

# UC Irvine

## UC Irvine Previously Published Works

### Title

Prediction of epidermal growth factor receptor (EGFR) mutation status in lung adenocarcinoma patients on computed tomography (CT) images using 3-dimensional (3D) convolutional neural network.

### Permalink

<https://escholarship.org/uc/item/5zs5d390>

### Journal

Quantitative Imaging in Medicine and Surgery, 14(8)

### ISSN

2223-4292

### Authors

Zhang, Guojin

Shang, Lan

Cao, Yuntai

et al.

### Publication Date

2024-08-01

### DOI

10.21037/qims-24-33

Peer reviewed



# Prediction of epidermal growth factor receptor (*EGFR*) mutation status in lung adenocarcinoma patients on computed tomography (CT) images using 3-dimensional (3D) convolutional neural network

Guojin Zhang<sup>1#^</sup>, Lan Shang<sup>1#</sup>, Yuntai Cao<sup>2#</sup>, Jing Zhang<sup>3#</sup>, Shenglin Li<sup>1,4</sup>, Rong Qian<sup>1</sup>, Huan Liu<sup>5</sup>, Zhuoli Zhang<sup>6</sup>, Hong Pu<sup>1</sup>, Qiong Man<sup>7</sup>, Weifang Kong<sup>1</sup>

<sup>1</sup>Department of Radiology, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, China; <sup>2</sup>Department of Radiology, Affiliated Hospital of Qinghai University, Xining, China; <sup>3</sup>Department of Radiology, Fifth Affiliated Hospital of Zunyi Medical University, Zhuhai, China; <sup>4</sup>Department of Radiology, Lanzhou University Second Hospital, Lanzhou, China; <sup>5</sup>Department of Pharmaceuticals Diagnosis, GE Healthcare, Beijing, China; <sup>6</sup>Department of Radiology, University of California Irvine, Irvine, CA, USA; <sup>7</sup>School of Pharmacy, Chengdu Medical College, Chengdu, China

**Contributions:** (I) Conception and design: G Zhang, Q Man, W Kong; (II) Administrative support: G Zhang, W Kong; (III) Provision of study materials or patients: G Zhang, L Shang, S Li; (IV) Collection and assembly of data: G Zhang, Y Cao, J Zhang, R Qian, H Pu; (V) Data analysis and interpretation: G Zhang, Q Man, H Liu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work.

**Correspondence to:** Weifang Kong, MS. Department of Radiology, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, No. 32, West Second Section, First Ring Road, Qingyang District, Chengdu 610072, China. Email: kongweifang@med.uestc.edu.cn; Qiong Man, MD. School of Pharmacy, Chengdu Medical College, No. 783 Xindu Avenue, Xindu District, Chengdu 610500, China. Email: gretam@163.com.

**Background:** Noninvasively detecting epidermal growth factor receptor (*EGFR*) mutation status in lung adenocarcinoma patients before targeted therapy remains a challenge. This study aimed to develop a 3-dimensional (3D) convolutional neural network (CNN)-based deep learning model to predict *EGFR* mutation status using computed tomography (CT) images.

**Methods:** We retrospectively collected 660 patients from 2 large medical centers. The patients were divided into training (n=528) and external test (n=132) sets according to hospital source. The CNN model was trained in a supervised end-to-end manner, and its performance was evaluated using an external test set. To compare the performance of the CNN model, we constructed 1 clinical and 3 radiomics models. Furthermore, we constructed a comprehensive model combining the highest-performing radiomics and CNN models. The receiver operating characteristic (ROC) curves were used as primary measures of performance for each model. Delong test was used to compare performance differences between different models.

**Results:** Compared with the clinical [training set, area under the curve (AUC) =69.6%, 95% confidence interval (CI), 0.661–0.732; test set, AUC =68.4%, 95% CI, 0.609–0.752] and the highest-performing radiomics models (training set, AUC =84.3%, 95% CI, 0.812–0.873; test set, AUC =72.4%, 95% CI, 0.653–0.794) models, the CNN model (training set, AUC =94.3%, 95% CI, 0.920–0.961; test set, AUC =94.7%, 95% CI, 0.894–0.978) had significantly better predictive performance for predicting *EGFR* mutation status. In addition, compared with the comprehensive model (training set, AUC =95.7%, 95% CI, 0.942–0.971; test set, AUC =87.4%, 95% CI, 0.820–0.924), the CNN model had better stability.

<sup>^</sup> ORCID: 0000-0002-4312-5212.

**Conclusions:** The CNN model has excellent performance in non-invasively predicting *EGFR* mutation status in patients with lung adenocarcinoma and is expected to become an auxiliary tool for clinicians.

**Keywords:** Deep learning; convolutional neural network (CNN); lung adenocarcinoma; epidermal growth factor receptor (*EGFR*); computed tomography (CT)

Submitted Jan 07, 2024. Accepted for publication Jun 28, 2024. Published online Jul 30, 2024.

doi: 10.21037/qims-24-33

View this article at: <https://dx.doi.org/10.21037/qims-24-33>

## Introduction

Lung cancer is the second most common cancer and the leading cause of cancer death worldwide (1). Non-small cell lung cancer (NSCLC) accounts for 80–85% of all lung cancer cases, among which lung adenocarcinoma is the most common histological type (2). With the continuous in-depth exploration of the field of molecular pathology, a series of oncogenic driver genes have been discovered in patients with NSCLC, transitioning the treatment of lung cancer from traditional chemotherapy to targeted therapy for specific molecules (3). In particular, the introduction of tyrosine kinase inhibitors (TKI) for epidermal growth factor receptor (*EGFR*) mutations has made targeted therapy possible. Compared with patients receiving standard chemotherapy, patients receiving *EGFR* TKI have a better objective remission rate, markedly longer progression-free survival, and experience effectively lower toxic effects (4,5). Therefore, determining the *EGFR* mutation status before treatment is a prerequisite for receiving *EGFR* TKI.

Currently, the detection of *EGFR* mutation status mainly relies on biopsy or post-operative tissue specimens; however, these methods have some limitations. First, NSCLC is a heterogeneous disease, and a small portion of tissue obtained from tissue specimens does not reflect intra-tumor or inter-tumor heterogeneity (6). Second, needle biopsy increases the risk of potential cancer metastasis (7). Finally, this is an invasive process that cannot be tolerated by some elderly or frail patients (8). In this case, a non-invasive method to predict *EGFR* mutation status is necessary to supplement the shortcomings of tissue sample analysis.

