BJR

https://doi.org/10.1259/bjr.20180334

Received: 10 April 2018 | Revised: 09 July 2018 | Accepted: 26 July 2018

Cite this article as:
Xiong J-F, Jia T-Y, Li X-Y, Yu W, Xu Z-Y, Cai X-W, et al. Identifying epidermal growth factor receptor mutation status in patients with lung adenocarcinoma by three-dimensional convolutional neural networks. *Br J Radiol* 2018; **91**: 20180334.

FULL PAPER

Identifying epidermal growth factor receptor mutation status in patients with lung adenocarcinoma by three-dimensional convolutional neural networks

[1]JUN-FENG XIONG, Ph.D., [2]TIAN-YING JIA, MD., [2]XIAO-YANG LI, Ph.D., [2]WEN YU, Ph.D., [2]ZHI-YONG XU, Ph.D., [2]XU-WEI CAI, Ph.D., [1]LING FU, MS, [3]JIE ZHANG, MS, [1]BIN-JIE QIN, Ph.D., [2]XIAO-LONG FU, MD and [1]JUN ZHAO, Ph.D.

[1]Department of Biomedical Engineering, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
[2]Department of Radiotherapy, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China
[3]Department of Pathology, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China

Address correspondence to:
Professor Jun Zhao
E-mail: *junzhao@sjtu.edu.cn*

E-mail: *xlfu1964@hotmail.com*

The authors Jun-Feng Xiong and Tian-Ying Jia contributed equally to the work.

Objective: Genetic phenotype plays a central role in making treatment decisions of lung adenocarcinoma, especially the tyrosine-kinase-inhibitors-sensitive mutations of the epidermal growth factor receptor (EGFR) gene. We constructed three-dimensional convolutional neural networks (CNN) to analyze underlying patterns in CT images that could indicate that EGFR gene mutation status but are invisible to human eyes.
Methods: From 2012 to 2015, 503 Chinese patients with lung adenocarcinoma that had underwent surgery were included. Pathological types and EGFR mutation status were tested from surgical resections. EGFR mutations (exon 19 deletion or exon 21 L858R) were found in 215/345 (62.3%) and 91/158 (57.6%) patients in the training and independent validation set, respectively. CT images were taken before any invasive operation. The patients were randomly chosen to train the CNNs or validate the CNNs' performance. The performance was quantified using area under receiver operating characteristic curve (AUC), sensitivity, specificity, and accuracy.
Results: The CNNs showed an AUC of 0.776 (range: 0.702–0.849, $p < 0.0001$) in the independent validation set and a fusion model of CNNs and clinical features (sex and smoking history) showed an AUC of 0.838 (range: 0.778–0.899, $p < 0.0001$), accuracy of 77.2%, sensitivity of 75.8% and specificity of 79.1% at the best diagnostic decision point.
Conclusion: The CNN exhibits potential ability to identify EGFR mutation status in patients with lung adenocarcinoma which might help make clinical decisions.
Advances in knowledge: The CNN showed some diagnostic power and its performance could be further improved by increasing the training set, optimizing the network structure and training strategy. Medical image based CNN has the potential to reflect spatial heterogeneity.

INTRODUCTION

The mutation status of the epidermal growth factor receptor (EGFR) gene is crucial for tailoring treatments for advanced lung adenocarcinoma, since the tyrosine kinase inhibitors (TKI) could improve the survival of patients who have sensitive mutations, such as exon 19 deletions (19 del) and exon 21 amino acid substitution at position 858 in the EGFR gene, from a leucine (L) to an arginine (R) (21L858R).[1–5] Currently, EGFR mutation tests requires invasive operations such as punctuation or resection, which could be difficult or impossible for some patients. Besides, invasive operations may not offer information of occurrence of new mutations (especially TKI resistive mutation) in time, this is because invasive operation may not be repeated frequently. As a result, a non-invasive test method for EGFR mutation is in need.

Two non-invasive techniques are under development for testing the EGFR gene mutation status, liquid biopsy and radiomics. Liquid biopsy tests circulating tumor DNA (ctDNA) or circulating tumor cells (CTC) from blood and other body liquid samples.

