European Journal of Nuclear Medicine and Molecular Imaging (2021) 48:361–371
https://doi.org/10.1007/s00259-020-04986-6

ORIGINAL ARTICLE

The deep learning model combining CT image and clinicopathological information for predicting ALK fusion status and response to ALK-TKI therapy in non-small cell lung cancer patients

Zhengbo Song[1] · Tianchi Liu[2,3] · Lei Shi[2,3] · Zongyang Yu[4] · Qing Shen[2,3] · Mengdi Xu[2,3] · Zhangzhou Huang[5] · Zhijian Cai[6] · Wenxian Wang[1] · Chunwei Xu[7] · Jingjing Sun[8] · Ming Chen[9]

Received: 2 May 2020 / Accepted: 31 July 2020 / Published online: 13 August 2020

Abstract

Purpose This study aimed to investigate the deep learning model (DLM) combining computed tomography (CT) images and clinicopathological information for predicting anaplastic lymphoma kinase (ALK) fusion status in non-small cell lung cancer (NSCLC) patients.

Materials and methods Preoperative CT images, clinicopathological information as well as the ALK fusion status from 937 patients in three hospitals were retrospectively collected to train and validate the DLM for the prediction of ALK fusion status in tumors. Another cohort of patients ($n = 91$) received ALK tyrosine kinase inhibitor (TKI) treatment was also included to evaluate the value of the DLM in predicting the clinical outcomes of the patients.

Results The performances of the DLM trained only by CT images in the primary and validation cohorts were AUC = 0.8046 (95% CI 0.7715–0.8378) and AUC = 0.7754 (95% CI 0.7199–0.8310), respectively, while the DLM trained by both CT images and clinicopathological information exhibited better performance for the prediction of ALK fusion status (AUC = 0.8540, 95% CI 0.8257–0.8823 in the primary cohort, $p < 0.001$; AUC = 0.8481, 95% CI 0.8036–0.8926 in the validation cohort, $p < 0.001$). In addition, the deep learning scores of the DLMs showed significant differences between the wild-type and ALK infusion tumors. In the ALK-target therapy cohort ($n = 91$), the patients predicted as ALK-positive by the DLM showed better performance of progression-free survival than the patients predicted as ALK-negative (16.8 vs. 7.5 months, $p = 0.010$).

Conclusion Our findings showed that the DLM trained by both CT images and clinicopathological information could effectively predict the ALK fusion status and treatment responses of patients. For the small size of the ALK-target therapy cohort, larger data sets would be collected to further validate the performance of the model for predicting the response to ALK-TKI treatment.

Keywords Deep learning model · Anaplastic lymphoma kinase · Computed tomography · Non-small cell lung cancer

Introduction

Non-small cell lung cancer (NSCLC) is one of the malignant tumors with the highest incidence and mortality [1]. At the time of the diagnosis, over 60% of the patients present with advanced lung cancer. Currently, treatments with targeting drugs can achieve substantial responses for advanced lung cancer patients with gene mutations. Anaplastic lymphoma kinase (ALK) fusion gene is the driver gene with the highest incidence of variation, following EGFR and KRAS mutations, which account for about 5–7% of all lung adenocarcinoma [2, 3]. To date, several ALK fusion gene-targeting inhibitors have

Zhengbo Song, Tianchi Liu and Lei Shi contributed equally to this work.

This article is part of the Topical Collection on Oncology – Chest

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00259-020-04986-6) contains supplementary material, which is available to authorized users.

✉ Ming Chen
chenming@zjcc.org.cn

Extended author information available on the last page of the article

already been applied in clinical practice, substantially improving the survival time and life quality of patients [4–8].

FISH, immunohistochemistry, and NGS are the currently recommended methods for the detection of ALK fusion gene [9–12]. Most of these techniques require tissue or liquid samples. In addition, the intra- and inter-tumor heterogeneities and tumor microenvironments could also influence the detection results. Therefore, new non-invasive, dynamic, low-price methods which could cover the global tissues and tumor microenvironment are needed for the detection of ALK fusion gene in clinical practice.

Previous studies have demonstrated that ALK inhibitors, including the first, second, and third generations of inhibitors, have high treatment responses [4–8]. However, there are still some patients who are not sensitive to ALK inhibitors. For instance, although treating ALK fusion gene-positive patients with crizotinib could achieve a high response rate of 60–70%, primary resistance to crizotinib is still found in about 10–20% of the patients [7, 8]. Nonetheless, there are no effective methods that could predict the responses to ALK inhibitors, and no non-invasive effective methods that could be used to dynamically monitor the responses to ALK inhibitors and predict drug resistance.

CT scanning is a commonly used technique for the diagnosis of lung cancer, which provides a non-invasive method for the analysis of lung cancer. Recent studies have shown that features obtained from CT images are associated with the status of genes, including EGFR and KRAS, and have predictive values for the ALK fusion gene [13–16]. In addition, CT imaging could cover the complete area of tumors, including the tumor microenvironments and attached tissues, which could allow us to include tumor heterogeneity when predicting ALK gene status. In addition, CT scanning is non-invasive, and data of CT imaging could be easily accessed during the overall treatment procedures.

Associations between radiomics and patient outcomes have been shown for NSCLC [17–19]. Previous studies have demonstrated that the characteristics obtained by radiomics methods have predictive values for genes, including ALK, EGFR, and KRAS [13–16, 20, 21]. In addition, there are several studies which identified the association between PDL1 and radiomics features [22, 23]. However, such radiomics-based methods based on feature engineering can only reflect the general features, which lack the specificity of the ALK fusion gene. Also, radiomics methods rely on precise outlining of tumor borders, which can only be done manually. Such radiomics characteristics only consider the tumor areas, while the microenvironments and attached tissues are largely ignored. In comparison, advanced artificial intelligence (AI), such as deep learning, could overcome such disadvantages due to the powerful learning capability of deep learning. Currently, deep learning models have shown expert-level performances in various medical image-based diagnostic tasks [24–26]. Deep learning methods have achieved high performance in predicting EGFR mutation [27, 28].

In this study, large-scale data (from 937 patients) were collected from three independent hospitals, and the CT images and clinicopathological information were used to train the DLMs for the prediction of ALK gene status and evaluate the model performances in an independent validation cohort. In addition, this study also analyzed the tumor areas strongly associated with ALK fusion on the CT images, which could validate the reliability of the models and help the clinicians to further make clinical decisions. Another group of patients undergoing ALK-TKI treatment was included in the study to evaluate the potential of the DLM in predicting the responses of patients to targeting therapy.