Computed tomography (CT) is the preferred imaging method for lung cancer screening, diagnosis, and prognosis assessment (2). Recently, some studies have used CT radiomics to predict *EGFR* mutation status and achieved varying degrees of success (6,9-11). However, the extraction

of radiomics features requires the accurate delineation of the tumor area of interest, which is extremely time-consuming and costly (12). In addition, radiomics features may be affected by different scanning equipment and parameters, resulting in poor repeatability of feature extraction (13). In clinical practice, CT imaging is suitable for deep learning because similar large data sets are available (14). Deep learning technology based on convolutional neural network (CNN) has recently attracted wide attention from researchers in the medical field, particularly in medical image analysis (15-17). Some studies have used deep learning to predict *EGFR* mutation status and have shown strong promise (8,18-20). For example, Wang *et al.* (19) used a deep learning model to predict *EGFR* mutation status, and the areas under the curve (AUCs) in the training and validation sets were 0.85 and 0.81, respectively. Zhao *et al.* (18) constructed a DenseNets model using CT images of 579 patients to predict *EGFR* mutation status, and the AUC in the test set was 0.75. Although the predictive performance of these research results needs to be further improved, these initial successes have raised expectations for the implementation of high-performance artificial intelligence in daily clinical practice.

This study aimed to establish a fully automated deep learning model to evaluate whether a CT images-based CNN model can improve the performance of predicting *EGFR* mutation status. We present this article in accordance with the TRIPOD reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-33/rc>).

## Methods

### Study design

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This retrospective study was approved by the Institutional Review Boards of Sichuan Provincial People's Hospital (Chengdu, China) (No. 2022-254) and Lanzhou University Second Hospital (Lanzhou,

China) (No. 2020A-180). The requirement for written informed consent was waived because the data were analyzed retrospectively and anonymously. Patients were included in this study if they (I) had a histological type of lung adenocarcinoma according to the 2021 World Health Organization classification of lung tumors, (II) underwent a thin-slice CT scan before biopsy or surgical treatment, (III) had no lung cancer-related treatment before the CT scan, and (IV) had complete clinical data, including sex, age, smoking history, and carcinoembryonic antigen (CEA). Patients were excluded in this study if they (I) had poor image quality due to severe motion artifacts, foreign body artifacts *in vitro*, or other technical deficiencies, (II) the interval between CT examination and biopsy or surgery was greater than 14 days, and (III) were less than 18 years old. All CT image data were obtained from the picture archiving and communication system, and clinical data were obtained from medical records.

In total, 528 patients (including 260 *EGFR* mutant and 268 *EGFR* wild-type patients) from Sichuan Provincial People's Hospital from January 2018 to December 2020 were included in the training set, and 132 patients (including 65 *EGFR* mutant and 67 *EGFR* wild-type patients) from Lanzhou University Second Hospital from January 2019 to March 2020 were included in the external test set. The training and external test sets were used to develop and validate the CNN model, respectively.

### *CT image acquisition*

Chest CT scans were performed by using one of three spiral CT systems (Philips iCT 256, Philips Medical Systems, Best, the Netherlands; Discovery CT750 HD, GE Healthcare, Milwaukee, WI, USA; Somatom Sensation 64, Siemens, Erlangen, Germany). (I) For 64-detector scanner, tube voltage, 120 kVp; tube current, 375 mA. (II) for the other 2 scanners, tube voltage, 120 kVp; tube current, 150–200 mA. For all scanners, reconstruction thicknesses and intervals were both 1.25 mm, and scan range, from the tip of the lung to the bottom of the lung.

### *Detection of *EGFR* mutation status*

Drug target-associated mutations on *EGFR* exons 18, 19, 20, and 21 were detected by pathologists on histological specimens. *EGFR* mutation status was detected using a polymerase chain reaction (PCR)-based amplification refractory mutation system (ARMS) with human *EGFR*

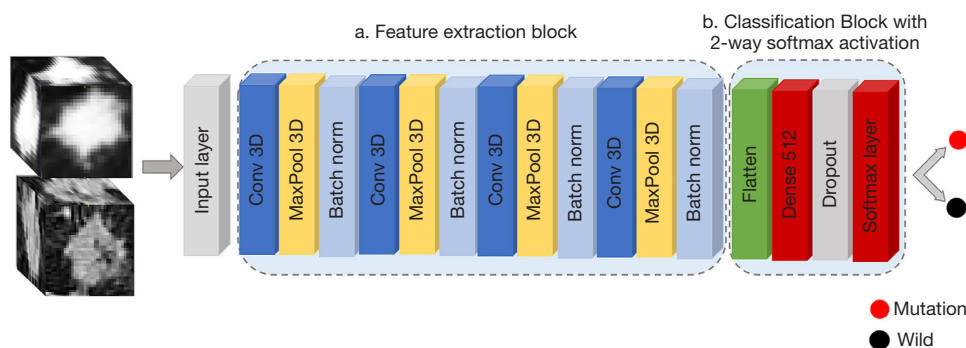
gene detection kit (Beijing SinoMD Gene Detection Technology Co., Ltd., Beijing, China; Amoy Diagnostics, Xiamen, China). If any mutation in exons 18 to 21 was detected, the tumor was identified as *EGFR*-mutant; otherwise, the tumor was identified as *EGFR*-wild type.

### *Data preprocessing*

The spacing of CT images was first resampled to  $1 \times 1 \times 1 \text{ mm}^3$  by third-order spline interpolation to avoid the image distortion. Then, the 3-dimensional (3D) target lesions were manually segmented slice by slice by the radiologists using the medical image processing software ITK-SNAP 3.8.0 (<https://www.itksnap.org>), and subsequently confirmed by an experienced physician. After the segmentation, the volumes of interest (VOIs) were exported in NII format for further analysis. Region of interest (ROI) segmentation is described in [Appendix 1](#) (Methods). Before feeding into the models, the VOIs were normalized according to the following methods: (I) the CT VOIs were automatically cropped to show only the lesion of interest using code written in the programming language Python 3.8.0 (Python Software Foundation; <https://www.python.org/>), (II) we then rotated the volumes by  $90^\circ$  to fix the orientation, (III) to minimize bias field effects, cropped images, with a threshold between  $-1,000$  and  $400$  Hounsfield units, were commonly used to normalize CT VOIs to be between 0 and 1, and (IV) we resized width, height and depth to  $64 \times 64 \times 32$ .