Much effort was contributed to the testing of ctDNA and CTC to detect EGFR gene mutations (exon 19 deletions, 21L858R) and some experiments showed remarkable results.[6–8] One report of droplet digital PCR reached a level of sensitivity and specificity, compared with the matched tumor tissues tested by amplification refractory mutation system, of 81.8% (59.7–94.81)] and 98.4%CI (91.6–99.96)], respectively,[9] a higher level compared with the matched tumor tissues tested by amplification refractory mutation system. However, liquid biopsy lacks the ability to show spatial information and is unable to locate the position of mutation.

Another approach is based on medical images. Radiomics was inspired by the idea that medical images contain information that reflects underlying pathophysiology and that these relationships can be revealed by quantitative image analyses.[10] By studying CT, MRI, or positron emission tomography-CT (PET-CT) images, hundreds of quantitative image features could be extracted and used for finding information of tumor phenotype, heterogeneity, sensitivity to treatment etc. In a study analyzing CT images of non-small cell lung cancer, researchers found that a prognostic radiomics signature which captures intratumor heterogeneity was associated with underlying gene-expression patterns[11] In two recent studies, quantitative image features extracted from CT images were found to be associated with the EGFR mutation status in lung adenocarcinomas.[12,13] These studies showed that, an underlying difference might exist in terms of radiomics between the two EGFR mutation statuses (wild type *vs* TKI sensitive mutation).

In this study, we used convolutional neural networks (CNN) as the analyzing tool for EGFR mutation status using non-contrast enhanced CT images.

In the recent years, CNNs has shown superior performance in pattern recognition problems to other algorithms in many worldwide computer vision-competitions. CNNs have been used in medical image research as well.[14] In a study of skin cancer diagnosis, CNNs achieved performance comparable to experienced doctors.[15] However, differences involving dimension, range of gray levels, size of data set etc. lies between medical and natural images. Much work is needed to gain the experience for applying CNNs in medical image analysis. To our knowledge, no research was found by us to be using CNNs for EGFR mutation status test.

In this study, we adopted a CNN that was developed in natural image problems to the analysis of three-dimensional CT images. The CNN was trained to identify patients with EGFR mutations from wild type and tested using an independent patient set.

