COPD identification and grading based on deep learning of lung parenchyma and bronchial wall in chest CT images

Zhang, Lin; Jiang, Beibei; Wisselink, Hendrik Joost; Vliegenthart, Rozemarijn; Xie, Xueqian

Published in:
British journal of radiology

DOI:
10.1259/bjr.20210637

BJR

https://doi.org/10.1259/bjr.20210637

Received: 24 May 2021 | Revised: 20 January 2022 | Accepted: 01 February 2022 | Published online: 16 February 2022

Cite this article as:
Zhang L, Jiang B, Wisselink HJ, Vliegenthart R, Xie X. COPD identification and grading based on deep learning of lung parenchyma and bronchial wall in chest CT images. *Br J Radiol* (2022) 10.1259/bjr.20210637.

FULL PAPER

COPD identification and grading based on deep learning of lung parenchyma and bronchial wall in chest CT images

[1]LIN ZHANG, MD, PhD, [1]BEIBEI JIANG, MD, [2]HENDRIK JOOST WISSELINK, MSc, [2]ROZEMARIJN VLIEGENTHART, MD, PhD and [1]XUEQIAN XIE, MD, PhD

[1]Radiology Department, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[2]Radiology Department, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Address correspondence to: Dr Xueqian Xie
E-mail: *xiexueqian@hotmail.com*

Objective Chest CT can display the main pathogenic factors of chronic obstructive pulmonary disease (COPD), emphysema and airway wall remodeling. This study aims to establish deep convolutional neural network (CNN) models using these two imaging markers to diagnose and grade COPD.

Methods Subjects who underwent chest CT and pulmonary function test (PFT) from one hospital (n = 373) were retrospectively included as the training cohort, and subjects from another hospital (n = 226) were used as the external test cohort. According to the PFT results, all subjects were labeled as Global Initiative for Chronic Obstructive Lung Disease (GOLD) Grade 1, 2, 3, 4 or normal. Two DenseNet-201 CNNs were trained using CT images of lung parenchyma and bronchial wall to generate two corresponding confidence levels to indicate the possibility of COPD, then combined with logistic regression analysis. Quantitative CT was used for comparison.

Results: In the test cohort, CNN achieved an area under the curve of 0.899 (95%CI: 0.853–0.935) to determine the existence of COPD, and an accuracy of 81.7% (76.2–86.7%), which was significantly higher than the accuracy 68.1% (61.6%–74.2%) using quantitative CT method (p < 0.05). For three-way (normal, GOLD 1–2, and GOLD 3–4) and five-way (normal, GOLD 1, 2, 3, and 4) classifications, CNN reached accuracies of 77.4 and 67.9%, respectively.

Conclusion CNN can identify emphysema and airway wall remodeling on CT images to infer lung function and determine the existence and severity of COPD. It provides an alternative way to detect COPD using the extensively available chest CT.

Advances in knowledge CNN can identify the main pathological changes of COPD (emphysema and airway wall remodeling) based on CT images, to infer lung function and determine the existence and severity of COPD. CNN reached an area under the curve of 0.853 to determine the existence of COPD in the external test cohort. The CNN approach provides an alternative and effective way for early detection of COPD using extensively used chest CT, as an important alternative to pulmonary function test.

BACKGROUND

Chronic obstructive pulmonary disease (COPD) is a common disease characterized by persistent respiratory symptoms and airflow limitation caused by abnormal alveoli and airways.[1] COPD is currently the fourth leading cause of mortality worldwide. The diagnosis of COPD depends on respiratory symptoms and pulmonary function test (PFT). However, symptoms associated with airflow limitation often occur after a certain degree of lung injury. Delayed hospital visits and PFT examinations can postpone treatment. Early diagnosis and classification of COPD are conducive to early self-management and effective treatment, to improve the prognosis and reduce the burden of COPD.

The global initiative for chronic obstructive lung disease (GOLD) does not recommend population-level PFT screening for asymptomatic individuals.[1] The main pathogenic factors of COPD, emphysema, and airway wall remodeling are the independent factors leading to airflow obstruction, which can be well displayed on thin-slice chest CT.[2,3] CT is extensively used in screening chest diseases, especially lung cancer,[4,5] and can quantify emphysema and airway wall in patients with COPD.[3,6] Because COPD and lung cancer share similar risk factors, additional detection of COPD will improve the benefits of CT scan.[7–9]

Quantitative CT (QCT) is valuable in the diagnosis of COPD. The percentage of voxels below −950 Hounsfield

units and wall area percentage were associated with forced expiratory volume in the first second as a predicted percentage (FEV_1%pred).[2,10] Xie et al found a positive correlation between the thickness of airway wall and respiratory symptoms.[11] Mets et al reported that CT emphysema, bronchial wall thickness, and air trapping have independent diagnostic values for COPD.[6] The integrated model of these imaging biomarkers improved the diagnostic value of COPD. However, interobserver and inter-technique variation limit the clinical application of QCT.

Recently, the convolutional neural network (CNN) is becoming the mainstream method of computer vision and has achieved remarkable results in medical imaging.[12–14] Due to the heterogeneous pathogenesis of COPD, the diagnosis and grading depend on the severity of emphysema and small airway remodeling and their combined contribution. We hypothesized that CNN could determine the respiratory function from chest CT images, according to the image features of lung parenchyma and airways. Thereafter, we trained and tested the deep CNNs based on CT images of lung parenchyma and bronchial wall to determine the presence of COPD and GOLD staging, using PFT as reference.

METHODS

Patients

We searched the consecutive patients in Hospital-A and Hospital-B as the training and external test cohorts, respectively. The inclusion criteria were as follows: (1) subjects who underwent CT examination in the two hospitals from December 2015 to December 2020; in order to balance the number of cases among groups, normal cases were collected until December 2018, and COPD cases were collected until December 2020. (2) CT images and clinical data were available on picture archiving and communication system (PACS) and electronic medical records (EMRs), respectively; (3) CT examination and PFT were performed within 3 days. For patients with acute COPD exacerbation, PFT was performed on the same day.