### *CNN architecture*

A 17-layer 3D CNN comprised 4 3D convolutional (Conv) layers with layers consisting of 32, 64, 128, and 256 filters all with a kernel size of  $3 \times 3 \times 3$ . Each Conv layer was followed by a max-pooling (MAXPOOL) layer with a stride of 2 and rectified linear unit (ReLU) activation, which ends with a batch normalization (BN) layer. Essentially, our feature extraction block consists of 4 Conv-MAXPOOL-BN modules. The final output from the feature extraction block was flattened and passed through a fully connected layer with 512 neurons. The output was then carried to a dense layer of 2 neurons with softmax activation for the binary classification problem. The network architecture is shown in [Figure 1](#). We considered keeping the network relatively simple to avoid overparameterization problems with only 1,297,090 learnable parameters. This was also motivated by the fewer number of training samples and memory challenges associated with



**Figure 1** 17-layer 3D CNN architecture. Conv, convolutional; 3D, 3-dimensional; CNN, convolutional neural network.

it. All training was conducted on the GeForce GTX 1060 (NVIDIA, Santa Clara, CA, USA) graphics processing unit and completed in 20 hours of wall-clock time using 1 GPU. The model was built using Python 3.8.0 and Keras 2.2 (<https://keras.io/>) running on a Tensorflow backend (Google, <https://www.tensorflow.org/>). The architecture of the 3D CNN used in this example was based on a previous study by Zunair *et al.* (21).

#### *Performance comparison of CNN model with radiomics and clinical models*

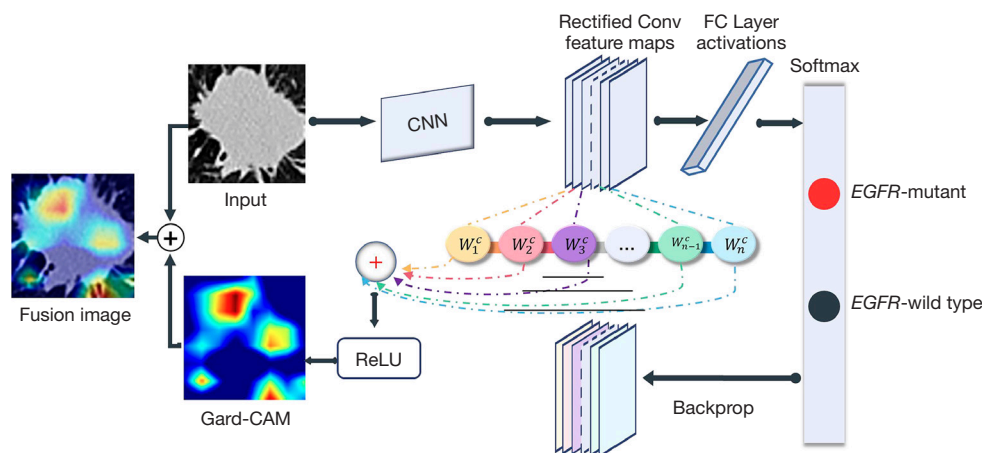
In our previous study, clinical characteristics (7) and radiomics features (9) were used to predict *EGFR* mutation status. Therefore, we built a clinical model and 3 radiomics models to compare with the proposed CNN model. The clinical model was established using logistic regression (LR), which was characterized by sex and smoking history. The radiomics model automatically extracted 1,727 radiomics features from the 3D VOIs of each patient using the Python 3.8.0 open-source software package PyRadiomics 3.0.1. Details of radiomics features are included in the [Appendix 1](#). The Mann-Whitney U test, Spearman correlation analysis, least absolute shrinkage and selection operator (LASSO) regression and univariate LR were used to reduce the dimensionality of the radiomics features, and 13 radiomics features were screened ([Figure S1](#)). Subsequently, 3 radiomics models were constructed to predict the *EGFR* mutation status using LR, support vector machine (SVM), and naïve Bayes (Bayes) classifiers, respectively. These classifiers have been shown to perform well in radiomics analysis (9,22). Finally, we used the radiomics model with the highest diagnostic performance combined with the CNN model to build a comprehensive model both in training and test sets.

#### *Visualization of the deep learning model*

Deep learning is an end-to-end process, and its reasoning process cannot be intuitively understood. To further understand the reasoning process of the CNN model, we used the Gradient-weighted Class Activation Mapping (Grad-CAM) technique to visualize the features learned by the CNN model. Grad-CAM measures the importance of each pixel to the prediction result by calculating the gradient of the model output (prediction category) relative to the middle Conv layer. Through Grad-CAM visualization, we can know which regions in the image the model focuses on during prediction are important for making the prediction, which helps us understand the black box of the CNN model (23). The architecture explaining the Grad-CAM technology is shown in [Figure 2](#). The weight calculation formula for Grad-CAM is provided in the [Appendix 1](#).

#### *Statistical analysis*

All statistical analyses were performed using IBM SPSS Statistics for Windows 23.0 (IBM Corp., Armonk, NY, USA). The chi-square or Fisher's exact tests were used to evaluate the differences in categorical data between the *EGFR* mutant and *EGFR* wild-type groups and is expressed as a percentage. The independent sample *t*-test or Mann-Whitney U-test were used to assess differences in continuous data and is expressed as mean  $\pm$  standard deviation (SD). A P value  $<0.05$  was considered statistically significant. AUC, sensitivity, specificity, and accuracy were used as primary measures of performance for each model. The DeLong test was used to evaluate the differences in AUC values between the various models.



**Figure 2** Grad-CAM architecture example diagram. Grad-CAM, Gradient-weighted Class Activation Mapping; CNN, convolutional neural network; Conv, convolutional; FC, fully connected; EGFR, epidermal growth factor receptor; ReLU, rectified linear unit.