Materials and methods

Patients

The institutional review board of Zhejiang Cancer Hospital, Fujian Cancer Hospital, and the 900th Hospital in China approved this retrospective study and waived the need to obtain informed consent from the patients. Patients older than 18 years who meet the following inclusion criteria were included in the study: (1) pathologic examination of tumor specimens carried out with proven records of ALK gene status; (2) preoperative CT data obtained; and (3) recorded clinicopathological information, including smoking history, age, gender, and histological type of lung cancer.

Patients were excluded if (1) clinical data including age, gender, and stage were missing; (2) preoperative treatment was received; and (3) the duration between CT examination and subsequent surgery exceeded 1 month. Finally, 937 patients from three hospitals were enrolled in this study. We allocated the patients into a primary cohort and an independent validation cohort, according to the hospital. The primary cohort included 651 patients from Zhejiang Cancer Hospital between January 2013 and July 2018. The validation cohort included 286 patients from Fujian Cancer Hospital and 900th Hospital between January 2016 and December 2018. The primary and validation cohorts were used for developing and validation of the proposed models, respectively.

Collection of ALK-targeting therapy data

Data of the patients with advanced NSCLC who were diagnosed as ALK fusion gene-positive in the Zhejiang Cancer Hospital between 2014 and 2018 were retrospectively collected. The inclusion criteria were as follows: (1) those diagnosed with NSCLC according to the WHO 2015 pathological criteria, and immunohistochemically (D5F3 antibody) proven with positive ALK fusion gene; and (2) those diagnosed with

advanced NSCLC or locally advanced NSCLC who could not be treated with radical chemoradiotherapy, according to the eighth edition of TNM criteria and (3) received crizotinib for the treatment; (4) whose complete follow-up data were available; and (5) who received chest CT scanning within 2 weeks before crizotinib treatment. The exclusion criteria were as follows: (1) those with metastatic tumors from extra-pulmonary tissues; (2) those with small cell lung cancers (SCLCs) or other types of lung cancer; and (3) the incomplete follow-up data. This study was approved by the Ethics Committee of the Zhejiang Cancer Hospital.

Development of deep learning model (DLM)

The architecture of the DLM is illustrated in Fig. 1. It consisted of three blocks, namely the CT image encoder, the clinicopathological information encoder, and the classifier. The CT image encoder aimed to generate a meaningful representation of a tumor based on CT scans. Two radiologists first manually annotated the tumor regions as follows: (1) a 2D tumor bounding box was drawn in the CT slice containing the largest tumor cross-section; (2) the indexes of other slices containing the tumor were recorded. A three-dimensional (3D) bounding box, as region of interest (ROI), was calculated based on the annotation, such that it contained the entire tumor region, including the edges of the tumor. Subsequently, the ROI was resized to $64 \times 64 \times 32$ pixels by trilinear interpolation; contrast stretching was applied using the lung window setting (window width 1600 and window level −450); and pixel intensity was normalized to (0, 1). The processed ROI image was used as the input to the CT image encoder, which has a modified 3D ResNet10 [29, 30] consisting of a sequence of convolution layers. The output of the last convolution layer was fed to an average pooling layer to generate a one-dimensional feature vector 512 in size as the final representation of the tumor's CT image.

The clinicopathological information encoder aimed to generate the representation of the clinical information of the patient and pathological information of the tumor. We first formed a one-dimensional feature vector of size eight based on four types of clinicopathological information, namely sex, age, smoking history, and tumor pathology. Except for one element representing the normalized age (using z-score with shifted mean 0.5), the other seven elements in the feature vector were binary indicators (with value 0 or 1) of male, female, light smoking, heavy smoking, adenocarcinoma, squamous, and other tumor pathological type, respectively.

We trained the three blocks in an end-to-end manner with cross-entropy loss function and applied modern techniques such as weight decay regularization, batch normalization [31], dropout [32], stochastic gradient descent optimization algorithm, Nesterov momentum [33], and data augmentation. Specifically, the "Kaiming uniform" method [34] was used to initialize the weights in the DLM. A Nesterov momentum SGD with momentum = 0.9 was used as a neural optimizer. Other training settings included a learning rate of 0.0005, batch size of 32, weight decay of 0.0001, and dropout rate of 0.1. Data augmentation techniques applied to CT images included random flipping, translation, rotation, and scaling. To alleviate the class imbalance problem, the minority class was upsampled to have the same number of samples as the majority class. We selected and evaluated the model on the primary cohort using fivefold cross-validation. The model was trained for up to 30 epochs and the best epoch was selected based on the AUC on the primary cohort. Training in the primary cohort resulted in five models; the averaged predictions from all five models were used for evaluation in the validation cohort.

Detection of ALK fusion gene

Immunohistochemistry using the D5F3 antibody, which has already been widely used for the detection of ALK fusion gene, was used for the detection in all the patients, and the results were confirmed by two senior pathologists. The detailed detection methods were described in previous studies [35, 36].

Statistical analysis

The independent sample t test was adopted to assess the statistical differences between the mean values of ages in the primary and validation cohorts. The same statistical analysis was performed to assess the statistical differences in deep learning scores between the ALK fusion and ALK wild-type groups. Fisher's exact test was used to evaluate the differences in categorical variables such as gender and smoking history between cohorts. DeLong's test was used to assess the statistical difference between AUCs of two models. Keras toolkit and Python 3.6 were used for the implementation of the DLM. The PFS from the patients after crizotinib treatment was defined as the time from the initiation of crizotinib treatment to tumor progression or last followed up. Data were analyzed with SPSS 20 software. Kaplan-Meier curve method was used for the analysis of PFS, and Cox regression analysis was used for the multivariate analysis. The last follow-up for all the patients was on December 20, 2019.