METHODS AND MATERIALS

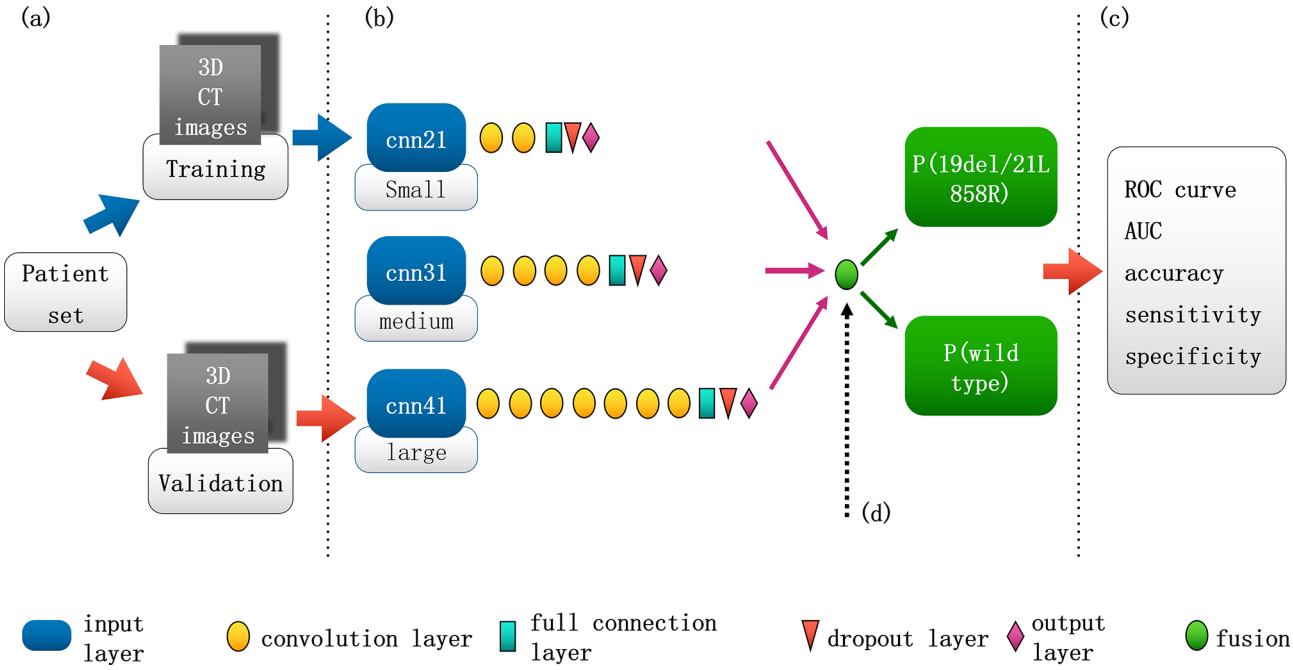
The workflow

The workflow of our study is shown in Figure 1. It contains the following major steps: (1) patient dataset construction (including the selection of patients and the gathering of clinical information); (2) The region of interest (ROI) definition and preprocessing of CT images; (3) the training of the CNN; (4) performance validation. Further details are described in the following sections.

Patients

Patients with Stage I–IV primary lung adenocarcinoma who were treated with surgery-based strategy at Shanghai Chest Hospital between 2012 and 2015 were enrolled in this study.

Figure 1. The workflow of our method. (a) The patient set was divided into training and validation sets. (b) The training set was studied by the multi CNN. (c) Performance of the CNN was test on the validation set. (d) In the fusion model, clinical features were fused to the CNN. AUC, area under ROC curve; CNN, convolutional neural network; ROC, receiver operating characteristic.



Patients with multiple lesions, or the only lesion being smaller than 8mm, or found having a ground glass component was excluded. Patients with tuberculosis or previous tumor history on other sites were also ruled out.

The patients were divided into training and validation groups using random stratified sampling considering: EGFR mutation status (19del or 21L858R *vs* wild type), sex, TNM staging (I~IIvs III~IV), and smoking history. The *rand()* function of MATLAB® was used to select the patients for different groups randomly.

In order to gain statistical significance for an area under receiver operating characteristic curve (AUC) larger than 0.72 (slightly higher than Ying Liu et al[13]) in a sample size of 503, 157 samples were needed for performance validation. Stratified sampling was applied to randomly select 158 samples for validation and 345 samples for training. The criteria for the stratified sampling included sex, smoking history, EGFR mutation status, and TNM staging. The TNM staging was based on the seventh edition of International Association for the Study of Lung Cancer non-small cell lung cancer TNM staging. The Stage IV was based on the latest follow-up.

From 2012 to 2015, 503 patients were included in our study. Among them, 254(50.5%) were male, 78(15.5%) were smokers, 307(61.0%) had EGFR mutation (19del or 21L858R). Most patients were in Stages I (237, 47.1%) and III (159, 31.6%). Tumor size (diameter) varied from 0.9to10.5cm (median, 2.5cm). Performance status score was 1 point for all the patients. The difference of clinical features of training and validation groups had no statistical significance, as is shown in Table 1.

The test for EGFR gene mutation status

The EGFR test was based on surgical resections via fluorescence PCR (ARMS) using Stratagene Mx3000PTM (Agilent) and the Human EGFR Gene Mutation Detection Kit (Amoy Diagnostics Co., Ltd) in the lab of Shanghai Chest Hospital.

Statistical analysis

The statistical analysis consisted of three parts.

First, we compared the clinical features of the training and validation groups to find out if there was any difference that reached statistical significance. Fisher's exact test was used to compare