The exclusion criteria were as follows: (1) patients with asthma history or clinical manifestations, such as allergy, rhinitis, urticaria, family history of asthma, reversible airflow limitation, and positive PFT bronchial dilation test; (2) history of chest surgery before CT examination; (3) CT image thickness >1 mm; (4) motion artifacts in CT images, such as respiratory motion, postural motion, and heartbeat and gastrointestinal motion, leading to blurred pulmonary blood vessels and bronchial tree structures; (5) difficulty to reconstruct images of lung parenchyma or bronchial wall.

The patient group (forced expiratory volume in 1 s [FEV_1] / forced vital capacity [FVC]<0.70) had COPD symptoms and risk factors (smoking, genetic factors, abnormal lung growth and development, exposure to particles, and infection). The normal group (FEV_1/FVC≥0.70) had no respiratory symptoms or COPD risk factors.

The inclusion and exclusion flowchart is shown in Figure 1. The local Institutional Review Board approved this retrospective study and waived the written informed consent.

CT scanning

Three CT systems in Hospital-A (Revolution and HD750, GE Healthcare; Somatom Force, Siemens) and two systems in

Figure 1. Inclusion and exclusion flow chart. COPD, chronic obstructive pulmonary disease; GOLD, the global initiative for chronic obstructive lung disease.

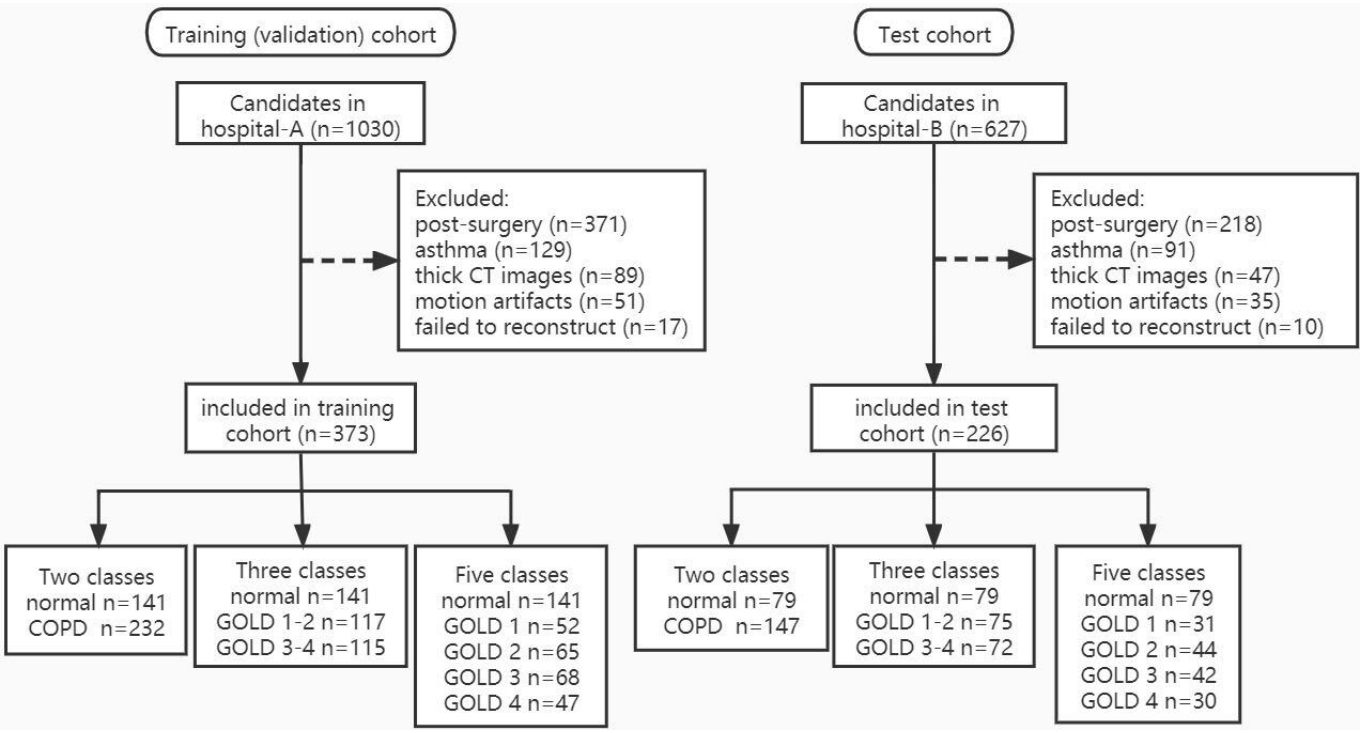
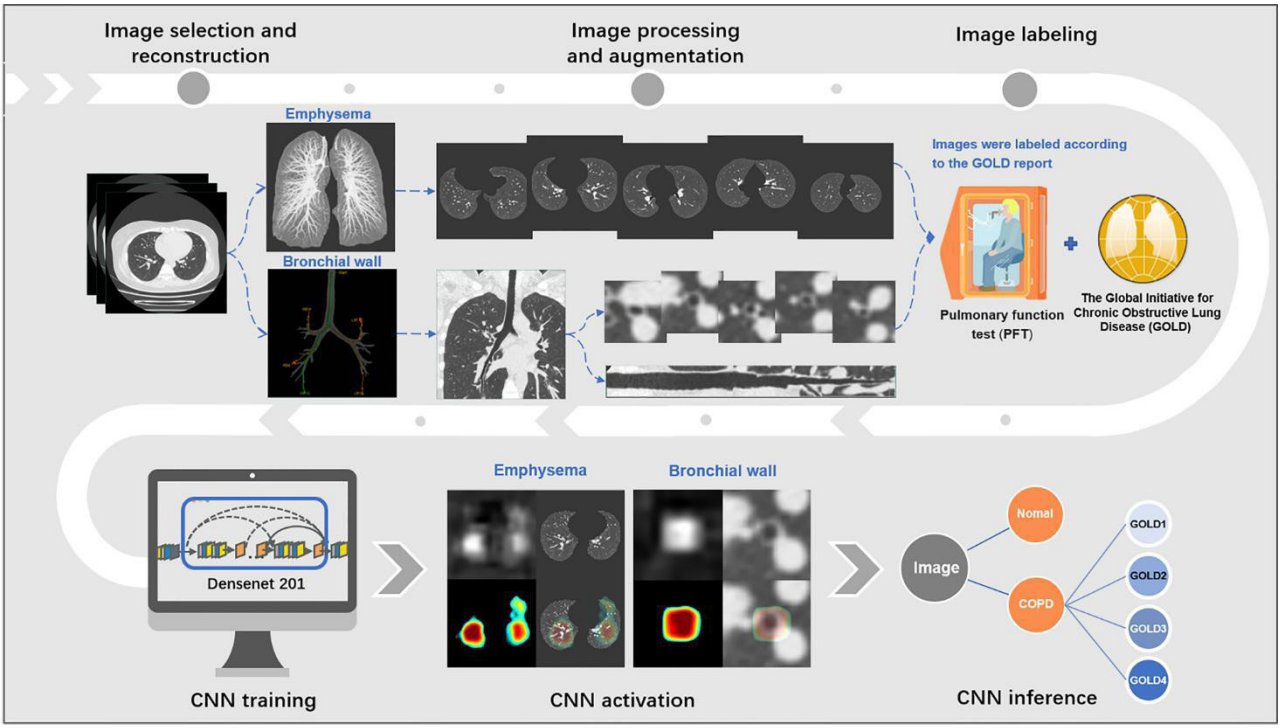