Results

Clinicopathological characteristic of patients

Table 1 includes the clinicopathological characteristics of the patients in the primary and validation cohorts. The age, sex, pathological types, and smoking history did not significantly

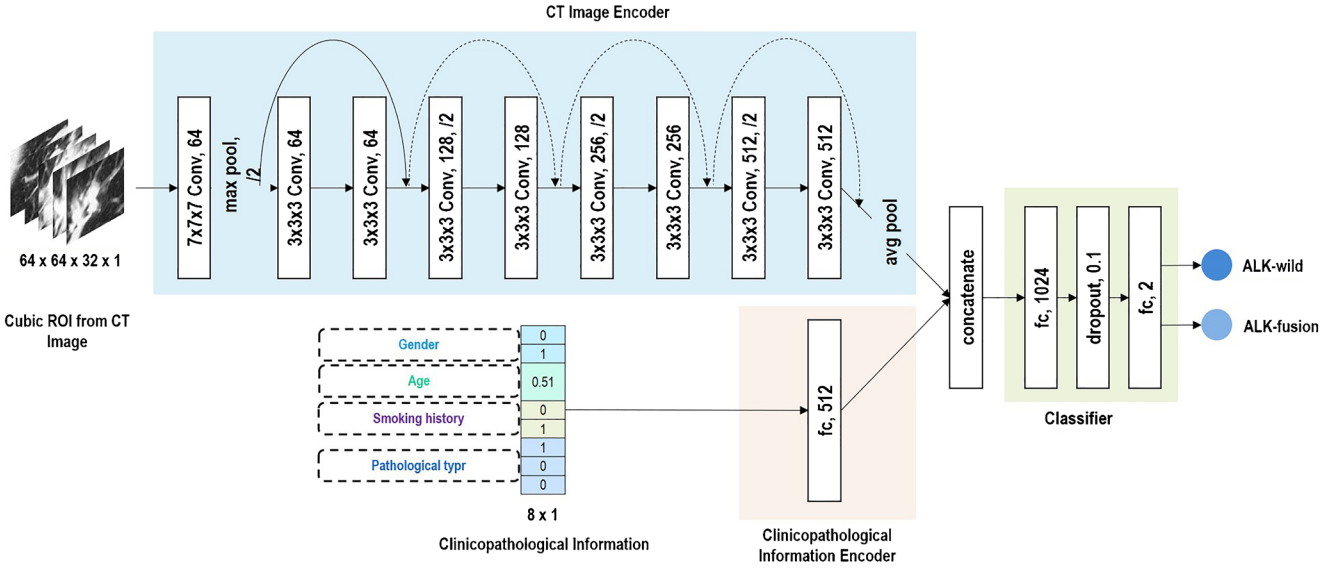

Fig. 1 Neural network architecture. The proposed DLM consists of three blocks, namely the CT image encoder, the clinicopathological information encoder, and the classifier. The CT image encoder has a three-dimensional (3D) ResNet10 architecture, which consists of an initial convolutional layer with kernel size $7 \times 7 \times 7$ and number of filters 64, a max pooling layer with kernel size $3 \times 3 \times 3$, and a sequence of residual blocks with kernel size $3 \times 3 \times 3$ and increasing number of filters (i.e., 64, 128, 256, 512). All convolution layers use stride 1 and are followed by batch normalization and ReLU activation. The output of the last convolution layer was fed to an average pooling layer to generate a one-dimensional feature vector of size 512 as the final representation of the tumor's CT image. The clinicopathological information encoder consists of a fully connected layer of size 512. The outputs of two encoders are combined by channel-wise concatenation to form the combined feature vector. The classifier includes one fully connected layer of size 1024, a dropout layer, and a softmax output layer with two nodes (corresponding to ALK wild-type and ALK fusion). Two variations of the models can be obtained with little architecture modification: the first is a CT image model consisting of only CT image encoder and classifier; the second is a clinicopathological model consisting of clinicopathological information encoder and classifier

differ between the primary and validation cohorts. Besides, the distributions of patients with ALK fusion and wild-type ALK gene were similar in both the primary cohort (195 ALK fusion vs. 456 ALK wild-type) and validation cohort (72 ALK fusion vs. 214 wild-type ALK). As clinicopathological information such as age, sex, smoking history, and histological types could reflect the differences between the patients with ALK fusion and wild-type ALK, a clinicopathological model was obtained by training with these characteristics and was then compared with the DLM.

Predictive performance of the DLMs

The area under the receiver operating characteristic (ROC) curve (AUC), accuracy, sensitivity, and specificity were used to evaluate the performances of the models (Table 2). The DLM trained by both CT images and clinicopathological characteristics showed good performance in the primary cohort (AUC = 0.8540, 95% CI 0.8257–0.8823), which was confirmed by the validation cohort (AUC = 0.8481, 95%CI0.8036–0.8926). The predictive performance of the model in the primary and validation cohorts was very similar, suggesting that the DLM had high feasibility in predicting the ALK fusion mutation in random patients. The model trained by both CT images and clinicopathological characteristics showed significantly higher performance compared with the DLM trained only by CT images (AUC = 0.8046, 95% CI 0.7715–0.8378 for the primary cohort, $p < 0.001$, and AUC = 0.7754, 95% CI 0.7199–0.8310 for the validation cohort, $p < 0.001$). In this study, a clinicopathological model was also built for comparison with the two deep learning models. This clinicopathological model used gender, age, smoking history, and pathological types as the characteristics to predict the ALK mutation. The performance of the DLM trained by both CT images and clinicopathological characteristics was significantly higher compared with the clinicopathological model (AUC = 0.7565, 95% CI 0.7193–0.7937 in the primary cohort, $p < 0.0001$, and AUC = 0.7547, 95% CI 0.6965–0.8129 in the validation cohort, $p < 0.005$). Figure 2a shows the ROC curves of the models. The deep learning scores of these two learning models in patients with ALK fusion and wild-type ALK were also significantly different in both the primary and validation cohorts (Fig. 3).

Figure 2b shows the decision curves of all the models. According to the decision curves, the DLM trained by both CT images and clinicopathological information had a higher beneficial effect than either the treat-all-patients scheme or the treat-none scheme, when the threshold probability was greater than 10% [37]. In addition, the DLM trained by both CT images and clinicopathological information had higher benefits than the DLM trained by either CT images or the clinical model.

Table 1 Clinical characteristics of patients in the primary and validation cohorts

Characteristic	Primary cohort	Validation cohort	*P* value
Subjects, *n*	651	286	
Sex			0.320
Male	310	126	
Female	341	160	
Age at treatment, year, median (range)	61 (24–82)	61 (33–85)	0.747
Tumor pathology			0.278
Adenocarcinoma	537	248	
Squamous	57	20	
Others	57	18	
Smoking history			0.268
Heavy	149	52	
Light	75	34	
No	427	200	
ALK			0.157
ALK fusion	195	72	
ALK wild	456	214	

Discovery of suspicious area

In order to explain the prediction process of the DLM and assess the reliability of the prediction, we adopted a visualization method Grad-CAM [38] that generates suspicious areas for a trained DLM on an input image. The suspicious area indicates the discriminative image regions used by the DLM for classifying the image; in this study, the highlighted regions associated with ALK fusion type are referred to as suspicious areas.