## Results

### Clinical characteristics

A total of 660 patients from 2 hospitals were included in this study. The mean age ( $\pm$  SD) for the entire dataset was  $57.64 \pm 9.30$  years, 52.27% (345/660) of patients were male, and 63.79% (421/660) of the patients were non-smokers.

There were no significant differences in age ( $P=0.448$ ), CEA ( $P=0.341$ ), tumor location ( $P=0.330$ ), and stage ( $P=0.234$ ) between the EGFR mutant and wild-type groups. Sex ( $P<0.001$ ) and smoking history ( $P<0.001$ ) were statistically different between the 2 groups; therefore, these characteristics were used to establish a clinical model. The clinical characteristics of the patients are summarized in Table 1.

### Diagnostic performance of the CNN model

The prediction performance of the CNN model is shown in Table 2 and Figure 3. In the training set, the CNN model showed good prediction performance (AUC and accuracy were 94.3% and 93.4%, respectively), which was confirmed in further independent test sets (AUC and accuracy were 94.7% and 93.8%, respectively). Figure 4 shows the loss and accuracy curves of the CNN model for predicting EGFR mutation status.

Figure 5 shows the decision curve of the CNN model. This curve indicated that if the threshold probability was between 0.14 and 0.97, the CNN model had more benefits than other single models in predicting EGFR mutation status.

### Model comparison

Although clinical [training set: AUC, 0.696; 95% confidence interval (CI), 0.661–0.732; test set: AUC, 0.684; 95% CI, 0.609–0.752] and radiomics [SVM (training set: AUC, 0.843; 95% CI, 0.812–0.873), (test set: AUC, 0.724; 95% CI, 0.653–0.794)] models could predict EGFR mutation status, their performance was inferior to that of the CNN model (Figure 3, Table 2). Decision curves confirmed this finding. Overall, using a CNN model for decision-making is a more robust approach than using clinical or radiomics models (Figure 5). Delong's test showed statistically significant AUC values between the CNN model and the clinical model and the 3 radiomics models in both the training and test sets (all  $P<0.05$ ; Figure S2). Although the comprehensive model performed better than the CNN model in the training set (AUC, 0.957; 95% CI, 0.942–0.971), there was no statistical difference between the 2 models (Figure S2). However, the comprehensive model performed poorly on the test set (AUC, 0.874; 95% CI, 0.820–0.924) (Figure 3, Table 2).

### Discover suspicious area with Grad-CAM

Grad-CAM visually interprets the area (suspicious area) that the CNN model focuses on when making predictions (Figure 6). For each tumor, the CNN model generated an attention map, and different colors represented the importance of the CNN model's attention. The dark red area is the area that attracts the highest attention of the CNN model, and it also represents the suspicious area found by the CNN model. In the bottom row of Figure 6A,

**Table 1** Demographics and clinical characteristics of patients

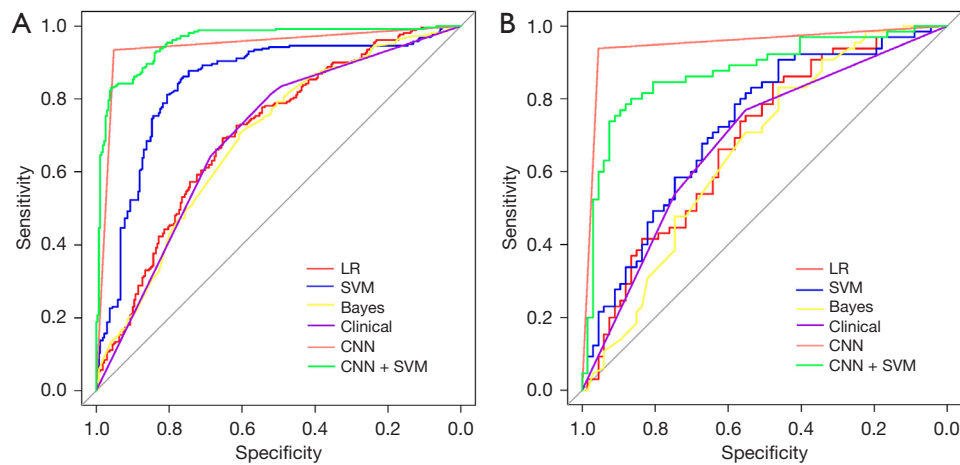
Characteristics	Total (n=660)	EGFR wild-type (n =335)	EGFR mutant (n =325)	P value
Age (years) (mean ± SD)	57.64±9.30	57.91±9.39	57.36±9.22	0.448
Sex, n (%)				<0.001
Male	345 (52.27)	227 (67.76)	118 (36.31)	
Female	315 (47.73)	108 (32.24)	207 (63.69)	
Smoking history, n (%)*				<0.001
Smoker	239 (36.21)	176 (52.54)	63 (19.38)	
Non-smoker	421 (63.79)	159 (47.46)	262 (80.62)	
CEA, n (%)				0.341
Normal	262 (39.70)	127 (37.91)	135 (41.54)	
High	398 (60.30)	208 (62.09)	190 (58.46)	
Lobe location, n (%)				0.330
Right upper	226 (34.24)	113 (33.73)	113 (34.77)	
Right middle	41 (6.21)	15 (4.48)	26 (8.00)	
Right lower	150 (22.73)	83 (24.78)	67 (20.62)	
Left upper	147 (22.27)	75 (22.39)	72 (22.15)	
Left lower	96 (14.55)	49 (14.63)	47 (14.46)	
Stage, n (%)				0.234
I	337 (51.06)	176 (52.54)	161 (49.54)	
II	91 (13.79)	49 (14.63)	42 (12.92)	
III	99 (15.00)	53 (15.82)	46 (14.15)	
IV	133 (20.15)	57 (17.01)	76 (23.38)	

\*, smoking history is defined as follows: smoker, former and current smokers; non-smoker, never smoked. *EGFR*, epidermal growth factor receptor; SD, standard deviation; CEA, carcinoembryonic antigen.