Figure 2. Deep learning workflow diagram. CNN, convolutional neural network.



Hospital-B (VCT, GE Healthcare; Somatom Flash, Siemens) were used for chest thin-slice CT scanning. All patients received only a non-contrast inspiratory CT scan during a single breath-hold. The tube voltage was 120 or 100 kV. The tube current was 50–200 mAs, and the dose modulation was on (Supplementary Material 1 - Table 1). The slice thickness was 0.6–1.0 mm. Because of the influence of image reconstruction on image texture, we used lung kernel and standard kernel to generate images to train CNN. All CT scanners were air-calibrated every day, and water-phantom-calibrated regularly according to the vendor manual.

Pulmonary function test

Post-bronchodilator PFT was performed with spirometer (MasterScreen, Jaeger). The breathing action was to exhale, inhale the most, and then exhale with maximum strength, to obtain the highest airflow. According to GOLD 2021,[1] post-bronchodilator $FEV_1/FVC < 0.70$ indicates the presence of COPD. The severity of COPD is graded as GOLD level 1 (FEV_1%pred≥80%), GOLD level 2 (50% ≤ FEV_1%pred<80%), GOLD level 3 (30% ≤ FEV_1%pred<50%), and GOLD level 4 (FEV_1%pred<30%). Since the diagnosis and severity of COPD depend on GOLD level, this study used the GOLD level as a reference standard to classify patients.

Image preprocessing for deep learning

Figure 2 illustrates the workflow of deep learning in this study. The lung parenchyma and bronchus images of each subject in the training and test cohorts were preprocessed to extract image patches instead of the surrounding irrelevant lung structures. A radiologist with 17 years of experience in thoracic imaging processed the images blinded to the PFT results. An image processing software package (Thoracic VCAR, GE Healthcare) automatically removed the chest wall and mediastinum to generate axial lung parenchyma images.[15] The software also automatically extracted the bronchial tree. Five segment-level bronchial trees were selected for analysis, including right upper lobe apex (RB_1), lower lobe posterior basal segment (RB_{10}), middle lateral segment (RB_4), left upper lobe apex (LB_{1+2}), and lower lobe posterior basal segment (LB_{10}). The cross-sectional subsegmental and infrabronchial images (a 1.5 cm square) were used to train CNN. Each subject's image preprocessing time was less than 5 min. After image augmentation of the training data set by geometric transformation (see Supplementary Material 1 - Appendix for details), CNN was trained with 163,200 lung parenchyma images and 816,000 bronchial wall images. The images were categorized based on two-way (normal and COPD), three-way (normal, GOLD level 1–2, and GOLD level 3–4), and five-way (normal, GOLD level 1, 2, 3, and 4).

CNN

We used the pretrained DenseNet-201 as the backbone to train CNN (Supplementary Material 1 - Figure 1). The classifier layer initially designed for 1000 categories was replaced by our defined categories (two, three, and five categories). Subsequently, we used the training data set to finetune the network parameters with a back-propagation method, and utilized the Adam optimization algorithm for mini-batch gradient descent (see Supplementary Material 1 - Appendix for details). The running environment of CNN was Deep learning toolbox (Matlab R2019a, MathWorks) and a computer with two graphics processing units (Titan RTX, nVidia).

CNN's inference for each input image was a one-dimensional array of confidence, corresponding to the output classification

(two, three, and five). The mean confidence of all input images of each subject was regarded as the final confidence.[16] Subsequently, the output category with the highest confidence was taken as the final prediction category.[16,17] Due to the establishment of two CNN models of lung parenchyma and bronchus images, each subject had two sets of predictive confidence.

