ORIGINAL ARTICLE

Early detection of COPD based on graph convolutional network and small and weakly labeled data

Zongli Li[1,2,3] · Kewu Huang[1,2] · Ligong Liu[4] · Zuoqing Zhang[3]

Abstract

Chronic obstructive pulmonary disease (COPD) is a common disease with high morbidity and mortality, where early detection benefits the population. However, the early diagnosis rate of COPD is low due to the absence or slight early symptoms. In this paper, a novel method based on graph convolution network (GCN) for early detection of COPD is proposed, which uses small and weakly labeled chest computed tomography image data from the publicly available Danish Lung Cancer Screening Trial database. The key idea is to construct a graph using regions of interest randomly selected from the segmented lung parenchyma and then input it into the GCN model for COPD detection. In this way, the model can not only extract the feature information of each region of interest but also the topological structure information between regions of interest, that is, graph structure information. The proposed GCN model achieves an acceptable performance with an accuracy of 0.77 and an area under a curve of 0.81, which is higher than the previous studies on the same dataset. GCN model also outperforms several state-of-the-art methods trained at the same time. As far as we know, it is also the first time using the GCN model on this dataset for COPD detection.

Keywords Chronic obstructive pulmonary disease · Graph convolution network · Deep learning · Early detection

1 Introduction

Chronic obstructive pulmonary disease (COPD) is a lung disease with high global incidence, high mortality, and high medical costs. The World Health Organization (WHO) predicted that COPD would be the third leading cause of death in the world by 2030 [1]. Nevertheless, patients with early COPD can be easily neglected as they have no symptoms or only mild symptoms [2, 3]. Most of the patients have often developed into the moderate-to-severe stage when diagnosed, seriously affecting the quality of life, and the cost of treatment also rises sharply [4]. Therefore, early detection of COPD is associated with a lower risk of exacerbations, fewer comorbidities, and lower costs. There is growing awareness of the need to identify COPD in patients at an early stage.

Spirometry is the cornerstone of COPD diagnosis. However, it is considered to be considerably underdiagnosed and limited by insensitivity to the early stages of COPD [5]. COPD is a highly heterogeneous disease, as shown in Fig. 1, which has different imaging phenotypes and histopathological features, such as emphysema, bronchial wall thickening, gas trapping, interstitial lung abnormality, bronchiectasis, and so on [6]. Computed tomography (CT) has been used to capture the presence, pattern, and extent of phenotypic abnormalities associated with COPD, and has become one of the most widely used imaging modalities for characterizing heterogeneities of COPD [7, 8]. With the widespread application of CT, there is an opportunity to use these scans to identify those with COPD, with subsequent confirmation using spirometry.

Previous COPD classification based on CT imaging was conventionally approached using traditional machine

✉ Kewu Huang
  kewu_huang@126.com

1   Department of Pulmonary and Critical Care Medicine, Beijing Chao-Yang Hospital, Capital Medical University, Beijing 100020, People's Republic of China

2   Beijing Institute of Respiratory Medicine, Beijing 100020, People's Republic of China

3   Department of Respiratory, Shijingshan Teaching Hospital of Capital Medical University, Beijing Shijingshan Hospital, Beijing 100043, People's Republic of China

4   Department of Enterprise Management, China Energy Engineering Corporation Limited, Beijing 100022, People's Republic of China

Fig. 1 COPD is a highly heterogeneous disease, with lesions distributed in a diffuse and irregular way. Two axial thin-section CT scans of a patient with COPD disease. **a** bronchial wall thickening; **b** centrilobular emphysema; **c** paraseptal emphysema; **d** interstitial lung abnormality; **e** bronchiectasis; **f** normal lung parenchyma. Asterisks and arrows represent emphysema and bronchiectasis, respectively
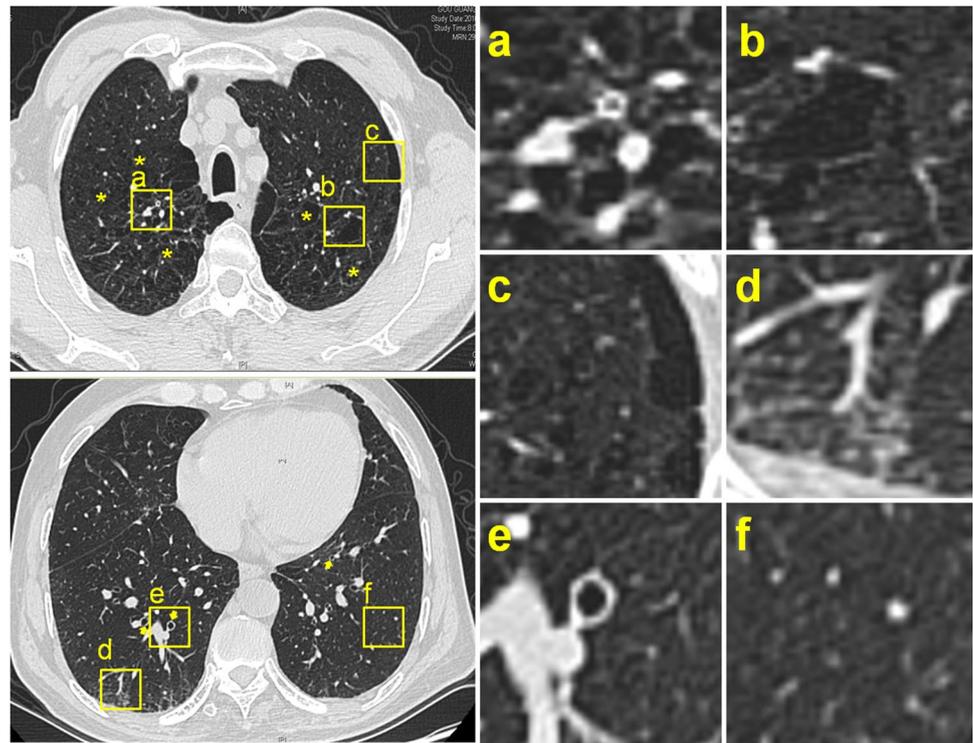
learning techniques. For example, Feragen et al. applied a support vector machine (SVM) to the airway tree information of 1996 subjects, including 893 with COPD. Their highest accuracy in the COPD classification task was 64.9% [9]. Bodduluri et al. evaluated the ability of the *k* nearest neighbor learning algorithm to detect COPD patients. The best area under the curve (AUC) was 0.89 on texture feature sets [10]. Cheplygina et al. adopted multiple instance learning (MIL) methods which were a kind of weakly supervised classification where only patient-level labels were known for COPD classification and obtained an AUC of 0.742 [11]. Subsequently, their team utilized instance-transfer leaning to classify COPD from different centers, scanners, or subject distributions and had an AUC of 0.790, 0.917, 0.956, and 0.953 for the 4 datasets of DLCST, COPDGene1, COPDGene2, and Frederikshavn [12].

