival After Tyrosine Kinase Inhibitor Therapy in Non-Small Cell Lung Cancer," *Cancer Imaging* 23 (2023): 9, https://doi.org/10.1186/s40644-023-00522-5.

8. J. M. Zhu, L. Sun, L. Wang, et al., "Radiomics Combined With Clinical Characteristics Predicted the Progression-Free Survival Time in First-Line Targeted Therapy for Advanced Non-Small Cell Lung Cancer With EGFR Mutation," *BMC Research Notes* 15 (2022): 140, https://doi.org/10.1186/s13104-022-06019-x.

9. M. Jiang, P. Yang, J. Li, et al., "Computed Tomography-Based Radiomics Quantification Predicts Epidermal Growth Factor Receptor Mutation Status and Efficacy of First-Line Targeted Therapy in Lung Adenocarcinoma," *Frontiers in Oncology* 12 (2022): 985284, https://doi.org/10.3389/fonc.2022.985284.

10. M. Ravanelli, G. M. Agazzi, B. Ganeshan, et al., "CT Texture Analysis as Predictive Factor in Metastatic Lung Adenocarcinoma Treated With Tyrosine Kinase Inhibitors (TKIs)," *European Journal of Radiology* 109 (2018): 130–135, https://doi.org/10.1016/j.ejrad.2018.10.016.

11. X. Tang, Y. Li, W. F. Yan, et al., "Machine Learning-Based CT Radiomics Analysis for Prognostic Prediction in Metastatic Non-Small Cell Lung Cancer Patients With EGFR-T790M Mutation Receiving

Third-Generation EGFR-TKI Osimertinib Treatment," *Frontiers in Oncology* 11 (2021): 719919, https://doi.org/10.3389/fonc.2021.719919.

12. J. Song, J. Shi, D. Dong, et al., "A New Approach to Predict Progression-Free Survival in Stage IV EGFR-Mutant NSCLC Patients With EGFR-TKI Therapy," *Clinical Cancer Research* 24 (2018): 3583–3592, https://doi.org/10.1158/1078-0432.CCR-17-2507.

13. K. Deng, L. Wang, Y. Liu, et al., "A Deep Learning-Based System for Survival Benefit Prediction of Tyrosine Kinase Inhibitors and Immune Checkpoint Inhibitors in Stage IV Non-Small Cell Lung Cancer Patients: A Multicenter, Prognostic Study," *EClinicalMedicine* 51 (2022): 101541, https://doi.org/10.1016/j.eclinm.2022.101541.

14. J. Song, L. Wang, N. N. Ng, et al., "Development and Validation of a Machine Learning Model to Explore Tyrosine Kinase Inhibitor Response in Patients With Stage IV EGFR Variant-Positive Non-Small Cell Lung Cancer," *JAMA Network Open* 3 (2020): e2030442, https://doi.org/10.1001/jamanetworkopen.2020.30442.

15. S. Wang, H. Yu, Y. Gan, et al., "Mining Whole-Lung Information by Artificial Intelligence for Predicting EGFR Genotype and Targeted Therapy Response in Lung Cancer: A Multicohort Study," *Lancet Digital Health* 4 (2022): e309–e319, https://doi.org/10.1016/S2589-7500(22)00024-3.

16. R. Hou, X. Li, J. Xiong, et al., "Predicting Tyrosine Kinase Inhibitor Treatment Response in Stage IV Lung Adenocarcinoma Patients With EGFR Mutation Using Model-Based Deep Transfer Learning," *Frontiers in Oncology* 11 (2021): 679764, https://doi.org/10.3389/fonc.2021.679764.

17. X. Zhang, B. Lu, X. Yang, et al., "Prognostic Analysis and Risk Stratification of Lung Adenocarcinoma Undergoing EGFR-TKI Therapy With Time-Serial CT-Based Radiomics Signature," *European Radiology* 33 (2023): 825–835, https://doi.org/10.1007/s00330-022-09123-5.

18. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection With Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2017): 1137–1149, https://doi.org/10.1109/TPAMI.2016.2577031.

19. E. D. Cubuk, B. Zoph, D. Mané, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning Augmentation Strategies From Data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), 113–123.

20. P. Lambin, E. Rios-Velazquez, R. Leijenaar, et al., "Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis," *European Journal of Cancer* 48 (2012): 441–446, https://doi.org/10.1016/j.ejca.2011.11.036.

21. Q. Zhang, J. Li, X. Nan, and X. Zhang, "Detail-Sensitive 3D-UNet for Pulmonary Airway Segmentation From CT Images," *Medical & Biological Engineering & Computing* 62 (2024): 3749–3762, https://doi.org/10.1007/s11517-024-03169-x.

22. X. Li, S. Zhang, X. Luo, et al., "Accuracy and Efficiency of an Artificial Intelligence-Based Pulmonary Broncho-Vascular Three-Dimensional Reconstruction System Supporting Thoracic Surgery: Retrospective and Prospective Validation Study," *EBioMedicine* 87 (2023): 104422, https://doi.org/10.1016/j.ebiom.2022.104422.

23. H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, et al., "Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach," *Nature Communications* 5 (2014): 4006, https://doi.org/10.1038/ncomms5006.

24. Y. Qin, H. Zheng, Y. Gu, et al., "Learning Tubule-Sensitive CNNs for Pulmonary Airway and Artery-Vein Segmentation in CT," *IEEE Transactions on Medical Imaging* 40 (2021): 1603–1617, https://doi.org/10.1109/TMI.2021.3062280.

25. A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier Nonlinearities Improve Neural Network Acoustic Models," in *Proceedings of the 30th International Conference on Machine Learning* (2013), 3.

26. W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-Margin Softmax Loss for Convolutional Neural Networks," arXiv preprint arXiv:1612.02295 (2016).

27. D. P. Kingma, "Adam: A Method for Stochastic Optimization," arXiv Preprint arXiv:1412.6980 (2014).

28. C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal* 27 (1948): 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

29. A. Paszke, S. Gross, F. Massa, et al., "Pytorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems 32*, eds. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'AlchÃ-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019), 8026–8037.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.