Quantitative CT

The same radiologist used a software package (Thoracic VCAR, GE Healthcare) to measure CT images (see Supplementary Material 1 - Appendix for details). The software automatically calculated %LAA-950 (percentage of low attenuation area <-950 HU) representing the severity of emphysema in the whole lung volume, and %WA (wall area percentage) and average lumen diameter indicating the severity of airway remodeling. Five representative bronchi (RB_1, RB_4, RB_{10}, LB_{1+2}, LB_{10}) were measured. Bronchial wall thickness is expressed as the square root of the wall area for a theoretical bronchus with a 10 mm lumen perimeter (AWT-Pi10).[18] The time of quantitative analysis was approximately 15 min per subject. We randomly selected 50 cases to evaluate the reproducibility of QCT method. Another radiologist with 10 years of experience who was blinded to the PFT results measured emphysema (%LAA-950) and airway wall (AWT-Pi10) using the same software.

Statistical analysis

Analysis of variance (ANOVA) was used to compare the body mass index (BMI) and PFT results among GOLD groups. To validate the performance of the CNN model in the training cohort, we used fourfold cross-validation to evaluate its generalizability. For two-way classification, the threshold of CNN-derived confidence was determined by the maximum value of area under the curve (AUC). Binary and multinomial logistic regression models were established to combine the prediction confidence of lung parenchyma and bronchus under two-, three-, and five-way classification conditions. Then, the regression models were used to discriminate or grade COPD. The diagnostic performance of CNN was evaluated by AUC, accuracy, sensitivity, specificity, and F1 score. F1 score is the harmonic average of model accuracy and recall in machine learning.

For the quantitative CT method, the intraclass correlation coefficient (ICC) and Bland–Altman analysis were used to evaluate the consistency of AWT-Pi10 between the two observers. In Bland–Altman analysis, the ratio of measured value of AWT-Pi10_1 divided by that of AWT-Pi10_2 was used, 95% limits of agreement was calculated by the mean of ratio ±1.96 × standard deviation of ratio. We used %LAA-950 and AWT-Pi10 to establish a diagnostic model of two-way classification. The cut-off value of reaching the maximum AUC determined in the training cohort was used to evaluate the data in the test cohort. The AUCs of CNN and quantitative CT method were compared by DeLong test,[19] and the diagnostic metrics were compared by Pearson's χ^2 test. A $p < 0.05$ was considered statistically significant. Software

Table 1. Characteristics of patients classified by GOLD level

	Training cohort (n = 373)	External test cohort (n = 226)	p-value
Age, yr	66.0 ± 11.5	63.5 ± 12.5	0.788
Male, n (%)	259 (69.4%)	164 (72.5%)	0.224
BMI, kg/m2	22.8 ± 3.3	22.7 ± 3.2	0.527
GOLD level			
Normal, n	141	79	
FEV_1/FVC	83.9 ± 6.0	83.6 ± 8.2	0.923
FEV_1%pred	98.6 ± 15.4	92.3 ± 13.2	0.973
GOLD 1, n	52	31	
FEV_1/FVC	66.6 ± 3.0	64.8 ± 3.5	0.058
FEV_1%pred	93.0 ± 11.4	89.3 ± 6.4	0.141
GOLD 2, n	65	44	
FEV_1/FVC	61.9 ± 5.5	59.8 ± 7.3	0.015
FEV_1%pred	64.8 ± 9.1	65.8 ± 7.6	0.403
GOLD 3, n	68	42	
FEV_1/FVC	49.1 ± 8.2	48.2 ± 9.0	0.391
FEV_1%pred	40.3 ± 7.0	39.9 ± 5.2	0.324
GOLD 4, n	47	30	
FEV_1/FVC	38.8 ± 11.0	44.2 ± 8.3	0.707
FEV_1%pred	24.8 ± 5.3	25.2 ± 3.0	0.076

BMI = body mass index; FEV_1%pred = forced expiratory volume in the first second as a predicted percentage; FVC = forced vital capacity;GOLD = the global initiative for chronic obstructive lung disease.

Data are indicated as mean ± standard deviation. p-value indicates the significance of values between the training and external test cohorts.

packages (MedCalc 20.03, MedCalc Software, and SPSS 22.0, IBM) were used for statistical analysis.

RESULTS

Patient characteristics

The search identified 599 subjects, including 373 from Hospital-A as the training cohort and 226 from Hospital-B as the test cohort (Figure 1). There was no significant difference in gender and BMI between the two cohorts (all $p > 0.05$) (Table 1). The 599 subjects were from health checkup (n = 185, 31%), outpatient (n = 256, 43%), inpatient (n = 91, 15%), and emergency department (n = 67, 11%).

Reproducibility of QCT method

The measurement of emphysema was automatic, %LAA-950 measured repeatedly by software were same between the two measurements. For AWT-Pi10, the ICC between the two observers was 0.991 (95% CI: 0.984–0.994), and Bland–Altman plots showed that the AWT-Pi10 measurement between the two observers was highly consistent (Supplementary Material 1 - Figure 1).

Determining the presence of COPD

Supplementary Material 1 - Table 2 shows the confusion matrix of two-way CNN classification. Combining the CNN prediction confidence based on lung parenchyma and bronchus with the binary logistic regression model (Supplementary Material 1 - Table 3), two formulas were established: $p = \frac{1}{1+e^{-(-7.366+6.858a+8.662b)}}$ for the training cohort, and $p = \frac{1}{1+e^{-(-2.259+2.135a+3.615b)}}$ for the test cohort, where a, b, and p represent the confidence of COPD based on lung parenchymal images, bronchial images, and integration regression model, respectively. In the training cohort, the AUC, accuracy, sensitivity, specificity, and F1 score of the regression model were 0.987 (95% CI: 0.969–0.996), 93.3% (90.3–95.6%), 94.8% (91.1%–97.3%), 90.8% (84.8–95.0%), and 0.946, respectively. In the test cohort, these metrics were 0.899 (0.853–0.935), 81.7% (76.2–86.7%), 81.0% (73.7%–87.0%), 83.5% (73.5–90.9%), and 0.853, respectively. Table 2 lists the diagnostic metrics of CNN for the presence of COPD. Figure 3 displays the receiver operating characteristic curves. Figure 4 shows two representative cases.