A major weakness of such feature engineering is that they require prior knowledge of the features, which makes them strongly application dependent. With the development of artificial intelligence, modern deep learning methods enable direct interpretation of image data, going directly from the raw image data to the clinical outcome without relying on the specification of radiographic features of interest, and have achieved excellent results in COPD detection. A convolutional neural network (CNN) is one of the most successful deep learning architectures in the field of computer vision. For example, González et al. trained deep CNN models with an accuracy (ACC) of 0.773 for the detection of COPD in the COPDGene testing cohort [13]. Hatt et al. had developed one CNN model with an accuracy of 0.777 for the COPDGene cohort and 0.762 for the National Lung Screening Trial (NLST) cohort [14]. In addition to CNN, recent work by Tang and colleagues further showed that residual neural networks can effectively diagnose COPD (AUC = 0.88) using data from the PanCAN cohort with stable replication results in ECLIPSE based on a subset of slices [15].

However, traditional machine learning methods usually failed to capture complex features, while modern deep learning methods, either from scratch or fine tuning, generally required a large amount of labeled training data and extensive computational and memory resources. In addition, on account of the constraints caused by the processing capabilities of existing graphical processing units, the full CT images from an individual were not used for the above deep learning models. These studies usually extract a subset of CT slices to build up a single montage for an individual and directly input this montage to a 2D-CNN [13, 14]. Because the disease heterogeneity of COPD also shows at the diverse spatial distributions of abnormalities. For instance, the majority of COPD subjects have the upper lobe dominant emphysema (80.6%) [16]. The above single montage image can not make full use of the spatial information of CT image and unavoidably lose information. Ahmed et al. proposed that 3D CNN can extract larger spatial context to preserve more discriminative information which subsequently could improve COPD classification [17]. Ho et al. compared the

performances of a 3D-CNN model with CT-based parametric response mapping to an alternative 2D-CNN model, and found that the proposed 3D approach significantly outperformed the 2D approach [18]. Furthermore, these studies above focused on the accuracy of the model but did not pay attention to the early identification of COPD.

To make full use of the spatial information of CT images and further achieve the purpose of early detection of COPD, we used the publicly available Danish Lung Cancer Screening Trial (DLCST) dataset, which contained nearly 90% of patients with early mild to moderate COPD, and which is small ($n = 600$) and weakly labeled (assign COPD or non-COPD to the entire image, and no information on where or how serious the lesions are available). In the dataset, each image is represented by 50 cubic 3D regions of interest (ROIs), sampled at random locations within the segmented lung parenchyma. Each ROI is described by histograms of responses of 8 filters (smoothed image, gradient magnitude, Laplacian of Gaussian, three eigenvalues of the Hessian, Gaussian curvature, and eigen magnitude) at 4 scales (0.6, 1.2, 2.4, and 4.8 mm), which aim to capture the texture of the image.

Because the distribution of these 3D-ROIs is disordered and irregular, and the dataset is imbalanced, which contains nearly 90% of mild-moderate COPD patients, we use the GCN-based method with focal loss to classify COPD in this study. Different from several well-known frameworks such as DNN, CNN, and RNN (LSTM and GRU), which are used to process data in Euclidean space, such as pictures, voice, and text, graph neural networks can be applied to more abundant topological data, such as social network, recommendation system, transportation network, etc. Recently, graph convolutional network (GCN) has attracted more attention, which intends to generalize CNN on non-Euclidean graph data, and integrates phenotypic information into a graph to establish interactions between individuals and populations, which can achieve an excellent effect by graph theory. In the field of medical image analysis, GCNs have demonstrated to be superior in learning network representations tailored for identifying specific brain disorders such as early mild cognitive impairment (EMCI) [19], Parkinson's disease [20], Alzheimer's disease (AD) [21] and autism spectrum disorder (ASD) [22]. In addition, GCNs have made significant breakthroughs in the diagnosis of COVID-19 pneumonia [23, 24], cervical cancer [25], breast cancer [26], and grading of colorectal cancer histology images [27]. All these studies validate the effectiveness of GCN for disease classification. However, while the use of graph-based representations is becoming more common in the medical domain, little work has been done to use GCN to study its application for COPD detection.

In our work, we formulated COPD detection as a graph classification problem and attempted to advance deep learning for graph-structured data with GCN. Specially, we first connected the ROIs from the same patient to construct the graph for modeling. Then we adopted the Chebyshev polynomials filter and input our graph into a GCN model for COPD detection. After that, ablation studies were conducted to evaluate the performance of the proposed model. Finally, we compared GCN with four successful classical CNN models (DenseNet121, VGG16, ResNet50, and InceptionV3) and a light gradient boosting machine (LightGBM) on the same dataset. The main contributions of this study can be summarized as follows:

1. We build a graph using all ROIs which are randomly selected and disorderly distributed, and then input it into the GCN model, hoping to better detect COPD with spatially heterogeneous by using the topological structure information between these ROIs.

2. Since the imbalanced dataset contains nearly 90% of mild-moderate COPD patients, which are easy to misclassify, we propose an optimization step for the GCN model using focal loss to improve the classification accuracy.

2 Materials and methods

2.1 Dataset

In this study, we employed the publicly available dataset, which contained derived features (320-dimensional feature vectors) from CT images of 300 COPD patients and 300 controls scanned at the Danish Lung Cancer Screening Trial (DLCST [National Clinical Trials identifier NCT00496977; ClinicalTrials.gov]) [28]. The patients with COPD were diagnosed by spirometry-based pulmonary function test (a post-bronchodilator $FEV_1/FVC < 70\%$) according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria [29]. The DLCST dataset included 197/93/10/0 COPD subjects with mild (GOLD 1)/moderate (GOLD 2)/severe (GOLD 3)/very severe (GOLD 4) COPD. Each CT image is assigned a global label according to the PFTs of the scanned subject that are acquired at the same time as the CT image, and ROIs are labeled with the global label of the CT image. Therefore, the images in this database are weakly labeled, i.e., per image, a diagnosis (COPD or no COPD) is given, but it is not known which parts of the lungs are affected.

In the small and weakly-labeled dataset, each chest image of these 600 participants was represented by 50 feature vectors, where each feature vector described a volumetric ROIs of size 41 * 41 * 41 voxels, extracted at random locations inside the lung mask. Then 30,000 ROIs were obtained in total. Next, the 3D-ROIs were represented by Gaussian scale space (GSS) features, or histograms of intensity values, which capture the image texture. After using eight filters,

four scales, and histograms of ten bins, $8 \times 4 \times 10 = 320$ features were generated. For details of the dataset, please refer to literature [12] and http://bigr.nl/research/projects/copd.

The DLCST was originally approved by the ethics committee of Copenhagen County, and all participants provided written informed consent. The present study did not require additional institutional review board approval.

2.2 Graph convolutional networks

Compared with the CNN-based methods, which operate convolution on local Euclidean structure, GCN generalizes the operation of convolution to non-Euclidean data (e.g., graph), and it can be categorized into spectral methods and spatial methods [30–32]. In this study, we use the spectral approach, which is based on the spectrum of graph Laplacian, and provides a well-defined localization operator on graphs to define graph convolutions. Specifically, it performs spectral graph convolution on the features of neighbor nodes, and learns the feature representation of each node combining the graph structure during the learning process. Here, we give a brief introduction of graph convolutional networks.