Table 1. Comparison of clinical features between training and validation groups

	Training group (*n*=345)	Validation group (*n*=158)	*p*-value
Sex (*n*/%)			0.443
Male	170(49.3)	84(53.2)	Fisher's exact test
Female	175(50.7)	74(46.8)	Two-sided Pr*p* 0.443
Smoking history	1.00		
Yes	55(15.9)	25(15.8)	Fisher's exact test
No	290(84.1)	133(84.2)	Two-sided Pr 1.00
EGFR mutation status	0.496		
Wild type	130(37.7)	67(42.4)	X^2
19del	103(29.8)	40(25.3)	
21L858R	112(32.5)	51(32.3)	
TNM staging			0.811
I	165(47.8)	74(46.8)	Wilcoxon two-sample test Normal approximation (Z 0.239) One-sided Pr>Z 0.406 Two-sided Pr > |Z| 0.811
II	47(13.6)	25(15.8)	
III	113(32.8)	45(28.5)	
IV[a]	20(5.8)	14(8.9)	
Age			0.797
Median	60	61	Wilcoxon two-sample test Normal approximation (Z 0.257) One-sided Pr>Z 0.399 Two-sided Pr > |Z| 0.797
Min	28	32	
Max	82	79	
Tumor size (cm)			0.462
Median	2.6	2.5	Wilcoxon two-sample test Normal approximation (Z −1.31) One-sided Pr <Z 0.095 Two-sided Pr > |Z| 0.191
Min	0.9	1	
Max	10.5	7	

EGFR, epidermal growth factor receptor;
[a]In the 34 Stage IV patients, 32 were found with pleural nodules, parietal pleura, lung, pericardium, or diaphragm metastases during surgery, the other 2 were found with brain or bone metastases by PET-CT and had wedge cutting.

sex and smoking history, Wilcoxon two sample test was used to compare age and TNM stages, and X^2 was used to compare EGFR mutations status.

Second, we compared the clinical features between the EGFR wild type group and the mutant groups to select the clinical features that had statistically significant difference between the groups. Such features were assumed to have discriminative power and would be used to build the classifier. $p<0.05$ was considered to be statistically significant.

Third, the receiver operating characteristic (ROC) curve and AUC was used to validate the performance of the models and the comparison of ROC curves were pairwise compared by using the DeLong test[16] and the corresponding AUC difference, standard error, 95%CI, z-statistic, and p-value were calculated.

The acquisition of CT images

The CT scans were taken about one week before surgery, most of the scans were contrast enhanced but 20–30% were not, these scans were used to build the classification model.

The scans were taken with a voltage between 120 and 140 kV, a current of 170 mA, scan layer thickness of 5mm, and a spatial resolution of about 1mm using the Brilliance iCT and Brilliance 64 CT from PHILIPS.

In three-dimensional analysis, the spatial resolution should be identical in all dimensions. However, the resolution in the x- and y-axis varied from 0.578to0.934mm/pixel and the resolution in $z$axis was 5mm/pixel. Linear interpolation was therefore applied and the resulting resolution was 2×2×2mm. The Hounsfield unit (HU) of each voxel was linearly uniformed to [0, 1].

Definition of ROI

Considering the input of the CNN was cubic which encloses the tumor area regardless of its actual shape, a precise definition of the tumor contour was not necessary. In other words, a roughly defined region could work just as well as a region that closely conforms to the tumor. On the Pinnacle 2 platform from Varian®, the ROI was defined by radio-oncologists in the lung window (−300~1301HU).

The CNN model construction

Our classification method consisted of a multilevel CNN (denoted as M_{CNN}) containing three CNNs denoted as M_{21}, M_{31}, and M_{41} with input patches of 21×21×21 voxels (42×42×42mm), 31×31×31 voxels (62×62×62mm), and 41×41×41(82×82×82mm) voxels, respectively. These CNNs were tailored to match different sized tumors. The input size of 42×42×42mm, 62×62×62mm, and 82×82×82mm covers 58%, 92%, and 99% of all the tumors in the data set, respectively. The tumor size varied in a large range of 8–103mm among the patients in our data set. The required input image size of traditional CNNs needs to be same. Hence, the image lose some tumor information or contain too much noise if we only use one CNN with only one specific size. Specifically, on one hand, the image cannot contain the whole tumor (the size>21 voxels) if we only used the CNN with input size of 21×21×21 voxels. The image may include too much background noise for the small tumor (size<21 voxels) if a CNN with input size of 41×41 × 41 voxels were to be used. To alleviate this problem, we used three CNNs with different input sizes to construct our multilevel CNN model. The input patches (ROI of the CT images) flow from left to right and the CNN would give probabilities of mutation (19del or 21L858R) or wild type. The three networks were combined to obtain a multilevel CNN based model denoted as M_{CNN}, as shown below:

$$M_{CNN} = w_{21}M_{21} + w_{31}M_{31} + w_{41}M_{41} \quad (1)$$

where w_{21}, w_{31}, and w_{41} denotes each network' contribution to the multilevel CNN model. These weight parameters were optimized by using a grid search method from the range 0.0–1.0 with a step size of 0.1.