For quantitative CT method, a binary logistic regression model was established to combine the measurement results of parenchyma and bronchus as follows: $p = \frac{1}{1+e^{-(-18.304+0.696a+3.432b)}}$ where a, b, and p represent the confidence of COPD based on %LAA-950, AWT-Pi10, and binary regression model, respectively. In the training cohort, AUCs of %LAA-950, AWT-Pi10, and the binary regression model were 0.942 (0.913–0.964), 0.922 (0.890–0.947), and 0.970 (0.947–0.985), respectively, lower than those of CNN (all DeLong's $p < 0.05$). In the test cohort, the accuracy, sensitivity, specificity and F1 score of the binary regression model were 68.1% (61.6–74.2%), 76.2% (68.4%–82.9%), 54.2% (42.9–65.2%) and 0.752, respectively, which were significantly lower than those of CNN (Pearson's χ^2 test, all $p < 0.05$).

Three- and five-way COPD classification

In the training cohort for three-way CNN classification, the accuracies of CNN based on lung parenchyma, bronchial wall, and the multinomial regression model were 81.7%, 83.1%, and 88.5%, respectively. In the five-way classification, the accuracies were 74.0%, 71.8% and 83.1% respectively.

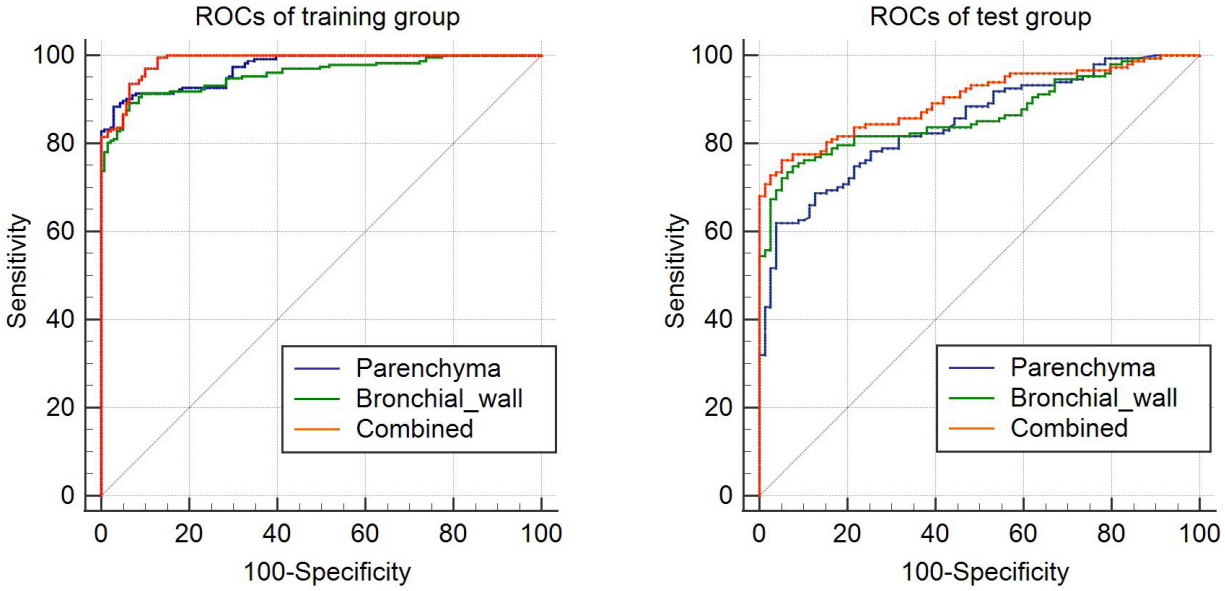
In the test cohort for three-way classification, the accuracies of CNN were 70.8%, 68.1%, and 77.4% based on lung parenchyma, bronchial wall, and the multinomial regression model, respectively. In the five-way classification, the accuracies were 54.4%, 57.1%, and 67.9%, respectively. The confusion matrix of three- and five-way classification are shown in Supplementary Material 1 - Tables 4 and 5.

Table 2. Diagnostic performance of CNN for the presence of COPD in the training and test cohorts

	Lung parenchyma	Bronchial wall	Logistic regression model
Training cohort			
AUC	0.972 (0.949–0.986)	0.957 (0.931–0.975)	0.987 (0.969–0.996)
Sensitivity	91.4% (87.0–94.7%)	89.2% (84.5–92.9%)	94.8% (91.1–97.3%)
Specificity	88.7% (82.2–93.4%)	92.2% (86.5–96.0%)	90.8% (84.8–95.0%)
Accuracy	90.4% (86.9–93.2%)	90.4% (86.9–93.2%)	93.3% (90.3–95.6%)
F1 score	0.922	0.920	0.946
External test cohort			
AUC	0.846 (0.792–0.891)	0.864 (0.813–0.906)	0.899 (0.853–0.935)
Sensitivity	74.8% (67.0–81.6%)	81.6% (74.4–87.5%)	81.0% (73.7–87.0%)
Specificity	78.5% (67.8–86.9%)	78.5% (67.8–86.9%)	83.5% (73.5–90.9%)
Accuracy	76.1% (70.0–81.5%)	80.5% (74.7–85.5%)	81.7% (76.2–86.7%)
F1 score	0.803	0.845	0.853

AUC, area under the curve; CNN, convolutional neural network; COPD, chronic obstructive pulmonary disease.

Figure 3. ROC curves for determining the presence of COPD in the training (A) and test (B) cohorts using CNN. ROCs of prediction confidence based on emphysema, bronchial wall, and the combination model (binary logistic regression model) are shown as blue, green, and orange lines, respectively. CNN, convolutional neural network; COPD, chronic obstructive pulmonary disease; ROC, receiver operating characteristic.