A graph G is a pair (V, E) with $V = \{v_1, \ldots, v_n\}$ the set of vertices, and $E \subseteq V \times V$ the set of edges. Each graph can be represented by an adjacency matrix A of size $n \times n$ describing the graph's connectivity. For GCN models, the goal is to learn a function of signals/features on a graph $G = (V, E)$ which takes as input [33]. Every neural network layer can be written as a nonlinear function

$$H^{(l+1)} = f\left(H^{(l)}, A\right),$$

with $H^{(0)} = X$ and $H^{(L)} = Z$ (or z for graph-level outputs). X is the feature matrix ($N \times D$ matrix), L is the number of layers, and Z is a node-level output. The specific models then differ only in how $f(\cdot, \cdot)$ is chosen and parameterized.

$$f\left(H^{(l)}, A\right) = \sigma\left(\mathrm{A}H^{(l)}W^{(l)}\right),$$

where $W^{(l)}$ is a weight matrix for the l-th neural network layer and $\sigma(\cdot)$ is a nonlinear activation function (i.e., ReLU in our experiments).

We usually use a normalized adjacency matrix in order to change the scale of the feature vectors. The adjacency matrix is processed to achieve a better effect and computational efficiency [31]. One normalized adjacency method is as follows:

$$f\left(H^{(l)}, A\right) = \sigma\left(\widehat{D}^{-\frac{1}{2}}\widehat{A}\widehat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right),$$

with $\widehat{A} = A + I$, where I is the identity matrix and $\widehat{D}$ is the diagonal node degree matrix of $\widehat{A}$.

GCN can be considered a Laplacian smoothing operator for node features over graph structures [34]. The architecture of GCN consists of a series of convolutional layers, each followed by the activation functions to increase nonlinearity. The first hidden layer is set to input original node features. All layers share the same adjacency matrix. To localize the filter and reduce the number of parameters, we employ the Chebyshev polynomials to approximate the convolutional kernels, which can efficiently decrease the computation complexity of eigendecomposition.

## 2.3 Classification based on GCN model

2.3.1 Graph construction

The distribution of 50 3D-ROIs per subject extracted at random locations from the segmented lung parenchyma is disordered, so we approach the classification of COPD from lung images as a graph problem. Before performing COPD identification using the GCN model, we should first construct a graph using all subjects. In this study, we define each ROI as a node or vertex, and the correlation between every 50 ROIs from the same subject as edges.

Among these 600 subjects (30,000 ROIs), 400 (20,000 ROIs) are randomly selected for training, 100 (5000 ROIs) for validation, and 100 (5000 ROIs) for testing. As shown in Fig. 2, we connect these ROIs from the same subject to construct the graph for modeling and make a further analysis using GCN. Specifically, we take 30,000 ROIs as vertexes, and if vertexes v_i and v_j are from the same subject, they are connected by an edge. We do not establish the connections between nodes in different subjects. In this way, every 50 ROIs are connected to each other, and then 1225 edges are generated. This leads to $1225 \times 600 = 735,000$ edges in total. Then the construction of the graph with a total of 30,000 vertexes and 735,000 edges is completed, and each vertex is characterized by a vector with 320 dimensions.

The graph can be represented by the adjacency matrix A, where $a_{i,j} = 1$ if there is an edge from vertex v_i to vertex v_j, and $\mathrm{a}_{i,j} = 0$ otherwise. This means we establish the connections between ROIs of the same subject. If the subject has COPD, the 50 ROIs belonging to this patient are all labeled as 1, otherwise as 0. As the dataset is weakly labeled, randomly sampled ROIs from COPD patients will therefore likely contain both diseased and healthy tissue where the healthy tissue ROIs still receives the label 1 that is COPD. As shown in Fig. 2, the green ROIs represent healthy tissue, while the red ROIs represent diseased tissue. Thus, the green ROI representing healthy tissue in the COPD image will be marked as 1. This is more common in mild COPD because their images have a higher proportion of green ROIs, that is, healthy tissue, which increases the probability of

Fig. 2 The graph generated by ROIs. If ROIs are from the same patient, they are connected to each other with an edge. A is the adjacency matrix of ROIs. L is the symmetric normalized Laplacian matrix of A. C is the Chebyshev polynomial of L, and F is the feature matrix of ROIs. n is the number of ROIs ($n = 30,000$), and m is the dimension of the feature vector ($m = 320$)
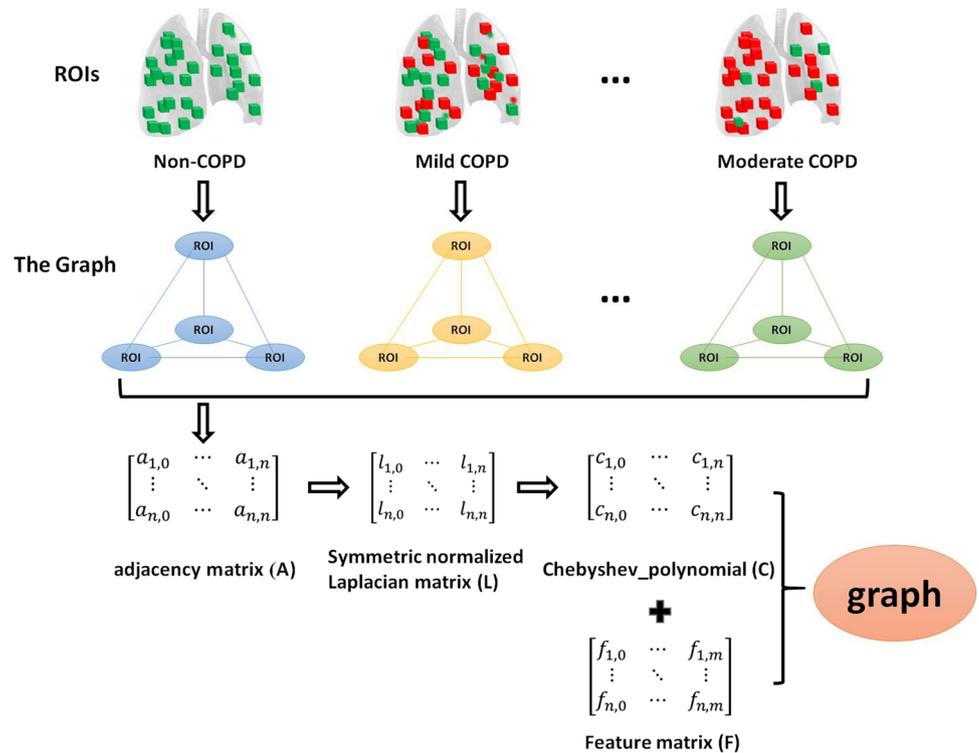
misclassification. For this reason, we call it a hard sample, thereby posing a challenge in accurate classification.

2.3.2 Classification using GCN model

The GCN model is trained, validated, and tested using the whole graph which was constructed with 30,000 ROIs from 600 cases. The input of the GCN model is the feature matrix X_{in} (30,000 * 320) of these ROIs and the Chebyshev polynomial approximation of adjacency matrix A (30,000 * 30,000), that is, the graph described in Fig. 2. Figure 3 shows the flowchart of the proposed approach. As can be seen from Fig. 3, A is the adjacency matrix of ROIs (30,000 * 30,000), which is symmetric. $L = D - A$ is the Laplacian matrix of A (where D is the degree matrix). $L_{sym} = D^{\left(-\frac{1}{2}\right)} L D^{\left(-\frac{1}{2}\right)}$ is symmetric normalized Laplacian. $L^\wedge = \frac{2}{\lambda_{max}} L_{sym} - I$ is the rescaled matrix of $L^\wedge_{sym}$ (where λ_{max} is the largest eigenvalue of L_{sym}, that is, the spectral radius. C is truncation (the order is 3) in order to obtain the approximation of L. X_{in} is the feature matrix of ROIs (30,000 * 320).