Figure 4 depicts the suspicious areas discovered by the DLM trained on both CT image and clinicopathological information. These suspicious areas varied among different tumors. For instance, the suspicious areas in Fig. 4a and b are central tumor areas. Based on patterns in these areas, the DLM explained these two tumors as ALK fusion. On the other hand, the DLM also registered the cavity areas in Fig. 4c and tumor and adjacent tissue in Fig. 4d and predicted them as wild-type ALK. Although some input images contained normal tissues, such as in Fig. 4b, the model discovered the suspicious areas inside the tumor without being influenced by normal tissues. The suspicious areas in accurate predictions could alert the clinicians to distinct patterns related to tumor ALK gene and may inspire future clinical research; unreasonable suspicious areas, such as completely off the tumor, can be an indicator of an unreliable prediction.

Predicting the response of ALK target therapy patients

Patients treated with crizotinib in Zhejiang Cancer Hospital between 2014 and 2018 who met the inclusion and exclusion criteria were consecutively included, and finally, 91 eligible patients were included. The general characteristics of the patients are shown in Table S1.

The median PFS of the 91 patients was 12.2 months (95% CI 9.2–15.2). When the ALK DL model was applied to the 91 patients, 79 were classified as ALK-positive, and 12 were classified as ALK-negative. Our findings showed that according to the classification results, the PFS of ALK-positive

Table 2 Predictive performance of various methods in the primary and validation cohorts

Models	Cohorts	AUC (95% CI)	Accuracy % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Clinicopathological information	Primary	0.7565 (0.7193–0.7937)	70.51 (67.00–74.01)	70.77 (64.39–77.15)	70.39 (66.20–74.58)
	Validation	0.7547 (0.6965–0.8129)	67.83 (62.42–73.25)	70.83 (60.33–81.33)	66.82 (60.51–73.13)
CT image	Primary	0.8046 (0.7715–0.8378)	72.35 (68.91–75.79)	73.33 (67.13–79.54)	71.93 (67.81–76.05)
	Validation	0.7754 (0.7199–0.8310)	69.23 (63.88–74.58)	79.17 (69.79–88.55)	65.89 (59.54–72.24)
CT image and clinicopathological information	Primary	0.8540 (0.8257–0.8823)	76.65 (73.40–79.90)	77.44 (71.57–83.30)	76.32 (72.41–80.22)
	Validation	0.8481 (0.8036–0.8926)	77.27 (72.42–82.13)	84.72 (76.41–93.03)	74.77 (68.95–80.59)

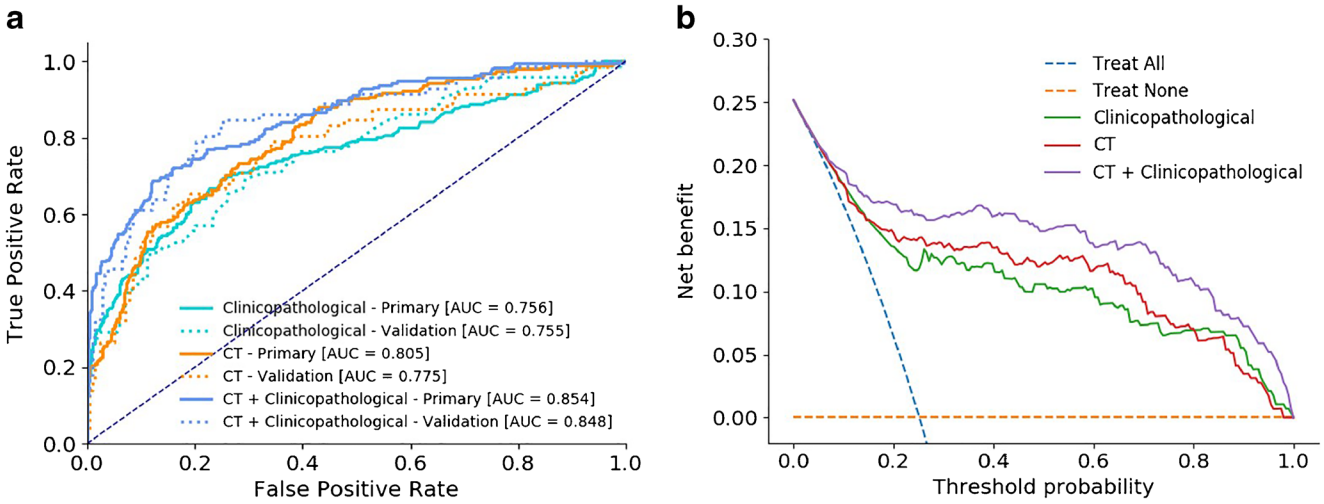

Fig. 2 Predictive performance of the deep learning model (DLM). **a** ROC curves of the three DLMs trained by clinicopathological information, CT images, and both in the primary/validation cohorts. **b** Decision curve of the DLMs. The blue line represents the benefit of treating all the patients as ALK wild-type, and the orange line represents the benefit of treating all the patients as ALK fusion. The purple line shows the benefit of using the DLM trained by both CT images and clinicopathological information. The red line and green line show the benefit of using DLMs trained by CT images and clinicopathological information, respectively

patients was significantly higher compared with that of ALK-negative patients after crizotinib treatment (16.8 vs. 7.5 months, $p = 0.010$)(Fig. 5). Also, the univariate analysis showed that age ($p = 0.012$) and performance status (PS) score ($p = 0.046$) were significantly associated with the response to crizotinib treatment, while smoking history ($p = 0.050$), sex ($p = 0.297$), stage ($p = 0.625$), pathological type ($p = 0.667$), and lines of crizotinib treatment ($p = 0.249$) were not significantly associated with PFS. Therefore, sex, PS score, and ALK status according to the model evaluation were included in the multivariate analysis, which showed that all these factors could influence the PFS of patients. The univariate and multivariate analysis results are shown in Table S2.