DISCUSSION

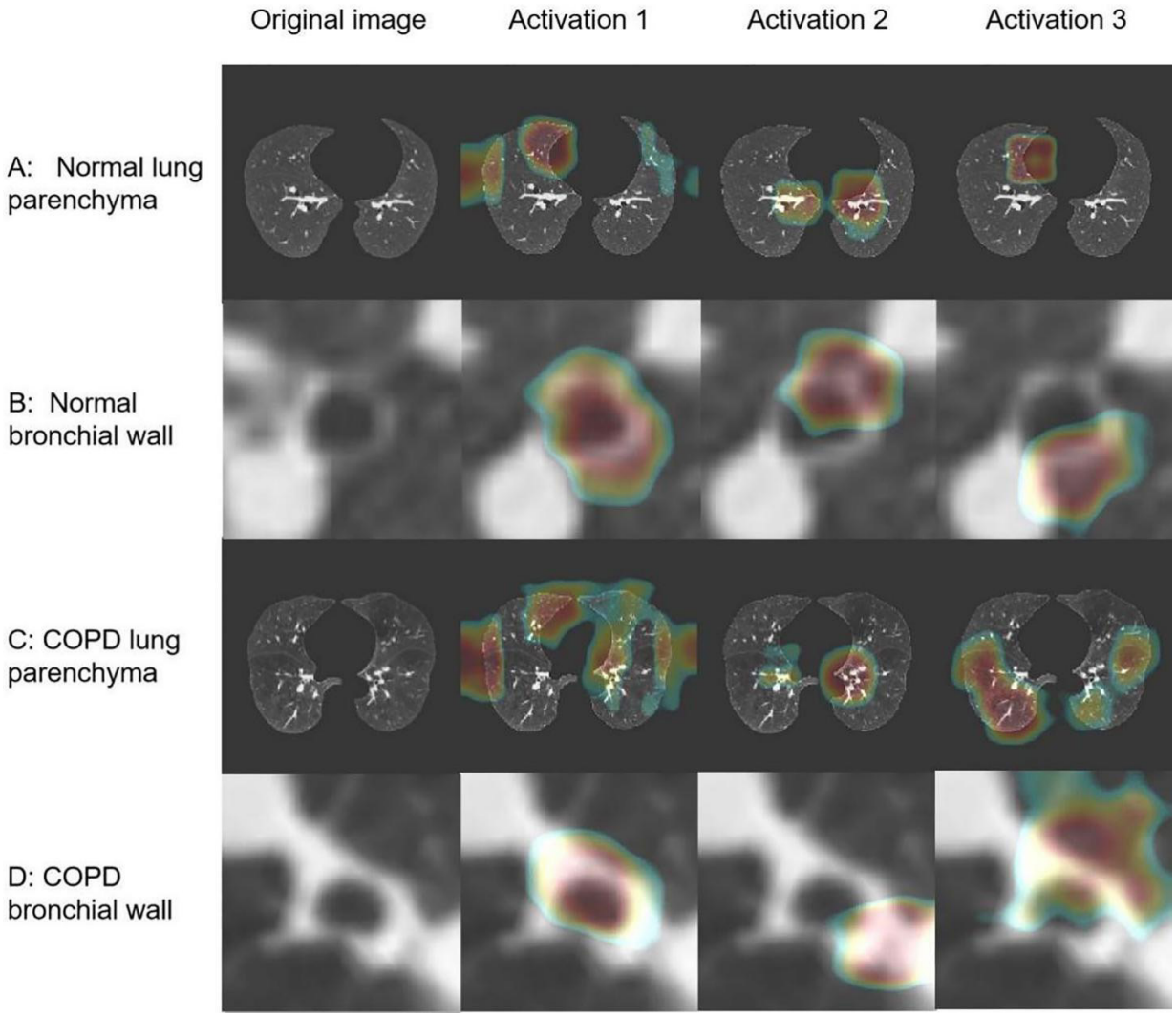
In this study, CNN can identify the main pathological changes of COPD (emphysema and airway wall remodeling) based on CT images, so as to infer lung function and determine the existence and severity of COPD. CNN reached AUCs of 0.946 and 0.853 to determine the presence of COPD in the training and external test cohorts, respectively. The diagnostic accuracies of COPD in the training and test cohorts were 93.3 and 81.7%, respectively, significantly higher than those of the quantitative CT method. In the external test cohort, the accuracies of CNN to determine COPD GOLD grade in three- and five- classifications were 77.4 and 67.9%, respectively. If the usefulness of this CNN algorithm would be translated into the clinic, *e.g.* be used as an opportunistic screening method at the patient level, presumably a large number of mild and moderate level COPD cases would be identified for smoking cessation and preventive treatment.

This study only applied single inspiratory CT instead of inspiratory and expiratory dual-phase scanning. Because of the increased radiation dose of the dual-phase scan, it is difficult to be used in clinical routine, so single inspiratory CT is more suitable and practical for diagnosing COPD. The CNN algorithm based on image feature learning of small bronchus effectively reflects airway limitation in COPD, because the airflow resistance of COPD mainly occurs in the small bronchus with a diameter of 2 mm,[20] where the quantitative CT method based on bronchial wall contour segmentation may be inaccurate. The image analysis of CNN does not rely on airway contour segmentation and airway measurement parameter selection, and the well-trained CNN model was not significantly affected by CT scanning methods, *i.e.* different CT equipment, radiation dose setting, and reconstruction kernels. These are the advantages of CNN over quantitative CT. Quantitative CT measurement requires manual operation to segment the bronchus. Different scanning schemes have a significant influence on the measurement of the bronchi. The low reproducibility limits the broad application of QCT.