The GCN model consists of four graph convolutional layers with a ReLU function as the activation function and the focal loss function as the final output layer. The output of the GCN model is the probability y that ROI is labeled as 1. If all 50 ROI labels of a patient are predicted to be 1 (y is greater than 0.5), we judge that the patient is a COPD patient, otherwise who is not. In Fig. 3, grapconv is the graph convolution that is employed directly

on graph-structured data to extract highly meaningful patterns and features in the space domain and was described in detail in [31]. The parameters of the proposed GCN structure are as follows: the learning rate is 0.001, the dropout rate is 0.5, the optimizer is Adam, and the kernel regularizer is l_2.

Considering that the DLCST dataset contains nearly 90% of mild-moderate COPD, which are difficult to be correctly classified, the focal loss is applied to the classification loss to further improve the accuracy of classification [35]. A common method for addressing class imbalance is to introduce a weighting factor $\alpha \in [0, 1]$ for class 1 and $1 - \alpha$ for class -1. While balances the importance of positive/negative examples, it does not differentiate between easy/hard examples. The loss function should be reshaped to down-weight easy examples and thus focus training on hard negatives. So a modulating factor to the cross-entropy loss was added, with tunable focusing parameter $\gamma \geq 0$. In our experiment, since COPD and no-COPD samples are balanced $\alpha = 1$. While $\gamma = 2$ because the dataset contains many mild-moderate COPD that is difficult to classify, the model needs to assign greater weight to hard samples.

A total of 30,000 ROIs from 600 cases constitute a graph as the input of GCN, so the batch = 1 in model training. In order to distinguish the training set, validation set, and test set in the training process, the 30,000 samples are given different weights. As shown in Fig. 4a, the weight of the

Fig. 3 Overview of the structure of the GCN. The input of the GCN is $C+F$ as shown in Fig. 2. n is the order of Chebyshev polynomial. In the focal loss function, α is 1 and γ is 2. The ellipsis indicates that the network structure is similar to the previous one, only the units are different

Fig. 4 The weight of samples in model training, validation, and predicting. The pink area has a weight of 1, while the green area has a weight of 0

training set samples is 1, and the weight of other samples is 0. Similarly, in model validation or testing, we set the weights of validation set or test set samples as 1 and the other two as 0 (Fig. 4 and Fig. 4). Then the prediction result on the test set is obtained in the end. In this way, we achieve the training, validation, and testing on the same graph by giving different weights to the samples.

3 Results

3.1 Classification performance

In our experiments, we apply our GCN model to DLCST dataset for binary classification tasks. To quantitatively analyze the classification performance of our method, we employed four metrics, including the area under the

receiver-operating characteristic curve (AUC), accuracy (ACC), precision (PR), and *F*-score. These metrics are computed from the true positive (TP), true negative (TN), false negative (FN), and false positive (FP) results.

Figure 5 shows the accuracy and loss of the training and validation dataset of the GCN model as the epochs proceeded. The training and validation loss decreased continuously, reaching approximately 0.45 after 400 iterations. Meanwhile, the training and validation accuracy increased gradually to over 0.76 after 400 iterations.

The confusion matrix of GCN on the test dataset (Fig. 5) reveals 10 false positives (FP, meaning that non-COPD is wrongly predicted as COPD) and 13 false negatives (FN, meaning that COPD is wrongly predicted as non-COPD). Therefore, the AUC, ACC, PR, and *F*-score of the model on the test dataset are 0.81, 0.77, 0.80, and 0.78, respectively.

3.2 Comparing with the-state-of-the-art methods

In this subsection, the proposed GCN method is compared with the-state-of-the-art methods on the same dataset, including the four classical convolutional neural networks (VGG16, DenseNet121, InceptionV3, and ResNet50) and lightGBM.

CNNs have made impressive success in image feature learning and play an important role in medical image classification. However, it is a challenging task for the small dataset to train deep CNN from scratch with proper convergence and without suffering from overfitting [36, 37]. Because there are only 600 samples in the public datasets, it is not enough to train models containing many deep convolution layers and involving more network parameters. So we use the CNNs with fine-tuning.

Take ResNet50 as an example. The architecture starts with an input layer, then there are five convolutions layers,

Fig. 5 Performance of COPD identification model by GCN. **a** Training/validation loss and accuracy of GCN model. **b** Confusion matrix of GCN on the test dataset
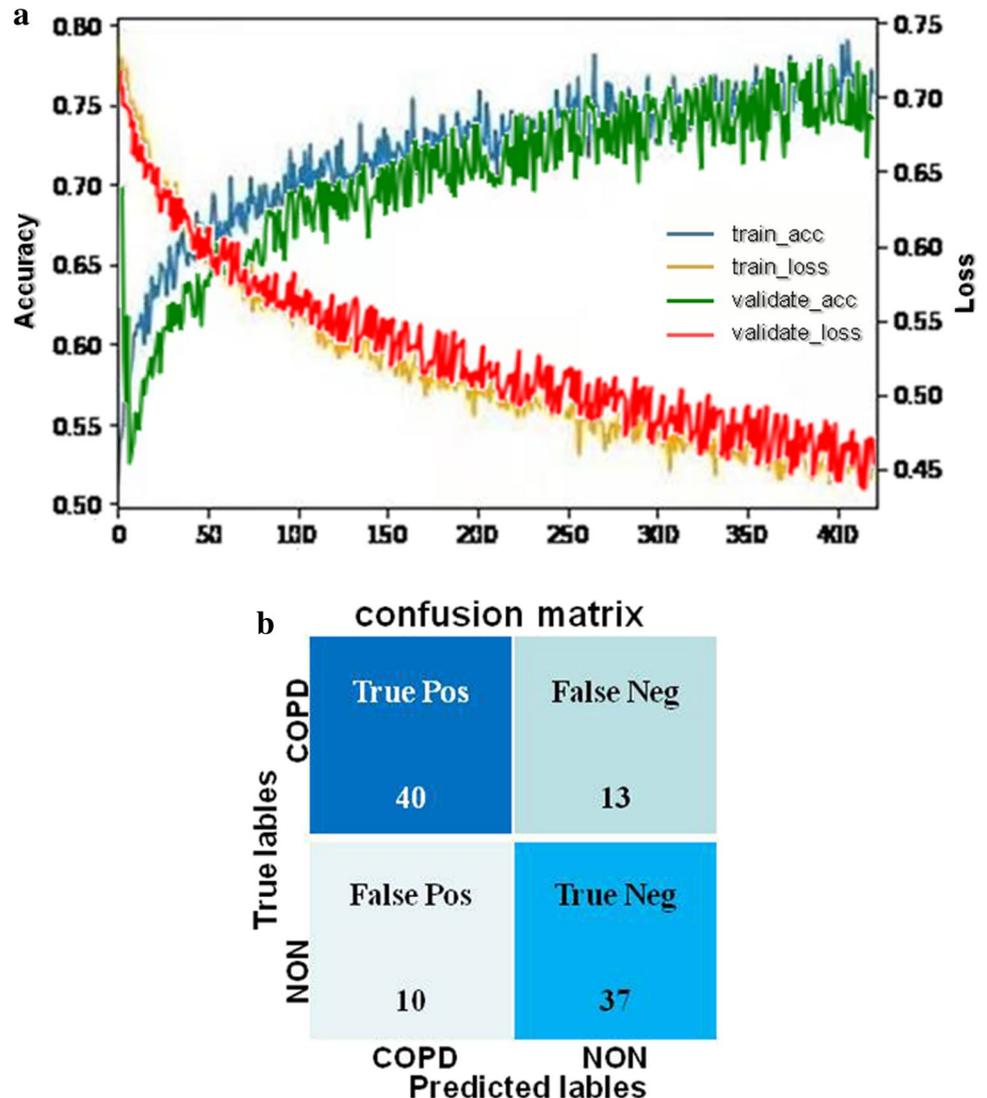
five ReLU layers, and five batch normalization layers, respectively. Two pooling layers are used after the first and second ReLU layers, respectively. A fully connected layer, softmax layer, and classification layer are also used at the end of the model. Fine tuning of this CNN model includes two aspects: (1) First, we remove the full connection layer at the top of the model and use two or more model blocks to retrain all parameters (for the ResNet50 model, the blocks are conv1, conv2_x, conv3_x, conv4_x, and conv5_x, as shown in Table 1). However, the results are not ideal due to overfitting (Fig. 6). (2) Second, in order to reduce overfitting, we use only one model block (for ResNet50, the block is conv1) and retrain the parameters, but the results are still unsatisfactory due to underfitting (Fig. 6).