Detailed description can be found in the supplementary material (The details of multilevel CNN). Smaller input patches corresponded with fewer hidden layers since they contained less information to analyze.

Clinical feature selection

According to the reports in,[12,13] sex and smoking history have a high correlation with the EGFR mutation status. Moreover, we assessed the classification potential of each clinical feature by analyzing their difference between the EGFR-wild subset and EGFR-mutant subset before we constructed a clinical feature based classifier.

The statistical analysis methods and results are shown in Table 2. Among these features, only sex and smoking history showed significant difference and would be used to build the clinical feature-based model.

We also built a simple clinical feature based model, which was based on sex and smoking history. We assigned scores to each sample according to the clinical feature, female non-smokers given 1.00, the female smokers and male non-smokers were given 0.50, and male smokers were given 0.00. The scoring for this clinical feature based model (denoted as $M_{Clinical}$) is summarized in the following equation:

$$M_{Clinical} = \begin{cases} 1.00 & \text{female \& non} - \text{smoking} \\ 0.50 & \text{either feamle or non} - \text{smoking} \\ 0.00 & \text{male \& smoking} \end{cases} \quad (2)$$

Another fusion Model of multilevel CNN and clinical feature was assessed to see if clinical features could help improve the performance of the CNNs. The fusion model (denoted as M_{Fusion}) is defined as a weighed summation of M_{CNN} and $M_{Clinical}$:

$$M_{Fusion} = w_{CNN}M_{CNN} + w_{Clinical}M_{Clinical} \quad (3)$$

where w_{CNN} and $w_{Clinical}$ denotes how much each model contributes to the fusion model. The weights were selected to optimize the performance of M_{Fusion}. They were optimized from 0.00to1.00 with step size 0.1.

Table 2. Comparison of clinical features between EGFR mutant and wild type

	Wild type (*n*=197)	Mutant (*n*=306)	*p*
Sex (n/%)			<.0001
Male	130(66.0)	124(40.5)	Fisher's exact test
Female	67(34.0)	182(59.5)	Two-sided Pr 2.32E-08
Smoking history			<0.0001
Yes	52(26.4)	28(9.2)	Fisher's exact test
No	145(73.6)	278(90.8)	Two-sided Pr 3.95E-07
TNM staging			0.664
I	90(45.7)	149(48.7)	Wilcoxon two-sample test Normal approximation (Z 0.435) One-sided Pr>Z 0.332 Two-sided Pr > |Z| 0.664
II	30(15.2)	42(13.7)	
III	65(33.0)	93(30.4)	
IV	12(6.1)	22 (7.2)	
Age			0.355
Median	61	61	Wilcoxon two-sample test Normal approximation (Z 0.435) One-sided Pr >Z 0.332 Two-sided Pr > |Z| 0.664
Min	32	28	
Max	82	82	

EGFR, epidermal growth factor receptor;

The performance was validated on the validation set using AUC, accuracy, sensitivity and specificity.

RESULTS

Model performance

The weight parameters, w_{21}, w_{31}, and $w_{41,}$ in Equation (1) were 0.3, 0.4, 0.3, respectively. The weight parameters, w_{CNN} and $w_{Clinical}$, in Equation (3) were 0.8 and 0.2, respectively.

Tested by the independent validation set, CNN resulted in an AUC of 0.776, accuracy of 0.715, sensitivity of 57.1%, and specificity of 91.0%, which is much better than using clinical data alone. Combining CNN and clinical features, the fusion model offered a performance upgrade. Detailed performance assessment is shown in Table 3 and Figure 2.