During an inspiratory scan, two CT biomarkers, emphysema and airway wall remodeling reflecting the main pathological changes of COPD, were combined to reflect the pathological changes of COPD. As far as we know, the previous deep learning studies of COPD have not simultaneously evaluated lung parenchyma and bronchial wall. Some studies aimed to improve the segmentation and extraction of distal airways.[21,22] Gonzalez et al[23,24] used four chest CT images as inputs for the CNN model, including one axial, one coronal, and two sagittal images, to predict the presence and staging of COPD. Their results showed that in the training cohort, the accuracy of two-way classification was 77.3%, while in the external test cohort, the accuracy dropped to 29.4%. As a pioneer study, this study showed the possibility of using CNN to discriminate between COPD and non-COPD patients. But, the four kinds of CT images mainly characterize the image feature of emphysema and air trapping, and the distal bronchi cannot be fully displayed and used as evidence for identifying COPD. Although the study sample is specific, our CNN approach based on lung parenchyma and distal bronchi improved the classification accuracy of the external test cohort to 81.7%, which showed the potential of COPD recognition on CT images, and inspire the development of automatic algorithms.

In deep learning studies, external test cohorts are used to verify the generalizability of the CNN model. Therefore, the results of deep learning outside the model development environment will not be too optimistic.[25] In our study, although the general information of the training and test cohorts matched well, performance degradation occurred in the test cohort. One possible

Figure 4. Two case examples in the test cohort, and the visualization of three most activated channels in CNN. The distribution of feature channel parameters in the last 2D multichannel layer of DenseNet-201 are associated with the activation of CNN model. This 2D layer transformed into a one-dimensional layer by a pooling algorithm and finally determined the classification layer through the fully connected layer and softmax layer. Thus, we visualized the three most activated channels and superimposed the activated areas into the original images; this was done to mark image features in details that were closely associated with the final classification results. (A, B) A female without COPD, 42 years old, FEV1/FVC = 90.86, FEV1%pred = 104.8. The activated areas represented by the three most active channels are the normal bronchial wall and the normal lung parenchyma, which are the image features that classify this subject as normal. (C, D) A male with COPD, 79 years old, FEV1/FVC = 35.22, FEV1%pred = 29.60. The activated areas represented by the three most active channels are the thick bronchial wall and extensive emphysema, which are the image features that classify this patient as COPD. CNN, convolutional neural network; COPD, chronic obstructive pulmonary disease; FEV_1/FVC, forced expiratory volume in the first second divided by forced vital capacity; FEV_1%pred, forced expiratory volume in the first second as a predicted percentage.



explanation is that the images of training and test cohorts came from different CT vendors and types. Performance degradation of the external test data set was also found in a previous study, which used CNN based on original CT images to stage COPD.[26] The accuracy of CNN is also affected by the network infrastructure. We applied a DenseNet structure[27] with 201 layers and feed-forward connection to improve the image processing performance by alleviating the vanishing-gradient problem. DenseNet has been widely used in medical imaging, such as using MR images to predict the genotype glioma of isocitrate dehydrogenase[28] and establishing cardiac segmentation and disease diagnosis system.[29]

This study has several limitations. First, this study was conducted based on several CT machines and two medical centers. The robustness and generalizability of our method need to be verified in more CT machines and a larger population. Second, in order to train solid CNN models, we included patients who underwent CT and PFT within 3 days. But, PFT is usually applied in patients with respiratory disorders. If the models are used for another test

set who only have CT examination, the incidence rate of disease may be different. It is necessary to conduct a prospective study to investigate the impact of CNN on patients from different sources, *i.e.* health checkup, outpatient, inpatient, or emergency. Third, this study used preprocessed CT image features (*i.e.* lung parenchyma and bronchial wall) as the standardized input images into CNN, future researches should be conducted on developing automatic CNN methods. Fourth, due to the inherent differences of CT scanners, the QCT method of simply measuring CT values is sometimes not comparable among different CT vendors.

CONCLUSION

CNN can identify the main pathological changes of COPD (emphysema and airway wall remodeling) based on CT images to infer lung function and determine the existence and severity of COPD. This approach provides an alternative and effective way for early detection of COPD using the extensively used chest CT, if pulmonary function test is not available. Patients with positive CT findings of COPD classified by artificial intelligence may have airway limitation; thus, they should be referred to specialists for further diagnosis and treatment. Early detection of COPD by extensively used chest CT may improve the prognosis of COPD.

CONTRIBUTORS

Lin Zhang carried out research conception, data collection and analysis, and manuscript drafting. Xueqian Xie supervised this study and provided deep learning algorithm support. Beibei Jiang conducted data visualization. Rozemarijn Vliegenthart and Hendrik Joost Wisselink revised the manuscript for important intellectual content. Xueqian Xie made final approval of the version to be published and agreed to be 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. All authors read and approved the final manuscript.