Similar to ResNet50, we remove the full connection layer at the top of the other three models and use two or more model blocks of VGG16/DenseNet121/InceptionV3. The results are also overfitting. Then we use only one block (for VGG16, DenseNet121, and InceptionV3, we use block conv1, dense block1, and conv1, respectively), but the results are still underfitting. This shows that these models are not suitable for the DLCST dataset. Table 2 details the performance of the fine-tuned deep CNN models on COPD classification. As can be seen in Table 2, the GCN approach significantly outperforms the four classical pre-trained CNNs with fine-tuning, demonstrating the effectiveness of the GCN model.

We also compare our model with lightGBM, which is a high-performance gradient lifting model based on the decision tree algorithm proposed by Microsoft. As shown in Table 2, the overall accuracy of this machine learning model is 64%, the area under the curve (AUC) is 0.68, the PR is 0.66, and the F-score is 0.63, all of which are weaker than our GCN model.

Table 1 The blocks of ResNet50

Layer name	Layer structure
conv1	7×7, 64, stride 2
conv2_x	3×3, max pool, stride 2
	$\begin{bmatrix} 1 \times 1.64 \\ 3 \times 3.64 \\ 1 \times 1.256 \end{bmatrix} \times 3$
conv3_x	$\begin{bmatrix} 1 \times 1.64 \\ 3 \times 3.64 \\ 1 \times 1.256 \end{bmatrix} \times 4$
conv4_x	$\begin{bmatrix} 1 \times 1.256 \\ 3 \times 3.256 \\ 1 \times 1.1024 \end{bmatrix} \times 6$
conv5_x	$\begin{bmatrix} 1 \times 1.512 \\ 3 \times 3.512 \\ 1 \times 1.2408 \end{bmatrix} \times 3$

ROCs are commonly used for binary classifiers, which is a graphical plot of the true positive rate (TPR = sensitivity) vs. the false positive rate (FPR = 1 specificity) for a classifier as the discrimination threshold is varied. Figure 7 describes the ROC curves of different methods based on the extracted features, and it shows that our GCN has better AUC value on the DLCST database.

3.3 Ablation studies

In this subsection, we conducted several experiments on this dataset to validate the effectiveness of each key component of our proposed model, including the Chebyshev polynomial kernel and focal loss.

3.3.1 Efficacy of the Chebyshev polynomial kernel

Chebyshev polynomial kernel plays an important role in our model. In many cases, it can improve the training speed and achieve better prediction accuracy [38]. To investigate the effectiveness of the Chebyshev polynomial kernel, we designed two sets of controlled trials. We firstly set up a baseline network without the Chebyshev polynomial kernel and focal loss called simple GCN. We then used the Chebyshev polynomial kernel to the network and denoted it as GCN_cheby_softmax_cross_entropy. The results of two sets of controlled trials are summarized in Table 3. It reveals that GCN_cheby_softmax_cross_entropy performs much better than simple GCN.

3.3.2 Efficacy of focal loss

DLCST dataset contains a large number of mild-moderate COPD, which are hard to classify. Focal loss can focus on hard examples and prevent the vast number of easy examples during training [35]. To verify the performance of focal loss, we performed another ablation study. As shown in Table 3, the performance of GCN_cheby_softmax_cross_entropy is improved after replacing focal loss. These experimental results indicated that both Chebyshev polynomial kernel and the focal loss could improve the GCN's performance.

3.4 Comparison with related works on DLCST dataset

Table 4 shows the diagnosis performance of our method and other related methods on the DLCST dataset. All these methods used Gaussian scale-space features or histograms of intensity values in the ROI after filtering the image. Sørensen et al. proposed a fully automatic, data-driven approach for texture-based quantitative analysis with k nearest neighbor classifier and got an AUC of 0.713 [39]. Cheplygina et al. adopted various multiple instance