Discussion

In this study, we built a deep learning model trained by both CT images and clinicopathological characteristics for the prediction of ALK gene status in NSCLC patients. The DLM was trained by the data from the primary cohort (which included 651 patients) and was validated by an independent validation cohort from another two hospitals (which included 286 patients). The DLM showed encouraging results in the primary cohort (AUC = 0.854, 95% CI 0.8257–0.8823) and excellent results in the validation cohort (AUC = 0.8481, 95% CI 0.8036–0.8926). The DLM showed that high-dimensional features of CT images were significantly associated with

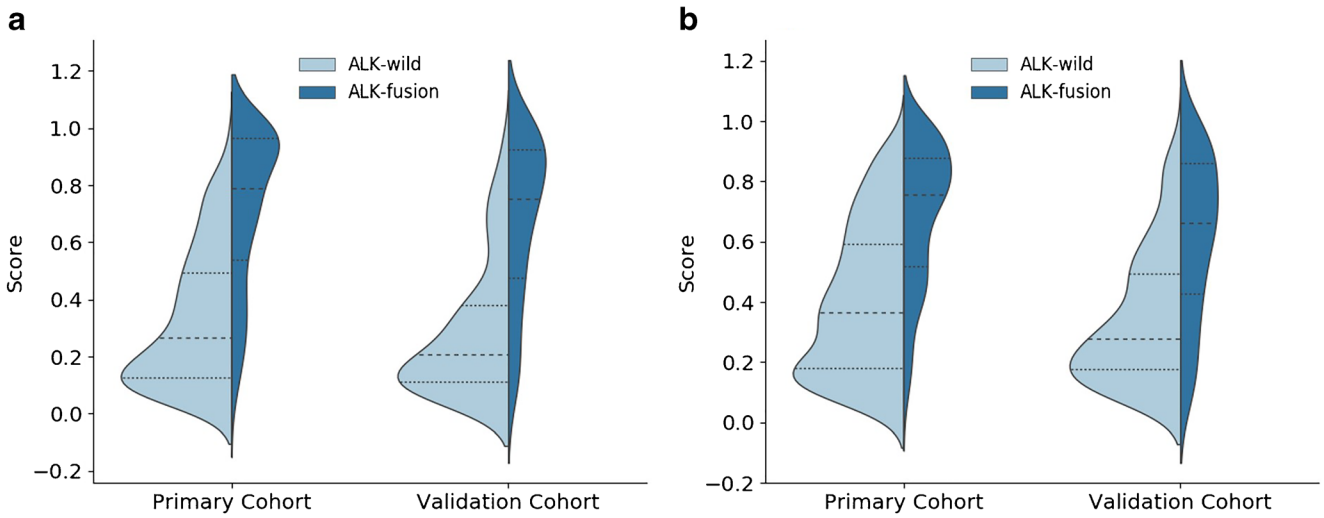

Fig. 3 Deep learning score of the DLMs. Deep learning scores of the DLM (**a**) trained by both CT images and clinicopathological information and the DLM (**b**) only trained by CT images between ALK fusion and ALK wild-type groups in the primary and validation cohorts. The horizontal dotted lines are the quartiles

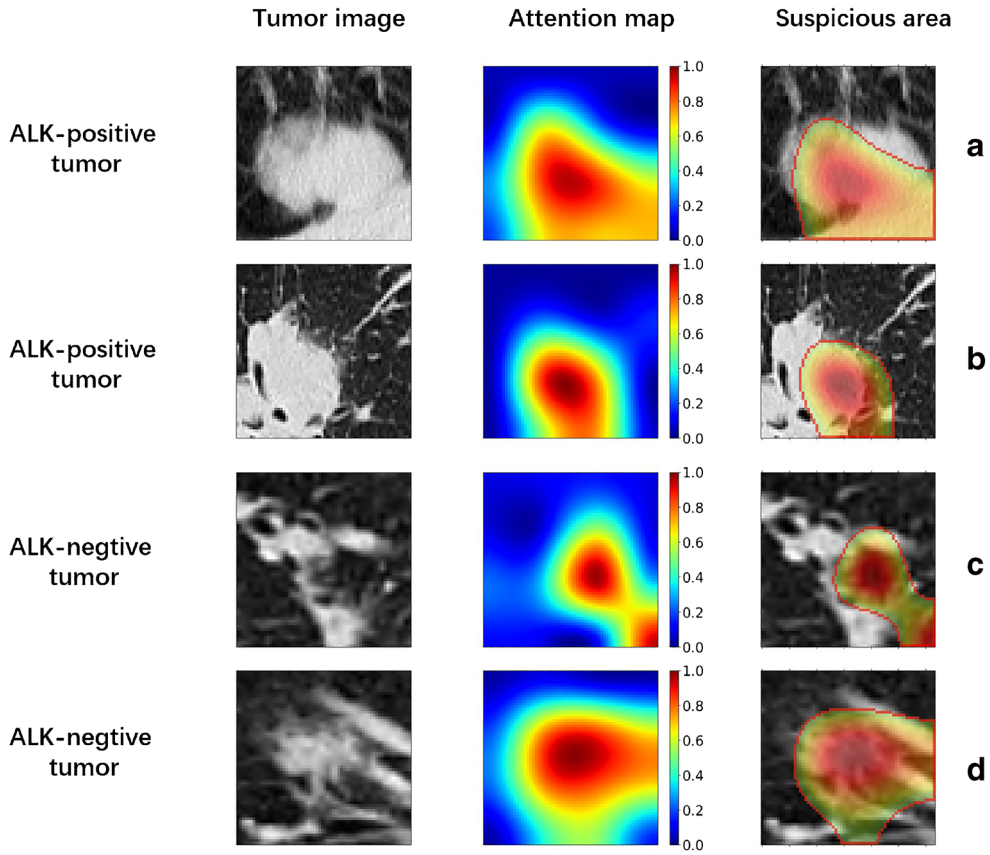

Fig. 4 Visualization of suspicious areas. **a**, **b** CT image predicted as ALK fusion by the DLM. **c**, **d** CT image that predicted as wild-type ALK by the DLM. The three columns show the original input image, the heat map for category of "ALK fusion" generated by Grad-CAM, and the heat map with a cut-off value of 0.5 superimposed on the input image, respectively

ALK gene status. The model trained by both CT images and clinicopathological characteristics had higher performances in predicting ALK gene status compared with the model trained only by CT images, which further suggested that clinicopathological information was also closely associated with ALK gene status, and the model based on both CT images and clinicopathological characteristics was more effective for the prediction of ALK gene status. To the best of our knowledge, this is the first study that combined CT image features and clinicopathological data with training DLMs for the prediction of ALK gene status. Besides, another prediction model was also built by training with clinicopathological information. Comparing with this clinicopathological model, the DLM showed significantly higher performance in predicting the ALK fusion gene. In this study, the clinicopathological factors achieved AUC = 0.7565 (95% CI 0.7193–0.7937) in the primary cohort and AUC = 0.7547 (95% CI 0.6965–0.8129) in a validation cohort. As clinicopathological factors have been widely applied, and nomogram is a linear model, the clinicopathological model could be interpreted. However, clinicopathological characteristics could only reflect relatively few general characteristics of tumors and lacked specificity in predicting the ALK gene as well as responses to ALK-TKI treatment.