FUNDING

This study was sponsored by Ministry of Science and Technology of China (2016YFE0103000), National Natural Science Foundation of China (81971612 and 81471662), Shanghai Municipal Education Commission- Gaofeng Clinical Medicine Grant Support (20181814), Shanghai Jiao Tong University (ZH2018ZDB10). The funders played no role in the study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Global initiative for chronic obstructive lung disease. Global strategy for the diagnosis, management, and prevention for chronic obstructive pulmonary disease 2021 REPORT. 2021. Available from: https://goldcopd.org
2. Patel BD, Coxson HO, Pillai SG, Agustí AGN, Calverley PMA, et al. Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2008; **178**: 500–505. https://doi.org/10.1164/rccm.200801-059OC
3. Labaki WW, Martinez CH, Martinez FJ, Galbán CJ, Ross BD, et al. The role of chest computed tomography in the evaluation and management of the patient with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2017; **196**: 1372–79. https://doi.org/10.1164/rccm.201703-0451PP
4. Oudkerk M, Devaraj A, Vliegenthart R, Henzler T, Prosch H, et al. European position statement on lung cancer screening. *Lancet Oncol* 2017; **18**: e754-66. https://doi.org/10.1016/S1470-2045(17)30861-6
5. Patz EF, Greco E, Gatsonis C, Pinsky P, Kramer BS, et al. Lung cancer incidence and mortality in national lung screening trial participants who underwent low-dose ct prevalence screening: a retrospective cohort analysis of a randomised, multicentre, diagnostic screening trial. *Lancet Oncol* 2016; **17**: 590–99. https://doi.org/10.1016/S1470-2045(15)00621-X
6. Mets OM, Schmidt M, Buckens CF, Gondrie MJ, Isgum I, et al. Diagnosis of chronic obstructive pulmonary disease in lung cancer screening computed tomography scans: independent contribution of emphysema, air trapping and bronchial wall thickening. *Respir Res* 2013; **14**: 59. https://doi.org/10.1186/1465-9921-14-59
7. de-Torres JP, Wilson DO, Sanchez-Salcedo P, Weissfeld JL, Berto J, et al. Lung cancer in patients with chronic obstructive pulmonary disease. development and validation of the copd lung cancer screening score. *Am J Respir Crit Care Med* 2015; **191**: 285–91. https://doi.org/10.1164/rccm.201407-1210OC
8. Mouronte-Roibás C, Leiro-Fernández V, Fernández-Villar A, Botana-Rial M, Ramos-Hernández C, et al. COPD, emphysema and the onset of lung cancer. a systematic review. *Cancer Lett* 2016; **382**: 240–44. https://doi.org/10.1016/j.canlet.2016.09.002
9. Young RP, Hopkins RJ. Chronic obstructive pulmonary disease (copd) and lung cancer screening. *Transl Lung Cancer Res* 2018; **7**: 347–60. https://doi.org/10.21037/tlcr.2018.05.04
10. Hasegawa M, Nasuhara Y, Onodera Y, Makita H, Nagai K, et al. Airflow limitation and airway dimensions in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2006; **173**: 1309–15. https://doi.org/10.1164/rccm.200601-037OC
11. Xie X, Dijkstra AE, Vonk JM, Oudkerk M, Vliegenthart R, et al. Chronic respiratory symptoms associated with airway wall thickening measured by thin-slice low-dose ct. *AJR Am J Roentgenol* 2014; **203**: W383-90. https://doi.org/10.2214/AJR.13.11536
12. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; **316**: 2402–10. https://doi.org/10.1001/jama.2016.17216
13. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**: 115–18. https://doi.org/10.1038/nature21056
14. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; **318**: 2199–2210. https://doi.org/10.1001/jama.2017.14585
15. Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med* 2018; **6**: 837–45. https://doi.org/10.1016/S2213-2600(18)30286-8

16. He Y, Guo J, Ding X, van Ooijen PMA, Zhang Y, et al. Convolutional neural network to predict the local recurrence of giant cell tumor of bone after curettage based on pre-surgery magnetic resonance images. *Eur Radiol* 2019; **29**: 5441–51. https://doi.org/10.1007/s00330-019-06082-2
17. Zhang Y, van der Werf NR, Jiang B, van Hamersvelt R, Greuter MJW, et al. Motion-corrected coronary calcium scores by a convolutional neural network: a robotic simulating study. *Eur Radiol* 2020; **30**: 1285–94. https://doi.org/10.1007/s00330-019-06447-7
18. Grydeland TB, Dirksen A, Coxson HO, Pillai SG, Sharma S, et al. Quantitative computed tomography: emphysema and airway wall thickness by sex, age and smoking. *Eur Respir J* 2009; **34**: 858–65. https://doi.org/10.1183/09031936.00167908
19. 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. https://doi.org/10.2307/2531595
20. Bhatt SP, Washko GR, Hoffman EA, Newell JD Jr, Bodduluri S, et al. Imaging advances in chronic obstructive pulmonary disease. insights from the genetic epidemiology of chronic obstructive pulmonary disease (copdgene) study. *Am J Respir Crit Care Med* 2019; **199**: 286–301. https://doi.org/10.1164/rccm.201807-1351SO
21. Charbonnier J-P, Rikxoort E van, Setio AAA, Schaefer-Prokop CM, Ginneken B van, et al. Improving airway segmentation in computed tomography using leak detection with convolutional networks. *Med Image Anal* 2017; **36**: 52–60. https://doi.org/10.1016/j.media.2016.11.001
22. Yun J, Park J, Yu D, Yi J, Lee M, et al. Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net. *Med Image Anal* 2019; **51**: 13–20. https://doi.org/10.1016/j.media.2018.10.006
23. González G, Washko GR, Estépar RSJ. Deep learning for biomarker regression: application to osteoporosis and emphysema on chest ct scans. *Proc SPIE Int Soc Opt Eng* 2018; **10574**: 105741H. https://doi.org/10.1117/12.2293455
24. González G, Ash SY, Vegas-Sánchez-Ferrero G, Onieva Onieva J, Rahaghi FN, et al. Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am J Respir Crit Care Med* 2018; **197**: 193–203. https://doi.org/10.1164/rccm.201705-0860OC
25. Carin L, Pencina MJ. On deep learning for medical image analysis. *JAMA* 2018; **320**: 1192–93. https://doi.org/10.1001/jama.2018.13316
26. González G, Ash SY, Vegas-Sánchez-Ferrero G, Onieva Onieva J, Rahaghi FN, et al. Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am J Respir Crit Care Med* 2018; **197**: 193–203. https://doi.org/10.1164/rccm.201705-0860OC
27. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Honolulu, HI. IEEE Computer Society; July 2017. https://doi.org/10.1109/CVPR.2017.243
28. Liang S, Zhang R, Liang D, Song T, Ai T, et al. Multimodal 3d densenet for idh genotype prediction in gliomas. *Genes (Basel)* 2018; **9**(8): E382. https://doi.org/10.3390/genes9080382
29. Khened M, Kollerathu VA, Krishnamurthi G. Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Med Image Anal* 2019; **51**: 21–45. https://doi.org/10.1016/j.media.2018.10.004