Fig. 6 The accuracy of the ResNet50 model with fine-tuning on training and validation dataset. **a** Overfitting when removing the full connection layer and using two blocks (conv1 and conv2_x). **b** Underfitting when using one block (conv1)
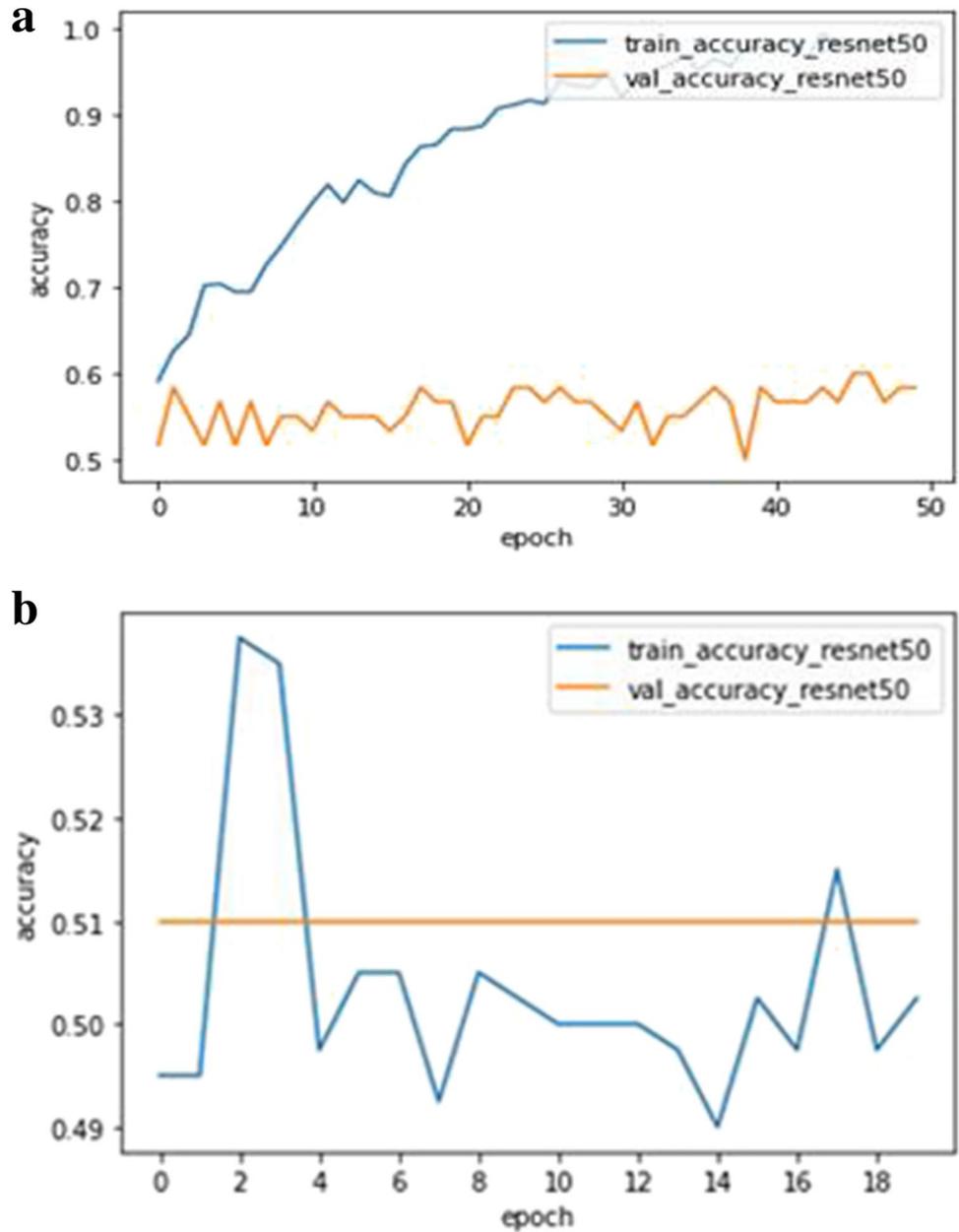

Table 2 Performance comparison of the applied methods

Model	AUC	ACC	PR	F-score
GCN	0.81	0.77	0.80	0.78
LightGBM	0.68	0.64	0.66	0.63
Fine tuning of ResNet50	0.50	0.53	0.53	0.69
Fine tuning of VGG16	0.50	0.47	0.00	0.00
Fine tuning of DenseNet121	0.50	0.53	0.53	0.69
Fine tuning of InceptionV3	0.50	0.47	0.00	0.00

learning (MIL) methods and obtained the best AUC of 0.742 by support vector machine [11]. Subsequently, their team utilized instance-transfer leaning to classify COPD from different centers, scanners, or subject distributions and had an AUC of 0.790 in the DLCST dataset [12]. It is observed that our GCN model achieves the best AUC, with it reaching 0.81. Compared to the other algorithms, it improves the AUC by 2.5–13.6%. Unfortunately, there is no more detailed comparative information about ACC, PR, and F-score because they are not discussed in these studies.
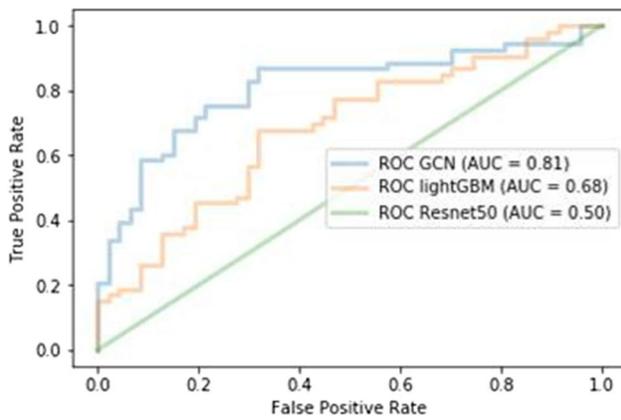
Fig. 7 ROC curves of the different classifiers based on the extracted features

Table 3 Ablation experiments for Chebyshev polynomial kernel and focal loss

Model	AUC	ACC	PR	F-score
Simple GCN model	0.76	0.71	0.74	0.72
GCN_cheby_softmax_ cross_entropy	0.78	0.74	0.77	0.75
GCN_cheby_focal loss	0.81	0.77	0.80	0.78

4 Discussion

This study developed a GCN architecture that can detect COPD on a small and weakly labeled CT imaging dataset. The proposed GCN model achieves an acceptable performance with an accuracy of 0.77 and area under curve of 0.81, which is higher than the previously studies on the same dataset. Furthermore, our GCN model also outperforms several state-of-the-art methods trained at the same time in this dataset. To the best of our knowledge, this is the first time to use GCN on small and weakly labeled data for detection of COPD.

With the development of artificial intelligence, deep learning, especially deep CNN, is a new and powerful

tool for machine vision and pattern recognition. Although CNNs have shown impressive performance in the medical field for imaging, their conventional formulation is limited to data structured in an ordered, grid-like fashion. Hence, they are inefficient when dealing with non-Euclidean data representations and when modeling global contextual information. GCNs have extended the theory of signal processing on graphs [40] and enhanced the data representation and classification capabilities of convolutional neural networks, which are highly effective for signals defined on regular Euclidean domains to irregular, graph-structured data defined on non-Euclidean domains such as topological structure [41, 42].

COPD is a complex and highly heterogeneous clinical entity. As shown in Fig. 1, its pathological abnormalities are multi-dimensional and multi-positional. The disease heterogeneity also shows in the diverse spatial distributions of abnormalities. Take emphysema as an example, which is the most common pathological type of COPD. Paraseptal emphysema is located in the periphery of the lung; centrilobular emphysema is predominantly in the upper lobes, while panlobular emphysema is predominantly in the lower lobes. In addition, each image in the public DLCST dataset is represented by 50 3D-ROIs, sampled at random locations within the segmented lung parenchyma. Therefore, the distribution of these ROIs is disordered and irregular, which makes COPD imaging diagnosis challenging.

Taking the above into consideration, we build a graph using a total of 30,000 ROIs from 600 cases and then input it into the GCN model with the Chebyshev polynomial kernel. Since the imbalanced dataset we used in the study contains nearly 90% of mild-moderate COPD patients, which are easy to be misclassified, we use the focal loss as an optimization step for the GCN model to improve the classification accuracy. The ablation study demonstrates that the proposed structures and modules contribute to the improvement of the performance. As shown in Table 4 and Fig. 7, the proposed GCN model obtains an AUC of 0.81, outperforms texture-based analysis with k nearest neighbor classifier [39] and various MIL classifiers [11, 12] with an AUC of 0.713, 0.742 and 0.79, respectively, on the same dataset.