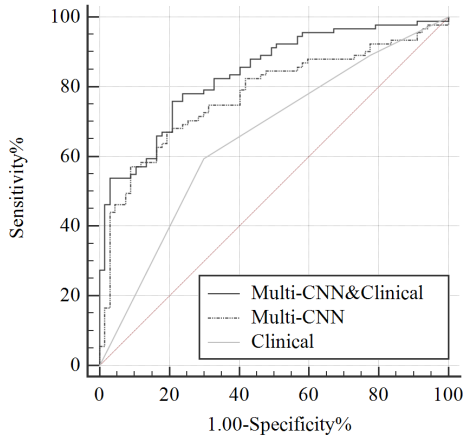
The fusion model gave the best performance with an AUC of 0.838 (accuracy 0.776), better than clinical model (0.652) and multi-CNN (0.778). Its sensitivity and specificity are 75.8and79.1% at the best diagnostic decision point. Table 3 shows the detailed performance of the multilevel CNN, clinical feature based, and fusion models. Figure 2 shows the ROC curves of the

Table 3. Performance of CNN, clinical feature-based model, and fusion model

Models	Multi-CNN&clinical features	Multi-CNN	Clinical features
AUC	0.838	0.776	0.654
SE	0.031	0.037	0.041
95%CI	0.778to0.899	0.702to0.849	0.575to0.734
p	<0.0001	<0.0001	<0.0001
Accuracy	77.2%	71.5%	63.9%
Sensitivity	75.8%	57.1%	59.3%
Specificity	79.1%	91.0%	70.1%
Comparison	**Multi-CNN&clinical features*vs* multi-CNN**	**Multi-CNN&clinical features*vs*clinical features**	**Multi-CNN*vs*clinical features**
AUC difference	0.063	0.184	0.121
SE	0.025	0.0418	0.0602
95%CI	0.014to0.111	0.102to0.266	0.003to0.239
Z statistic	2.504	4.393	2.016
p- value	0.0123	<0.0001	0.044

AUC, area under receiver operating characteristic curve; CI, confidence interval; CNN, convolutional neural network; SE, standard error.

Figure 2. The ROC curves of the clinical feature based model (gray solid), multi-CNN (black dotted), and the fusion model (black solid). CNN, convolutional neural network; ROC, receiver operating characteristic curve.



clinical feature-based model (orange), multi-CNN (green), and the fusion model (blue).

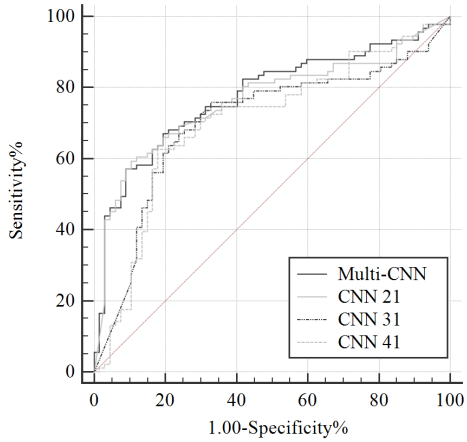
We also assessed how much benefit we could get by using three CNNs. Figure 3 shows the ROC curves of CNN 21, CNN 31, and CNN 41 and the corresponding performance data are listed in Supplemental Table S1. The test result confirmed our assumption that multiple CNNs that are adapted to the size of the tumor could improve performance.

DISCUSSION

Performance regards

The multi-CNN showed diagnostic power but its current sensitivity and specificity is not good enough for clinical use of the multi-CNN.

Figure 3. The ROC curves of the multi-CNN (black solid), CNN-21 (gray solid), CNN-31 (black dotted), CNN 41 (gray dotted). CNN, convolutional neural network; ROC, receiver operating characteristic.



Three former reports used radiomics to identify EGFR mutations. Velazquez E R's study[13] (n=258) used independent validation and reached an AUC of 0.67. Ying Liu's study[12] (n=298) did not use independent validation and reached an AUC of 0.709 in the training set. The study[17] using CNNs (each CNN has the same convolution layers) with a smaller sample size (n=405) had an AUC of 0.767 (p<0.001). In this study, the best model reached an AUC of 0.838 and used an independent set for performance validation. The improvement of performance gave us hope that CNN and other medical image-based analysis could achieve better results and tackle more difficult tasks.

CNN vs liquid biopsy

CNN and liquid biopsy are two non-invasive testing methods for the test of EGFR gene. To our knowledge, ddPCR had reached a sensitivity of 81.8% and specificity of 98.4%,[9] which is obviously superior to the CNN-clinical model in our study. However, liquid biopsy lacks the ability to show spatial information and could not tell the position of the mutation. In medical image analysis, using CNN or other modeling algorithms, the spatial heterogeneity was reserved and could even be quantified. Imaging analysis has the potential to suggest which lesion needs close follow-up or immediate treatment.