This study provided a new method for the assessment of ALK fusion in patients. In addition, the presented model could also be used to analyze the suspicious tumor areas dominating the prediction of ALK gene status. Such analysis provided a visualized explanation for understanding the prediction results by CT data. The DLM provided an easy-to-use method for the prediction of the ALK fusion status.

To the best of our knowledge, this is the first study that used DLM to predict the ALK gene status. A previous study used CT images, positron emission tomography, and four clinical parameters from 128 NSCLC patients (47 ALK-positive, 17 ROS1/RET-positive, and 64 negative patients) to train radiomics model for the prediction of ALK fusion status. The sensitivity and specificity of their radiomics model were 0.7344 and 0.7031, respectively [15]. The DLM built in our study was substantially better than the radiomics model reported in the previous study. In addition, several other studies have also used clinical characteristics, semantic features, imaging features, and deep learning to build models for the prediction of EGFR mutation [13, 20, 21, 27]. In such studies, Wang and colleagues reached a new level by building a DLM to predict the status of EGFR mutation [27]. Their findings demonstrated that DLM had substantial advantages compared with radiomics models. In this study, deep learning was adopted to predict the status of ALK fusion, achieving high predictive performances. In addition, the present study also modified the deep neural network to integrate clinical information into the model, thus substantially improving the predictive performance of the DLM.

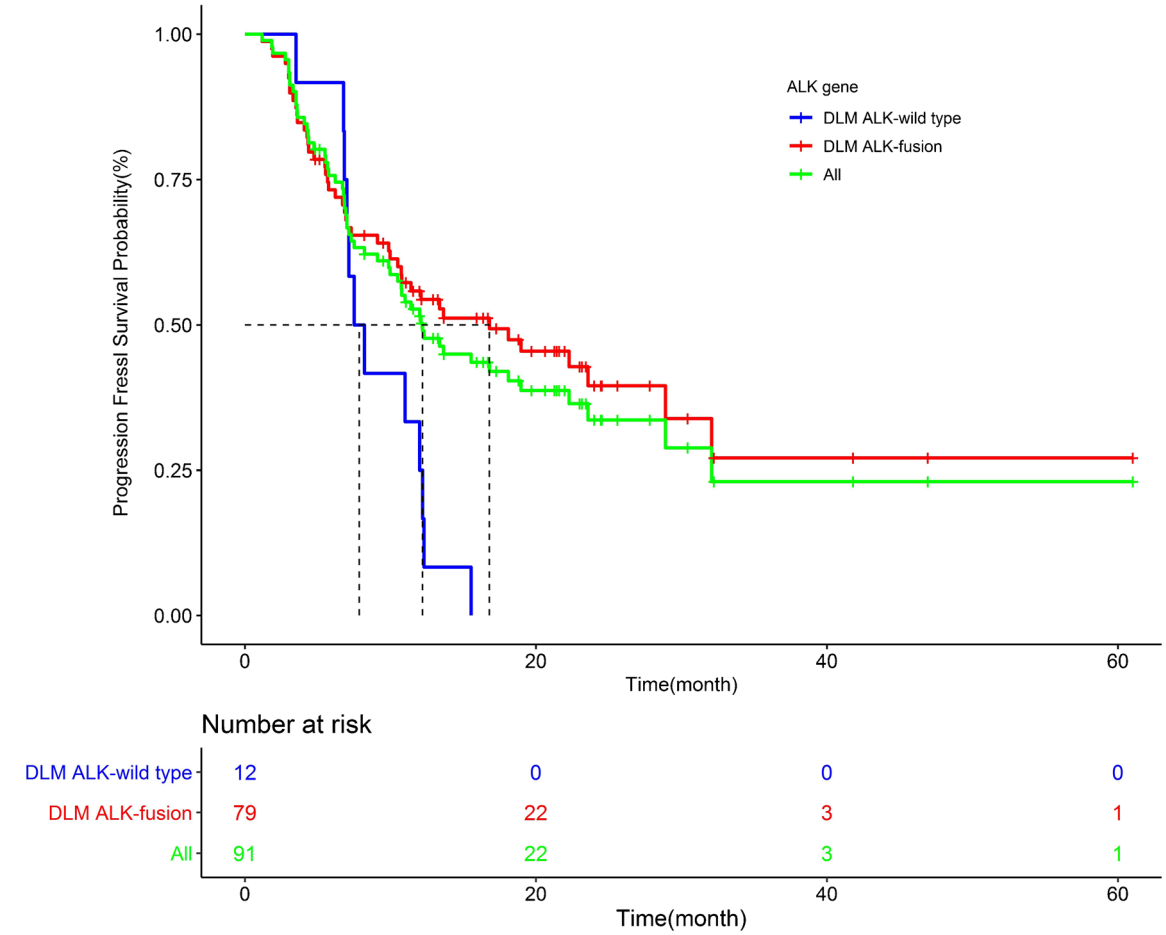

Fig. 5 Comparison of crizotinib treatment efficacy in ALK-positive, ALK-negative patients classified by DLM, and all ALK-TKI therapy cohort. (*P* value was 0.010 between ALK-positive and ALK-negative patients)

This study further evaluated the potential of the DLM in predicting the responses to ALK-TKI treatment. Drugs for targeting treatment substantially increased the survival time in patients with advanced lung cancer. For ALK-positive lung cancer patients, the median PFS after treatment with crizotinib could reach 10–12 months, and the objective response rate in patients was 60–70%. However, the responses in some patients were still suboptimal, although no effective methods are available to identify such patients. Previous studies have shown that the "abundance of tumor genes" could influence the responses to target treatment drugs [39]; however, there are still no non-invasive methods for evaluation of "abundance of tumor genes." In this study, we firstly used the deep learning method to further evaluate the pathologically suggested ALK-positive patients, which showed that 13.2% of such patients were classified as "imaging ALK-negative patients," of whom the responses were significantly poorer than the "imaging ALK-positive patients." Therefore, the model provided a non-invasive method that could effectively classify the responses to targeted treatment methods. Combining the gene classification results by CT image-based deep learning and pathological results to further clarify the association between them could be a new study direction.