Table 4 Performance comparison between our method and the related works on DLCST dataset

References	Sample size	Key points	Performance (AUC)
Sørensen, L. et al. [39]	300 COPO vs. 300 non-COPD	The κ nearest neighbor classifier	0.713
Cheplygina, V. et al. [11]	100 COPO vs. 100 non-COPD	Various MIL classifiers A total of 2296 feature vectors for each ROI	0.742 (the best obtained by support vector machine)
Sørensen, L. et al. [12]	300 COPO vs. 300 non-COPD	SimpleMIL logistic classifier A total of 420 feature vectors for each ROI	0.79 with average assumption 0.748 with noisy-or assumption
Our method	300 COPO vs. 300 non-COPD	GCN A total of 420 feature vectors for each ROI	0.81

Our work is not directly comparable to other studies using state-of-the-art systems because of different model parameters, training strategies, and data splits. Therefore, we also trained several other methods in the DLCST dataset at the same time, including the four classical CNNs and lightGBM. However, Fig. 6 shows that although the CNNs with fine-tuning are used, the results are still not ideal. The reason can be due to two aspects. On the one hand, each cubic 3D ROI is described by histograms of responses of 8 filters at 4 scales, which aims to capture the texture of the image. Therefore, we use the texture features of the preprocessed images rather than the raw images, which is different from the previous studies using a single montage image [13, 14]. This preprocessing method may miss potentially valuable information to a certain extent. For instance, Xu et al. used deep CNN to extract the automatically learned features, which are expected to be more discriminative and diverse than these texture features, and achieved an accuracy of 99.29% and an AUC of 0.9826 by transferring MIL for COPD identification [43]. We believe that if we can use CNN to extract features from raw images directly and then use GCN for classification, the performance of the model will be further improved. On the other hand, traditional CNNs analyze local areas based on fixed connectivity (determined by the convolutional kernel), leading to limited performance and difficulty in interpreting the spatial heterogeneity of COPD among diverse lung regions. Topological relations among these ROIs can be used to construct the graph, which can be analyzed to better integrate the correlation among these ROIs and improve the accuracy of classification. Meanwhile, we also trained a machine learning model based on the LightGBM algorithm and achieved an ACC of 0.64, AUC of 0.68, PR of 0.66, and F-Score of 0.63, which are all lower than our GCN model as well, confirming that our GCN model outperforms other methods.

It is worth mentioning that there are still some limitations. First, the public DLCST dataset does not disclose more detailed information such as age, gender, and smoking history of each patient, so we cannot further analyze the effect of GCN model on specific patients. Second, references [11, 12, 39] only analyzes the AUC value of other existing methods on the DLCST dataset, but does not analyze ACC, PR, and F-Score values. Therefore, there is no more detailed comparison between GCN and other related works in Table 4. Third, the size of the DLCST dataset is small, and it is collected from one single medical center. The generalization capabilities of the obtained GCN model are unknown.

In the future, large-scale and multi-center trials are required to prove the wide applicability of the present prediction algorithms in clinical practice. Moreover, we can use FCN or U-net models to accurately extract focus points from raw images directly, instead of randomly selecting ROIs

from lung mask. In addition, we can also use another graphical neural network, graphical attention network (GAT), for COPD detection. It applies an attention mechanism on graph neighborhoods to aggregate node information, which can assign larger weights to the more important nodes and guides us to study which ROIs are more important for COPD detection. Furthermore, GAT model does not need to use Laplace matrix for complex calculation and only updates the node characteristics by representing adjacent nodes, so it runs more efficiently on small datasets.

5 Conclusion

Previous studies on COPD detection using the DLCST dataset only focused on the feature information of ROIs themselves but not on the topological structure information around these ROIs. Moreover, COPD is a highly heterogeneous disease with various manifestations and diverse spatial distributions of abnormalities. To capture and explore such important information, we propose GCN based model for COPD classification. The obtained GCN model demonstrates superior performance in discriminating between subjects with and without COPD compared to the CNNs with fine-tuning, lightGBM and other classifiers used in previous researches on the DLCST dataset, which is small and weakly labeled. Furthermore, since the dataset contains nearly 90% mild-moderate COPD patients, this GCN model with focal loss can better realize the early detection of COPD. In addition, ablation experiments show the benefit of using the Chebyshev polynomial kernel and focal loss during training for model performance. Finally, we believe that it can help finding subgroups with high risk of COPD from large populations through CT scans ordered doing lung cancer screening.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Mathers CD, Loncar D (2006) Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 3(11):e442. https://doi.org/10.1371/journal.pmed.0030442

2. Zhong N, Wang C, Yao W et al (2007) Prevalence of chronic obstructive pulmonary disease in China. Am J Respir Crit Care Med 176(8):753–760. https://doi.org/10.1164/rccm.200612-1749OC

3. Mapel DW, Dalal AA, Blanchette CM et al (2011) Severity of COPD at initial spirometry-confirmed diagnosis: data from medical charts and administrative claims. Int J Chron Obstruct Pulmon Dis 6:573–581. https://doi.org/10.2147/COPD.S16975

4. Bellamy D, Smith J (2007) Role of primary care in early diagnosis and effective management of COPD. Int J Clin Pract 61:1380–1389. https://doi.org/10.1111/j.1742-1241.2007.01447.x

5. Gurney JW, Jones KK, Robbins RA et al (1992) Regional distribution of emphysema: correlation of high-resolution CT with pulmonary function tests in unselected smokers. Radiology 183(2):457–463. https://doi.org/10.1148/radiology.183.2.1561350

6. Lynch DA, Austin JH, Hogg JC et al (2015) CT-definable subtypes of chronic obstructive pulmonary disease: a statement of the Fleischner Society. Radiology 277(1):192–205. https://doi.org/10.1148/radiol.2015141579

7. Kauczor HU, Wielpütz MO, Jobst BJ et al (2019) Computed tomography imaging for novel therapies of chronic obstructive pulmonary disease. J Thorac Imaging 34(3):202–213. https://doi.org/10.1097/RTI.0000000000000378

8. Ostridge K, Wilkinson TM (2016) Present and future utility of computed tomography scanning in the assessment and management of COPD. Eur Respir J 48(1):216–228. https://doi.org/10.1183/13993003.00041-2016

9. Feragen A, Petersen J, Grimm D et al (2013) Geometric tree kernels: classification of COPD from airway tree geometry. Inf Process Med Imaging 23:171–183. https://doi.org/10.1007/978-3-642-38868-2_15

10. Bodduluri S, Newell JD Jr, Hoffman EA et al (2013) Registration-based lung mechanical analysis of chronic obstructive pulmonary disease (COPD) using a supervised machine learning framework. Acad Radiol 20(5):527–536. https://doi.org/10.1016/j.acra.2013.01.019

11. Cheplygina V, Sorensen L, et al (2014) Classification of COPD with multiple instance learning in 2014 22nd International Conference on Pattern Recognition. https://doi.org/10.1109/ICPR.2014.268

12. Cheplygina V, Pena IP, Pedersen JH et al (2018) Transfer learning for multicenter classification of chronic obstructive pulmonary disease. IEEE J Biomed Health Inform 22(5):1486–1496. https://doi.org/10.1109/JBHI.2017.2769800

13. González G, Ash SY, Vegas-Sánchez-Ferrero G et al (2018) Disease staging and prognosis in smokers using deep learning in chest computed tomography. Am J Respir Crit Care Med 197(2):193–203. https://doi.org/10.1164/rccm.201705-0860OC