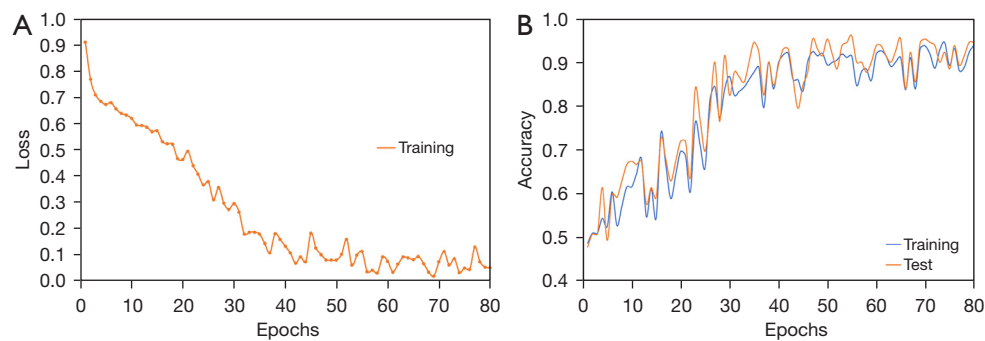
**Table 2** Predictive performance of clinical, radiomics, and deep learning models

Models	Training set (n=528)				Test set (n=132)			
	AUC (95% CI)	Sen	Spe	Acc	AUC (95% CI)	Sen	Spe	Acc
LR	0.706 (0.669–0.744)	0.688	0.653	0.670	0.685 (0.608–0.755)	0.708	0.567	0.636
SVM	0.843 (0.812–0.873)	0.854	0.769	0.811	0.724 (0.653–0.794)	0.785	0.567	0.674
Bayes	0.690 (0.652–0.727)	0.769	0.519	0.642	0.658 (0.583–0.733)	0.754	0.478	0.614
Clinical	0.696 (0.661–0.732)	0.642	0.687	0.665	0.684 (0.609–0.752)	0.538	0.746	0.644
CNN	0.943 (0.920–0.961)	0.934	0.951	0.934	0.947 (0.894–0.978)	0.938	0.955	0.938
Com	0.957 (0.942–0.971)	0.827	0.959	0.894	0.874 (0.820–0.924)	0.754	0.896	0.826

AUC, area under the curve; CI, confidence interval; Sen, sensitivity; Spe, specificity; Acc, accuracy; LR, logistic regression; SVM, support vector machine; Bayes, naïve Bayes; CNN, convolutional neural network; Com, comprehensive model.



**Figure 3** ROC curves of different models in the training (A) and test (B) sets were used to predict the mutation status of *EGFR* molecular subtypes. LR, logistic regression; SVM, support vector machine; Bayes, naïve Bayes; CNN, convolutional neural network; ROC, receiver operating characteristic; *EGFR*, epidermal growth factor receptor.



**Figure 4** Loss (A) and accuracy (B) curves for the CNN model with epochs. As the number of epochs increases, the loss in the training dataset decreases, indicating that the trained model converges. Simultaneously, as the number of epochs increases, the accuracy of the training dataset increases. The best training loss occurred at 80 epochs. At that epoch, the training loss and accuracy were 0.0505 and 94.1%, respectively. CNN, convolutional neural network.

the suspicious areas for all tumors were almost inside the tumors, and based on these observations, the deep learning model considered these 3 tumors *EGFR*-mutant tumors. In contrast, in the bottom row of *Figure 6B*, the cavity area of the tumor in the first 2 images caught the attention of the deep learning model and regarded it as an *EGFR* wild-type tumor. In the last image in the bottom row of *Figure 6B*, the area between the tumor and pleura was brought to the attention of the deep learning model, which similarly identified it as an *EGFR* wild-type tumor.

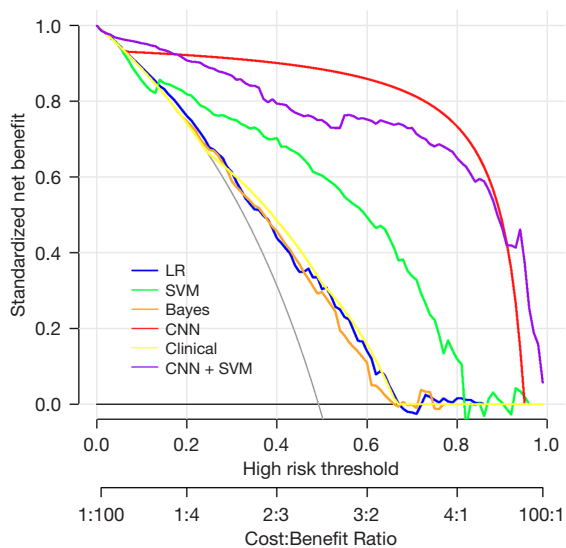
## Discussion

In this study, we presented a 3D CNN-based deep learning

method for non-invasively predicting of *EGFR* mutation status in patients with lung adenocarcinoma. The proposed CNN was successfully trained and tested using a manually segmented multicenter dataset. Compared with clinical, traditional radiomics, and comprehensive models, the CNN model showed superior performance in the training set (AUC =94.3%) and perfect performance in the further independent test set (AUC =94.7%). Furthermore, we used the Grad-CAM visualization technique to visually explain the prediction process of the CNN model, which helps us to better understand the black box of the CNN model.

Previous studies have shown that *EGFR* mutation status can be predicted based on clinical characteristics and radiomics features of patients with lung adenocarcinoma