Clinical significance

This study showed that it is feasible to use CT images and CNNs to identify EGFR mutations (19del/21L858R) in lung adenocarcinoma. This medical image-based method has many advantages over other testing methods. First, it had the potential to support treatment decisions when surgery resection and puncture biopsy are difficult or not available since it is non-invasive. Second, it could be applied during the entire treatment as a monitoring tool for EGFR mutation status since CT images are repeatedly captured in clinical practice. Third, this CNN tool analyzed the entire tumor instead of just tissue samples and might have the ability to overcome the problem of heterogeneity. Fourth, using a CNN is fast, convenient and low cost.

As the field of radiomics is expanding new methods of extracting information from medical images, these methods benefit the diagnosis, treatment, and prognosis of diseases. Tumor phenotypes, including aggressiveness, driver gene, heterogeneity, response to treatment, and recurrence, could be identified and defined more precisely with the help of quantified analysis of medical images.

Impact on future studies

A unique advantage of CNN compared to the radiomics method is its loose requirement of ROI delineation, because the data for analysis consist of a cube that bounds the region of the tumor.

Apart from the gene mutation status, other characteristics of the disease are also probable to be revealed by the CNN analysis of medical images (CT, MR or PET/CT). These include tumor proliferative activity, hypoxia status and the surrounding tumor environment. We believe the development of CNN could vigorously advance the development and clinical application of individualized medicine.

This study also sheds light on the use of CNNs to analyze medical images for gene tests of malignant tumors. Compared to other radiomics approaches that uses pre-defined features to capture the texture of medical images, CNN dynamically extracts image features and can change its "definition" of the features in the training process to improve its performance. Besides, the other radiomics approaches needs an accurate determination of target volume to compute the pre-defined features (especially the shape- and size-based features) and some features may be sensitive to the rotation or translation of the target volume. The CNN only need a rough target volume to crop a cube patch and each patch is augmented by rotations and translation, which elicits rotational and translational invariance.

Several factors can influence the performance of the CNN and improvements can be made in the following aspects. First, CNN is a tool for big data analysis, which means larger data set would bring better performance. It is necessary to use a larger data set for performance improvement.

Second, the quality of CT images could influence the performance of CNN in several aspects. Differences in reconstruction algorithms and spatial resolutions caused by the use of different CT machines from different manufacturers could influence the HU of the voxels. In this study, the CT images were taken by two machines from Siemens and Philips, we did not rule out either of them because it is common to use different CT scanners in clinical practice. This influence could be reduced since the difference among the pixels was what matters to CNNs, rather than the actual gray levels. As long as the HU reflects the density difference among different pixels, this difference could be learned by a CNN. As for the reduction of performance caused by different spatial resolutions, the unification of the resolution was necessary for the analysis but introduced errors through interpolation. Additionally, the 5-mm thickness of the CT scan layers could also reduce the CT images' ability to reflect the characteristics of the tumors. According to a study done by Zhao BS,[18] quantitative image features are more stable with a scan layer thickness of 1mm rather than 5mm. Although this study assessed only radiomics features, their findings might be true for the features learned by CNN as well. In our study, the CT scan layer thickness was 5mm, thus making the features unstable. Besides, for tumors smaller than 20mm, there may only be two to three layers, which could make the features insufficient for describing tumor characteristics.

CONCLUSION

Our CNN classifier for EGFR mutation exhibited a favorable AUC. We also showed that the sensitivity and specificity could be further improved by adding clinical features.

Our research showed that genetic differences among tumors could be identified by CNNs. For patients where puncture biopsy of the primary tumor is dangerous, this medical image-based approach might provide an alternative. We believe the CT image-based approach for EGFR mutation test has room for improvement and could be useful in clinics in the future.

FUNDING

This study is partly supported by National Key Research and development Program of China (2016YFC0104608 and 2016YFC0905500), the National Natural Science Foundation of China (No. 81371634), and Shanghai Jiao Tong University Medical Engineering Cross Research Funds (YG2017ZD10, YG2014ZD05, and YG2013MS30).