14. Hatt C, Galban C, Labaki W, Kazerooni E, Lynch D, Han M (2018) Convolutional neural network based COPD and emphysema classifications are predictive of lung cancer diagnosis. In: Stoyanov D. et al. (eds) Image analysis for moving organ, breast, and thoracic images. RAMBO 2018, BIA 2018, TIA 2018. Lecture Notes in Computer Science, vol 11040. Springer, Cham. https://doi.org/10.1007/978-3-030-00946-5_30

15. Tang L, Coxson HO, Lam S et al (2020) Towards large-scale case-finding: training and validation of residual networks for detection of chronic obstructive pulmonary disease using low-dose CT. Lancet Digit Health 2(5):e259–e267. https://doi.org/10.1016/S2589-7500(20)30064-9

16. Ju J, Li R, Gu S et al (2014) Impact of emphysema heterogeneity on pulmonary function. PLoS ONE 9(11):e113320. https://doi.org/10.1371/journal.pone.0113320

17. Ahmed J, Vesal S, Durlak F, et al (2020) COPD Classification in CT images using a 3D convolutional neural network. arXiv:2001.01100

18. Ho TT, Kim T, Kim WJ et al (2021) A 3D-CNN model with CT-based parametric response mapping for classifying COPD subjects. Sci Rep 11(1):34. https://doi.org/10.1038/s41598-020-79336-5

19. Liu J, Tan G, Lan W et al (2020) Identification of early mild cognitive impairment using multi-modal data and graph convolutional networks. BMC Bioinformatics 21(Suppl 6):123. https://doi.org/10.1186/s12859-020-3437-6

20. Zhang X, He L, Chen K et al (2018) Multi-view graph convolutional network and its applications on neuroimage analysis for Parkinson's disease. AMIA Annu Symp Proc 5(2018):1147–1156

21. TA Song S Roy Chowdhury F Yang et al 2019 Graph convolutional neural networks for Alzheimer's disease classification Proc IEEE IntSymp Biomed Imaging 414–417 https://doi.org/10.1109/ISBI.2019.8759531

22. Jiang H, Cao P, Xu M et al (2020) Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction. Comput Biol Med 127(1):104096. https://doi.org/10.1016/j.compbiomed.2020.104096

23. Liang X, Zhang Y, Wang J, Ye Q, Liu Y, Tong J (2021) Diagnosis of COVID-19 pneumonia based on graph convolutional network. Front Med 7:612962. https://doi.org/10.3389/fmed.2020.612962

24. Wang SH, Govindaraj VV, Górriz JM et al (2021) Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. Inf Fusion 67:208–229. https://doi.org/10.1016/j.inffus.2020.10.004

25. Li Y, Chen J, Xue P et al (2020) Computer-aided cervical cancer diagnosis using time-lapsed colposcopic images. IEEE Trans Med Imaging 39(11):3403–3415. https://doi.org/10.1109/TMI.2020.2994778

26. Ye H, Wang DH, Li J, et al (2019) Improving histopathological image segmentation and classification using graph convolution network. in ICCPR '19: 2019 8th International Conference on Computing and Pattern Recognition. https://doi.org/10.1145/3373509.3373579

27. Zhou Y, Graham S, Koohbanani NA, et al (2019) CGC-net: cell graph convolutional network for grading of colorectal cancer histology images. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), pp. 388–398. https://doi.org/10.1109/ICCVW.2019.0050

28. Pedersen JH, Ashraf H, Dirksen A et al (2019) The Danish randomized lung cancer ct screening trial—overall design and results of the prevalence round. J Thorac Oncol 4(5):608–614. https://doi.org/10.1097/JTO.0b013e3181a0d98f

29. Vogelmeier CF, Criner GJ, Martinez FJ et al (2017) Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary 53(3):128–149. https://doi.org/10.1016/j.arbr.2017.02.001

30. Bruna J, Zaremba W, Szlam A, et al (2013) Spectral networks and locally connected networks on graphs. Computer Science. arXiv:1312.6203

31. Defferrard M, Bresson X, Vandergheynst P (2016) Convolutional neural networks on graphs with fast localized spectral filtering. arXiv:1606.09375

32. Wang Z, Zheng L, Li Y, et al (2019) Linkage based face clustering via graph convolution network. arXiv:1903.11306

33. Niepert, M, Ahmed M and Kutzkov K (2016) Learning convolutional neural networks for graphs. arXiv:1605.05273

34. Li Q, Han Z, Wu XM (2018) Deeper insights into graph convolutional networks for semi-supervised learning. arXiv:1801.07606

35. Lin TY, Goyal P, Girshick R et al (2020) Focal loss for dense object detection. IEEE Trans Pattern Anal Mach Intell 42(2):318–327. https://doi.org/10.1109/TPAMI.2018.2858826

36. Shin HC, Roth HR, Gao M et al (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures,

dataset characteristics and transfer learning. IEEE Trans Med Imaging 35(5):1285–1298. https://doi.org/10.1109/TMI.2016.2528162

37. Tajbakhsh N, Shin JY, Gurudu SR et al (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging 35(5):1299–1312. https://doi.org/10.1109/TMI.2016.2535302

38. Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. arXiv:1609.02907

39. Sorensen L, Nielsen M, Lo P et al (2012) Texture-based analysis of COPD: a data-driven approach. IEEE Trans Med Imaging 31(1):70–78. https://doi.org/10.1109/tmi.2011.2164931

40. Shuman DI, Narang SK, Frossard P et al (2013) The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Process Magazine 30(3):83–98. https://doi.org/10.1109/MSP.2012.2235192

41. Li, RY, Yao JW, Zhu XL, et al (2018) Graph CNN for survival analysis on whole slide pathological images. In: Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, 11071. Springer, Cham. https://doi.org/10.1007/978-3-030-00934-2_20

42. Chen, ZM, Wei XS, Wang P, et al (2019) Multi-label image recognition with graph convolutional networks. in 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/CVPR.2019.00532

43. Xu C, Qi S, Feng J et al (2020) DCT-MIL: Deep CNN transferred multiple instance learning for COPD identification using CT images. Phys Med Biol 65(14):145011. https://doi.org/10.1088/1361-6560/ab857d

Kewu Huang Received his medical doctor's degree from Capital Medical University in 1998. Prof. Huang is currently a professor and a doctoral supervisor of the Department of Respiratory Disease at Capital Medical University. His research interests are diagnosis and management of chronic obstructive pulmonary disease. Prof. Huang serves on the editorial boards of several journals, as well as committees for a number of international conferences.

Ligong Liu Earned a PhD in mathematics from Chern Institute of Mathematics, NankaiUniversity, Tianjin, China in 2010. He has experience and expertise in utilizing deep learningapproach with big data for predictions.

Zuoqing Zhang Graduated from Harbin Medical University with a bachelor's degree ininternal medicine in 1987. She is currently the director of respiratory department of Shijingshanteaching hospital of Capital Medical University. Her research interests are diagnosis andcomprehensive treatment of internal disease.

Zongli Li Graduated from Capital Medical University with a master's degree in internalmedicine in 2011. She is currently a M.D. student of the department of respiratory disease atCapital Medical University. Her research interests focus on the application of artificialintelligence in chronic obstructive pulmonary disease.