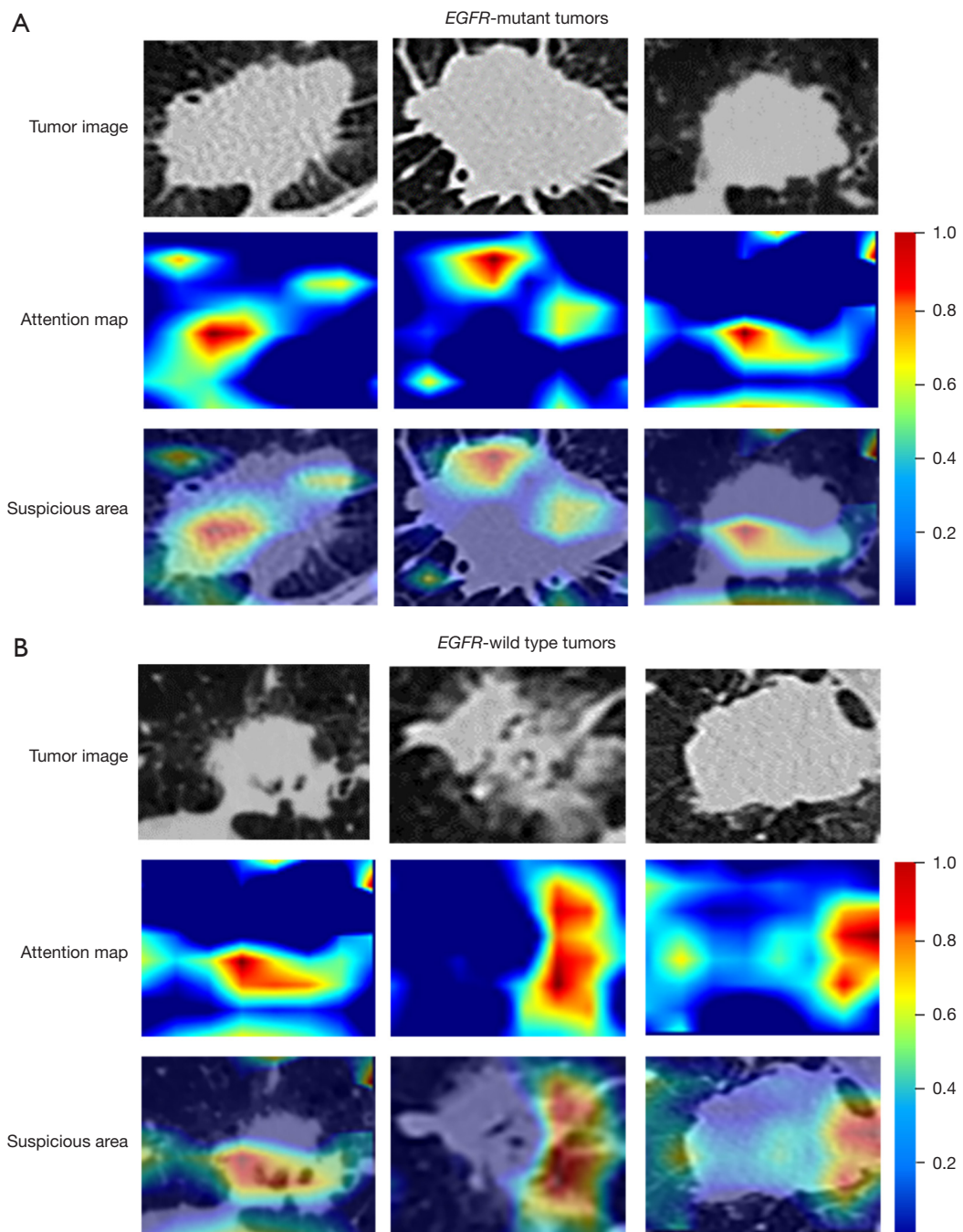


**Figure 5** Decision curves for different models in the training set. The X-axis represents threshold probability, and the Y-axis represents net income. The gray line represents the hypothesis that all patients have *EGFR* mutations, and the black line represents the hypothesis that no patients have *EGFR* mutations. LR, logistic regression; SVM, support vector machine; Bayes, naïve Bayes; CNN, convolutional neural network; *EGFR*, epidermal growth factor receptor.

(24,25). This finding was confirmed in our previous studies (7,9,26). In this study, a clinical model and 3 radiomics models were constructed for comparison with the proposed CNN model. The clinical model achieved an acceptable performance on the training and testing sets (AUCs of 69.6% and 68.4%, respectively). However, clinical characteristics reflect little information at the tumor pathology level. Moreover, the diagnostic performance of the model needs to be further improved. In contrast, radiomics utilizes computers to extract a large amount of biological and prognostic information from medical images that are unrecognizable to the human eye, and this hidden information in images has the potential to reflect tumor phenotypes (27). In this study, we constructed 3 radiomics models based on 3 classifiers (LR, SVM, and Bayes). Among them, the radiomics model based on SVM had the highest prediction performance, with AUCs of 84.3% and 72.4% in the training and test sets, respectively. Although radiomics models have evident advantages over a clinical model in predicting *EGFR* mutation status, in the face of massive medical image data, delineating tumor VOIs not only is time-consuming and labor-intensive but also

differs between different delineators. In addition, radiomics involves complex processes, such as feature extraction and screening (12). In this study, although the performance of the comprehensive model was improved in the training set, there was no statistical difference between the performance of the combined and CNN models. Simultaneously, the performance of the combined model in the test set did not increase, and over-fitting occurred. It can be seen that our proposed CNN model shows better prediction performance compared to the other models.

Recently, researchers have used deep learning methods to predict *EGFR* mutation status with encouraging results (18-20). However, the predictive performance of these studies varied between 75% and 81%. Compared with these studies, the prediction performance obtained in this study is currently the highest, with AUCs of 94.3% and 94.7% in the training and test sets, respectively. The superior performance of our CNN model may be attributed to the high-quality labeled data obtained. In this study, in order to prevent the loss of important volume information and avoid additional information interference, all VOIs were manually delineated layer-by-layer on CT slices. Wang *et al.* (19) and Zhao *et al.* (18) manually selected a cubic VOI that contained the entire tumor. Although this approach saves time and effort, it results in the inclusion of tissues other than the tumor, which increases the difficulty of training the model, particularly for datasets with relatively small sample sizes. Wang *et al.* (28) proposed a semi-supervised learning framework for 2-dimensional (2D) semantic segmentation technology in 3D CT segmentation, which not only reduces the requirement of high-level computational resources, but also improves the efficiency of manual annotation, as it only requires annotation on a small number of slices. However, the performance of this framework in medical image segmentation is still unknown. Another important factor may be related to our selection of network architecture that is relatively easy to train. In this study, we selected a 3D CNN with 17 layers, which includes 4 3D Conv layers consisting of 32, 64, 128, and 256 filters. 3D CNN models are widely used for object classification and detection under different data patterns, and CNN models with different network backbones can lead to different results. For example, Wang *et al.* (29) attempted for the first time to extend CNN to the classification and detection of prohibited items, and the results showed that the Voxception ResNet model has comparable performance in classification tasks compared to other networks such as Faster R-CNN and RetinaNet, but still needs improvement