The DLM has several potential clinical applications: (1) the DLM provides a non-invasive method to predict the status of ALK gene, which could be conveniently applied in routine CT diagnosis; (2) the wild-type ALK suggested by pathological examinations of biopsies can include false-negative results due to the inter-tumor heterogeneity, which could be overcome by our DLM. For patients whom the DLM predicted as ALK-positive, biopsies should be obtained for validation; (3) the DLM only requires the routine CT images and general clinicopathological information, without any additional expenses. Therefore, the model could be repeatedly applied during the treatment processes; and (4) although in this study we only investigated the status of ALK fusion, the model trained by both images and clinicopathological characteristics also showed substantial performance in predicting the outcomes in patients after ALK-TKI treatment. Future models combining chest CT images, clinicopathological information, and molecular pathological results could better predict the gene status and guide the targeting therapy in clinical practices.

The present study has several limitations: (1) This was a retrospective study and the presented model needs to be further validated by prospective studies. (2) The DLM mainly used

surgical samples, while only a few small biopsy samples were used, which could influence the overall stability of the model. However, according to the drug therapy results of the 91 patients with advanced lung cancer, the sensitivity of the model could be repeated in patients with advanced lung cancer. (3) The sample size of the clinical validation cohort was relatively low ($n = 91$), especially for the "image ALK-negative patients" ($n = 12$). More studies with larger sample size are needed to further validate the reliability of the model. (4) As new generation ALK inhibitors such as alectinib and ceritinib have been approved in clinical practice, validating the stability of the model to predict the response to these ALK inhibitors in the future studies is needed. Last but not least, other gene fusions, such as ROS1, RET, and NTRK, have similar clinicopathological characteristics with ALK fusion, and it is unclear that whether the DLM could distinguish different fusion genes for rare frequency of these genes in lung cancer.

In summary, the deep learning model combining CT images and clinicopathological information could effectively classify the status of the ALK gene and predict the responses to ALK inhibitor treatment in clinical practice. This study provided a non-invasive, rapid, and simple complementary method to guide clinical genetic diagnosis and targeting therapy. If validated by prospective studies with large sample sizes, this method could substantially increase the precision of targeting therapy in clinical practice.

Authors' contributions ZBS, ZZH, ZYY, CXW JJS, and WXW recruited the patients, collected and analyzed the data, and did the statistical analyses. LS and MC were responsible for the management of patients and conducted the study. ZBS and TCL and QS and MDX conducted the study and analyzed the data. QS and TCL interpreted CT data. LS and ZJC planned and led the study, analyzed the data, interpreted the results, and wrote the first draft of the article. All authors contributed to the article's revision, agreed to its submission, and had full access to the original data.

Funding This study was funded by the National Natural Science Foundation of China (grant 81802276).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This study was conducted in accordance with the Declaration of Helsinki and approved by the institutional ethics committee of Zhejiang Cancer Hospital, Fujian Cancer Hospital, and the 900th Hospital.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. Wiley. 2020;70(1):7–30 https://pubmed.ncbi.nlm.nih.gov/31912902/. Accessed 10 July 2020.
2. Blackhall FH, Peters S, Bubendorf L, et al. Prevalence and clinical outcomes for patients with ALK-positive resected stage I to III adenocarcinoma: results from the European Thoracic Oncology Platform Lungscape project. J Clin Oncol. American Society of Clinical Oncology. 2014;32(25):2780–7.
3. Soda M, Choi YL, Enomoto M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561–6 http://www.ncbi.nlm.nih.gov/pubmed/17625570. Accessed 11 Apr 2020.
4. Peters S, Camidge DR, Shaw AT, et al. Alectinib versus crizotinib in untreated ALK-positive non–small-cell lung cancer. N Engl J Med. Massachussetts Medical Society; 2017;377(9):829–838. http://www.nejm.org/doi/10.1056/NEJMoa1704795. Accessed 11 Apr 2020.
5. Solomon BJ, Kim DW, Wu YL, et al. Final overall survival analysis from a study comparing first-line crizotinib versus chemotherapy in alk-mutation-positive non–small-cell lung cancer. J Clin Oncol. American Society of Clinical Oncology. 2018;36(22):2251–8.
6. Shaw AT, Gandhi L, Gadgeel S, et al. Alectinib in ALK-positive, crizotinib-resistant, non-small-cell lung cancer: a single-group, multicentre, phase 2 trial. Lancet Oncol. Lancet Publishing Group; 2016;17(2):234–242. http://www.ncbi.nlm.nih.gov/pubmed/26708155. Accessed 11 Apr 2020.
7. Ou SHI, Ahn JS, De Petris L, et al. Alectinib in crizotinib-refractory alk-rearranged non-small-cell lung cancer: a phase II global study. J Clin Oncol. American Society of Clinical Oncology. 2016;34(7):661–8.
8. Shaw AT, Solomon BJ, Besse B, et al. ALK resistance mutations and efficacy of lorlatinib in advanced anaplastic lymphoma kinase-positive non–small-cell lung cancer. J Clin Oncol. American Society of Clinical Oncology; 2019;37(16):1370–1379. https://pubmed.ncbi.nlm.nih.gov/30892989/. Accessed 10 July 2020.
9. Lu S, Lu C, Xiao Y, et al. Comparison of EML4-ALK fusion gene positive rate in different detection methods and samples of non-small cell lung cancer. J Cancer. Ivyspring International Publisher; 2020;11(6):1525–1531. https://pubmed.ncbi.nlm.nih.gov/32047559/. Accessed 10 July 2020.
10. McLeer-Florin A, Duruisseaux M, Pinsolle J, et al. ALK fusion variants detection by targeted RNA-next generation sequencing and clinical responses to crizotinib in ALK-positive non-small cell lung cancer. Lung Cancer Elsevier Ireland Ltd. 2018;116:15–24.
11. Detection of ALK, RET, ROS1, NTRK1 and MET rearrangements and actionable mutations using next generation sequencing in patients with non-small cell lung cancer. Ann Oncol. https://www.annalsofoncology.org/article/S0923-7534(19)32386-5/fulltext. Accessed 11 Apr 2020.
12. McCoach CE, Blakely CM, Banks KC, et al. Clinical utility of cell-free DNA for the detection of ALK fusions and genomic mechanisms of ALK inhibitor resistance in non-small cell lung cancer. Clin Cancer Res NIH Public Access. 2018;24(12):2758.
13. Gevaert O, Echegaray S, Khuong A, et al. Predictive radiogenomics modeling of EGFR mutation status in lung cancer. Sci Rep Nature Publishing Group. 2017;7(1):1–8.
14. Yip SSF, Kim J, Coroller TP, et al. Associations between somatic mutations and metabolic imaging phenotypes in non-small cell lung cancer. J Nucl Med. Society of Nuclear Medicine Inc. 2017;58(4):569–76.
15. Yoon HJ, Sohn I, Cho JH, et al. Decoding tumor phenotypes for ALK, ROS1, and RET fusions in lung adenocarcinoma using a radiomics approach. Med (United States). 2015;94(41):1–8.
16. Radiol CJ. Preliminary value of CT radiomics in predicting anaplastic lymphoma kinase fusion gene expression in lung adenocarcinoma. Chin J Radiol. 2015;49(2):89–94.
17. Grove O, Berglund AE, Schabath MB, et al. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma.