ETHICS APPROVAL

The standards in the Code of Ethics of the World Medical Association (Declaration of Helsinki) have been adhered. The research procedures were approved by the responsible committee on human or animal experimentation (institutional or regional). Ethical approval for this investigation was obtained from the Research Ethics Committee, Shanghai Jiao Tong University, School of Medicine. Informed consent was obtained for experimentation with human subjects.

REFERENCES

1. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004; **350**: 2129–39. doi: https://doi.org/10.1056/NEJMoa040938
2. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004; **304**: 1497–500. doi: https://doi.org/10.1126/science.1099314
3. Fukuoka M, Yano S, Giaccone G, Tamura T, Nakagawa K, Douillard JY, et al. Multi-institutional randomizedphase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected]. *J Clin Oncol* 2003; **21**: 2237–46. doi: https://doi.org/10.1200/JCO.2003.10.038
4. Kris MG, Natale RB, Herbst RS, Lynch TJ, Prager D, Belani CP, et al. Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *JAMA* 2003; **290**: 2149–58. doi: https://doi.org/10.1001/jama.290.16.2149
5. Mitsudomi T, Yatabe Y. Mutations of the epidermal growth factor receptor gene and related genes as determinants of epidermal growth factor receptor tyrosine kinase inhibitors sensitivity in lung cancer. *Cancer Sci* 2007; **98**: 1817–24. doi: https://doi.org/10.1111/j.1349-7006.2007.00607.x
6. Maheswaran S, Sequist LV, Nagrath S, Ulkus L, Brannigan B, Collura CV, et al. Detection of mutations in EGFR in circulating lung-cancer cells. *N Engl J Med* 2008; **359**: 366–77. doi: https://doi.org/10.1056/NEJMoa0800668
7. Sundaresan TK, Sequist LV, Heymach JV, Riely GJ, Jänne PA, Koch WH, et al. Detection of T790M, the acquired resistance EGFR mutation, by tumor biopsy versus noninvasive blood-based analyses. *Clin Cancer Res* 2016; **22**: 1103–10. doi: https://doi.org/10.1158/1078-0432.CCR-15-1031

8. Ran R, Li L, Wang M, Wang S, Zheng Z, Lin PP. Determination of EGFR mutations in single cells microdissected from enriched lung tumor cells in peripheral blood. *Anal Bioanal Chem* 2013; **405**: 7377–82. doi: https://doi.org/10.1007/s00216-013-7156-y
9. Zhu G, Ye X, Dong Z, Lu YC, Sun Y, Liu Y, et al. Highly sensitive droplet digital PCR method for detection of EGFR-activating mutations in plasma cell–free DNA from patients with advanced non–small cell lung cancer. *J Mol Diagn* 2015; **17**: 265–72. doi: https://doi.org/10.1016/j.jmoldx.2015.01.004
10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016; **278**: 563-77. doi: https://doi.org/10.1148/radiol.2015151169
11. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; **5**: 4006. doi: https://doi.org/10.1038/ncomms5006
12. Rios Velazquez E, Liu Y, Parmar C, Narayan V, Gillies R, Aerts H. MO-DE-207B-08: radiomic CT features complement semantic annotations to Predict EGFR mutations in lung adenocarcinomas. *Med Phys* 2016; **43**–3706–06. doi: https://doi.org/10.1118/1.4957257
13. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. *Clin Lung Cancer* 2016; **17**: 441. doi: https://doi.org/10.1016/j.cllc.2016.02.001
14. Dou Q, Chen H, Yu L, Qin J, Heng PA, Qi D, Hao C, Jing Q. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. *IEEE Trans Biomed Eng* 2017; **64**: 1558. doi: https://doi.org/10.1109/TBME.2016.2613502
15. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**: 115–8. doi: https://doi.org/10.1038/nature21056
16. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**: 837–45. doi: https://doi.org/10.2307/2531595
17. Junfeng X, Tianying J, Xiaoyang L, Zhiyong X, Xuwei C, Jie Z. P3. 13-022 3D CNNs for recognition of epidermal growth factor receptor mutation status in patients with lung adenocarcinoma. *Journal of Thoracic Oncology* 2017; **11**: S2324. doi: https://doi.org/10.1016/j.jtho.2017.09.1757
18. Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* 2016; **6**: 23428. doi: https://doi.org/10.1038/srep23428