**Figure 6** Using the Grad-CAM technique to find tumor suspicious areas. (A) *EGFR*-mutant; (B) *EGFR*-wild type. The first column is the original CT image. The second column is the attention map for classifying *EGFR* mutation status. The third column is the fused image generated by fusing the original image and the attention map to find suspicious regions of the tumor. *EGFR*, epidermal growth factor receptor; Grad-CAM, Gradient-weighted Class Activation Mapping; CT, computed tomography.

in detection tasks. Currently, many neural networks tend to be larger and deeper; however, deeper networks can be more difficult to train and can lead to performance degradation and increased training error (30-32). Wang *et al.* (19) performed transfer learning using the ImageNet dataset, which has recently emerged as a potentially effective method for alleviating data requirements. However, ImageNet is a dataset of relatively low-resolution 2D color photographs. Network applications of transfer learning are limited for processing volumetric images with higher resolution in radiology (33). Therefore, choosing an appropriate network architecture is crucial to improve the performance of the model.

Deep learning promises to be an adjunct to clinicians; therefore, visual interpretation of specific areas in images is critical to understanding these typical black-box models (30). This not only increases the confidence of clinical users in the system but also indirectly assesses the accuracy of the model. In this study, we used Grad-CAM technology for visual interpretation. Grad-CAM does not need to modify the structure of the original network, does not need to retrain the model, and can be applied to many different tasks (e.g., image classification, image captioning, and visual question answering) (23). Through visual analysis, it can intuitively show which area is the focus of the CNN model (Figure 6). Therefore, it may provide clinicians with an advantageous biopsy site to accurately detect *EGFR* mutation status and avoid missed diagnoses caused due to intra-tumor heterogeneity.

The potential clinical applications of the CNN model include the following: (I) the proposed CNN model is a non-invasive method for predicting *EGFR* mutation status; thus, it can effectively alleviate the pain of patients; (II) the CNN model directly uses the CT images of patients; therefore, this is effective in saving medical costs; (III) if the biopsy shows that the tumor is *EGFR* wild-type and the CNN model indicates *EGFR* mutation, the results may contain false negatives due to intra-tumor heterogeneity. In this case, clinicians may need to re-biopsy to avoid missed diagnosis (19,34); and (IV) the proposed CNN model can be reused in the diagnosis and treatment of lung cancer.

This study has some limitations. First, despite the superior performance of our CNN model in predicting *EGFR* mutation status, the study population was only from China. In future studies, we will include populations of different ethnicities and geographic areas to improve the generalizability of the model. Second, to truly reflect the tumor's own information and avoid interference from

surrounding tissues, we manually delineated the tumor boundary. However, in the face of massive medical data, delineating ROI more effectively, automatically, and accurately is a direction for future studies. Third, this study only predicted the mutation status of *EGFR*, and the mutation status of other driver genes of lung adenocarcinoma (such as *ALK*, *KRAS*, and *ROS1*) will be explored in future studies. Finally, due to possible sampling bias during the biopsy process, there may be false negative results in *EGFR* wild-type cases. Therefore, further prospective validation of this result is needed in future study.

## Conclusions

The CNN model proposed in this study showed excellent performance in non-invasively predicting *EGFR* mutation status and is expected to become an auxiliary tool for clinicians. Furthermore, we provide a visual perspective explanation of the inference process of deep learning, which will help us understand typical black-box models and increase the confidence of clinical users in the system.

## Acknowledgments

**Funding:** This work was supported by the National Natural Science Foundation of China (No. 82202147), the Sichuan Provincial Medical Research Project Plan (No. S23086), the Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital Research Fund (No. 2022QN25), Qinghai Province "Kunlun Talents High end Innovation and Entrepreneurial Talents" Top Talent Cultivation Project, Qinghai Provincial Department of science and technology of China (No. 2023 ZJ 918M), the Medical Science and Technology Research Fund Project of Guangdong Province (No. B2022144), and the Science and Technology Plan Fund of Guizhou Provincial [Qiankehe Foundation-ZK (2022) General 634].

## Footnote

**Reporting Checklist:** The authors have completed the TRIPOD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-24-33/rc>

**Conflicts of Interest:** All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-33/coif>). H.L. is an employee of GE Healthcare and has no financial

or other conflicts with respect to this study. The other authors have no conflicts of interest to declare.

**Ethical Statement:** The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This retrospective study was approved by the institutional review boards of Sichuan Provincial People's Hospital (No. 2022-254) and Lanzhou University Second Hospital (No. 2020A-180), and the requirement for individual consent for this retrospective analysis was waived.

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

## References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-49.
2. National Comprehensive Cancer Network. The NCCN clinical practice guidelines in oncology for non-small cell lung cancer (version 4. 2023). Available online: <https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1450>
3. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014;511:543-50.
4. Ramalingam SS, Vansteenkiste J, Planchard D, Cho BC, Gray JE, Ohe Y, et al. Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. *N Engl J Med* 2020;382:41-50.
5. Mok TS, Cheng Y, Zhou X, Lee KH, Nakagawa K, Niho S, Lee M, Linke R, Rosell R, Corral J, Migliorino MR, Pluzanski A, Sbar EI, Wang T, White JL, Wu YL. Improvement in Overall Survival in a Randomized Study That Compared Dacomitinib With Gefitinib in Patients With Advanced Non-Small-Cell Lung Cancer and EGFR-Activating Mutations. *J Clin Oncol* 2018;36:2244-50.
6. Yang X, Dong X, Wang J, Li W, Gu Z, Gao D, Zhong N, Guan Y. Computed Tomography-Based Radiomics Signature: A Potential Indicator of Epidermal Growth Factor Receptor Mutation in Pulmonary Adenocarcinoma Appearing as a Subsolid Nodule. *Oncologist* 2019;24:e1156-64.
7. Zhang G, Zhang J, Cao Y, Zhao Z, Li S, Deng L, Zhou J. Nomogram based on preoperative CT imaging predicts the EGFR mutation status in lung adenocarcinoma. *Transl Oncol* 2021;14:100954.
8. Yin G, Wang Z, Song Y, Li X, Chen Y, Zhu L, Su Q, Dai D, Xu W. Prediction of EGFR Mutation Status Based on (18)F-FDG PET/CT Imaging Using Deep Learning-Based Model in Lung Adenocarcinoma. *Front Oncol* 2021;11:709137.
9. Zhang G, Cao Y, Zhang J, Ren J, Zhao Z, Zhang X, Li S, Deng L, Zhou J. Predicting EGFR mutation status in lung adenocarcinoma: development and validation of a computed tomography-based radiomics signature. *Am J Cancer Res* 2021;11:546-60.
10. Li S, Ding C, Zhang H, Song J, Wu L. Radiomics for the prediction of EGFR mutation subtypes in non-small cell lung cancer. *Med Phys* 2019;46:4545-52.
11. Tu W, Sun G, Fan L, Wang Y, Xia Y, Guan Y, Li Q, Zhang D, Liu S, Li Z. Radiomics signature: A potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology. *Lung Cancer* 2019;132:28-35.
12. Aerts HJ. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. *JAMA Oncol* 2016;2:1636-42.
13. Orhac F, Soussan M, Maisonobe JA, Garcia CA, Vanderlinden B, Buvat I. Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med* 2014;55:414-22.
14. Willeminck MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, Folio LR, Summers RM, Rubin DL, Lungren MP. Preparing Medical Imaging Data for Machine Learning. *Radiology* 2020;295:4-15.
15. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, Goo JM, Aum J, Yim JJ, Cohen JG, Ferretti GR, Park CM; DLAD Development and Evaluation Group. Development and Validation of a Deep Learning-Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs. *JAMA Netw Open* 2019;2:e191095.
16. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH,