Muñoz-Barrutia A, editor. PLoS One. Public Library of Science; 2015;10(3):e0118261. https://doi.org/10.1371/journal.pone.0118261. Accessed 10 Mar 2020.
18. Fried DV, Tucker SL, Zhou S, et al. Prognostic value and reproducibility of pretreatment ct texture features in stage III non-small cell lung cancer. Int J Radiat Oncol Biol Phys. Elsevier Inc. 2014;90(4):834–42.
19. Song J, Shi J, Dong D, et al. A new approach to predict progression-free survival in stage IV EGFR-mutant NSCLC patients with EGFR-TKI therapy. Clin Cancer Res. American Association for Cancer Research Inc.; 2018;24(15):3583–3592. https://pubmed.ncbi.nlm.nih.gov/29563137/. Accessed 10 July 2020.
20. Yang X, Dong X, Wang J, et al. 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(11):1156–64.
21. Velazquez ER, Parmar C, Liu Y, et al. Somatic mutations drive distinct imaging phenotypes in lung cancer 2018;77(14):3922–3930.
22. Tang C, Hobbs B, Amer A, et al. Development of an immune-pathology informed radiomics model for non-small cell lung cancer. Sci Rep Springer US. 2018;8(1):1–9.
23. Yoon J, Suh YJ, Han K, et al. Utility of CT radiomics for prediction of PD-L1 expression in advanced lung adenocarcinomas. Thorac Cancer. 2020:1–12.
24. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA J Am Med Assoc. 2016;316(22):2402–10.
25. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. Nature Publishing Group. 2017;542(7639):115–8. https://doi.org/10.1038/nature21056.
26. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med. Springer US. 2019;25(6):954–61. https://doi.org/10.1038/s41591-019-0447-x.
27. Wang S, Shi J, Ye Z, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. Eur Respir J. 2019;53(3).
28. Zhao W, Yang J, Ni B, et al. Toward automatic prediction of EGFR mutation status in pulmonary adenocarcinoma with 3D deep learning. Cancer Med Blackwell Publishing Ltd. 2019;8(7):3532–43.
29. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). Springer Verlag; 2016;9908 LNCS:630–645. http://arxiv.org/abs/1603.05027. Accessed 10 Apr 2020.
30. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. IEEE Computer Society; 2016. p. 770–778.
31. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. 32nd Int. Conf. Mach. Learn. ICML 2015. 2015.
32. Srivastava N, Hinton G, Krizhevsky A, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 2014.
33. Nesterov Y. Gradient methods for minimizing composite functions. Math Program, Ser B. 2013;140:125–61.
34. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. Proc IEEE Int Conf Comput Vis. 2015;2015 Inter:1026–34.
35. Conklin CMJ, Craddock KJ, Have C, Laskin J, Couture C, Ionescu DN. Immunohistochemistry is a reliable screening tool for identification of ALK rearrangement in non-small-cell lung carcinoma and is antibody dependent. J Thorac Oncol. Lippincott Williams and Wilkins. 2013;8(1):45–51.
36. (16) (PDF) Diagnostic value of a novel fully automated immunochemistry assay for detection of ALK rearrangement in primary lung adenocarcinoma. https://www.researchgate.net/publication/254261824_Diagnostic_value_of_a_novel_fully_automated_immunochemistry_assay_for_detection_of_ALK_rearrangement_in_primary_lung_adenocarcinoma. Accessed 11 Apr 2020.
37. Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol. 2016;34(18):2157–64.
38. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. Proc IEEE Int Conf Comput Vis. Institute of Electrical and Electronics Engineers Inc.; 2017. p. 618–626.
39. Zhou Q, Zhang X-C, Chen Z-H, et al. Relative abundance of EGFR mutations predicts benefit from gefitinib treatment for advanced non-small-cell lung cancer. J Clin Oncol. American Society of Clinical Oncology; 2011;29(24):3316–3321. http://www.ncbi.nlm.nih.gov/pubmed/21788562. Accessed July 11, 2020.

Affiliations

Zhengbo Song[1] · Tianchi Liu[2,3] · Lei Shi[2,3] · Zongyang Yu[4] · Qing Shen[2,3] · Mengdi Xu[2,3] · Zhangzhou Huang[5] · Zhijian Cai[6] · Wenxian Wang[1] · Chunwei Xu[7] · Jingjing Sun[8] · Ming Chen[9]

[1] Department of Clinical Trial, Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou 310022, Zhejiang, China

[2] Shanghai Key Laboratory of Artificial Intelligence for Medical Image and Knowledge Graph, Shanghai 200336, China

[3] YITU AI Research Institute for Healthcare, Hangzhou 310000, Zhejiang, China

[4] Department of Medical Oncology, 900th Hospital, Fuzhou 350000, Fujian, China

[5] Department of Medical Oncology, Fujian Cancer Hospital, Fuzhou 350001, China

[6] Institute of Immunology, Zhejiang University School of Medicine, Hangzhou 310009, Zhejiang, China

[7] Department of Pathology, Fujian Cancer Hospital, Fuzhou 350001, China

[8] Department of Radiology, Cancer Hospital of the University of Chinese Academy of Sciences(Zhejiang Cancer Hospital), Hangzhou 310022, Zhejiang, China

[9] Department of Radiotherapy, Cancer Hospital of the University of Chinese Academy of Sciences(Zhejiang Cancer Hospital), Hangzhou 310022, Zhejiang, China