- Goo JM, Aum J, Yim JJ, Park CM; Deep Learning-Based Automatic Detection Algorithm Development and Evaluation Group. Development and Validation of a Deep Learning-based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. *Clin Infect Dis* 2019;69:739-47.
17. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, Tse D, Etemadi M, Ye W, Corrado G, Naidich DP, Shetty S. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019;25:954-61.
  18. Zhao W, Yang J, Ni B, Bi D, Sun Y, Xu M, Zhu X, Li C, Jin L, Gao P, Wang P, Hua Y, Li M. Toward automatic prediction of EGFR mutation status in pulmonary adenocarcinoma with 3D deep learning. *Cancer Med* 2019;8:3532-43.
  19. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, Liu Y, Gevaert O, Wang K, Zhu Y, Zhou H, Liu Z, Tian J. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J* 2019;53:1800986.
  20. Wang C, Xu X, Shao J, Zhou K, Zhao K, He Y, Li J, Guo J, Yi Z, Li W. Deep Learning to Predict EGFR Mutation and PD-L1 Expression Status in Non-Small-Cell Lung Cancer on Computed Tomography Images. *J Oncol* 2021;2021:5499385.
  21. Zunair H, Rahman A, Mohammed N, Cohen JP. Uniformizing Techniques to Process CT scans with 3D CNNs for Tuberculosis Prediction. In: Rekik I, Adeli E, Park SH, Valdés Hernández MdC. editors. *Predictive Intelligence in Medicine. PRIME 2020. Lecture Notes in Computer Science*, Springer, 2020;12329:156-68.
  22. He B, Song Y, Wang L, Wang T, She Y, Hou L, Zhang L, Wu C, Babu BA, Bagci U, Waseem T, Yang M, Xie D, Chen C. A machine learning-based prediction of the micropapillary/solid growth pattern in invasive lung adenocarcinoma with radiomics. *Transl Lung Cancer Res* 2021;10:955-64.
  23. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int J Comput Vis* 2020;128:336-59.
  24. Liu Y, Kim J, Qu F, Liu S, Wang H, Balagurunathan Y, Ye Z, Gillies RJ. CT Features Associated with Epidermal Growth Factor Receptor Mutation Status in Patients with Lung Adenocarcinoma. *Radiology* 2016;280:271-80.
  25. Choi CM, Kim MY, Hwang HJ, Lee JB, Kim WS. Advanced adenocarcinoma of the lung: comparison of CT characteristics of patients with anaplastic lymphoma kinase gene rearrangement and those with epidermal growth factor receptor mutation. *Radiology* 2015;275:272-9.
  26. Zhang G, Zhao Z, Cao Y, Zhang J, Li S, Deng L, Zhou J. Relationship between epidermal growth factor receptor mutations and CT features in patients with lung adenocarcinoma. *Clin Radiol* 2021;76:473.e17-24.
  27. Park H, Sholl LM, Hatabu H, Awad MM, Nishino M. Imaging of Precision Therapy for Lung Cancer: Current State of the Art. *Radiology* 2019;293:15-29.
  28. Wang Q, Breckon TP. On the evaluation of semi-supervised 2D segmentation for volumetric 3D computed tomography baggage security screening. 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 2021:1-8.
  29. Wang Q, Bhowmik N, Breckon TP. On the Evaluation of Prohibited Item Classification and Detection in Volumetric 3D Computed Tomography Baggage Security Screening Imagery. 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020:1-8.
  30. Cheng PM, Montagnon E, Yamashita R, Pan I, Cadrin-Chênevert A, Perdigón Romero F, Chartrand G, Kadoury S, Tang A. Deep Learning: An Update for Radiologists. *Radiographics* 2021;41:1427-45.
  31. Sun S, Chen W, Wang L, Liu X, Liu TY. On the Depth of Deep Neural Networks: A Theoretical View. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016. doi: 10.1609/aaai.v30i1.10243.
  32. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016:770-8.
  33. Chartrand G, Cheng PM, Vorontsov E, Drozdal M, Turcotte S, Pal CJ, Kadoury S, Tang A. Deep Learning: A Primer for Radiologists. *Radiographics* 2017;37:2113-31.
  34. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, Ye Z, Gillies RJ. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. *Clin Lung Cancer* 2016;17:441-448.e6.

**Cite this article as:** Zhang G, Shang L, Cao Y, Zhang J, Li S, Qian R, Liu H, Zhang Z, Pu H, Man Q, Kong W. Prediction of epidermal growth factor receptor (*EGFR*) mutation status in lung adenocarcinoma patients on computed tomography (CT) images using 3-dimensional (3D) convolutional neural network. *Quant Imaging Med Surg* 2024;14(8):6048-6059. doi: 10.21037/qims-24-33