Automatic Diagnosis of COPD in Lung CT Images based on Multi-View DCNN

Yin Bao[1], Yasseen Al Makady[1] and Sasan Mahmoodi[1]
[1]*School of Electronics and Computer Science , University of Southampton, University Road, Southampton, UK*
{*yb1a19, yham1n15, sm3y07*}*@soton.ac.uk*

Keywords: COPD, Deep Convolutional Neural Network, Multi-View, Classification

Abstract: Chronic obstructive pulmonary disease (COPD) has long been one of the leading causes of morbidity and mortality worldwide. Numerous studies have shown that CT image analysis is an effective way to diagnose patients with COPD. Automatic diagnosis of CT images using computer vision will shorten the time a patient takes to confirm COPD. This enables patients to receive timely treatment. CT images are three-dimensional data. The extraction of 3D texture features is the core of classification problem. However, the classification accuracy of the current computer vision models is still not high when extracting these features. Therefore, computer vision assisted diagnosis has not been widely used. In this paper, we proposed MV-DCNN, a multi-view deep neural network based on 15 directions. The experimental results show that compared with the state-of-art methods, this method significantly improves the accuracy of COPD classification, with an accuracy of 97.7%. The model proposed here can be used in the medical institutions for diagnosis of COPD.

1 INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a disease in which changes in bronchial epithelial cells cause inflammation of the lungs, resulting in decreased lung function (Lapperre et al., 2007). The main clinical manifestation of this disease is continuous irreversible airflow limitation or obstruction of lung tissue due to functional small airway disease, atmospheric disease and destruction of lung tissue (emphysema) (Han et al., 2010). Individuals with this disease frequently develop complications such as muscular atrophy, cardiovascular disease, osteopenia and chronic infections. COPD has long been one of the leading causes of morbidity and mortality worldwide (Mannino and Buist, 2007). COPD will also be one of the leading causes of death in humans in the future. Currently, there are two mainstream diagnostic methods used to diagnose COPD. The first one is the Pulmonary Function Test (PFT) whereas the other is by Computed Tomography (CT) medical image. The PFT includes spirometry, maximal respiratory pressures and diffusing capacity. However, a major problem with PFT is that it is not sensitive to the early diagnosis of COPD (Dirksen et al., 1998). CT is an alternative to PFT and has proven to be an effective method for detecting COPD. So far, the dominant diagnostic approach has been for professional physicians to analyze the specific manifestations of emphysema by looking at high-resolution CT images. However, not all patients have the opportunity to be examined by a professional doctor due to the cost of medical treatment. Therefore, automatic analysis by computer may greatly improve the accuracy of diagnosis, reduce the threshold of patient detection and release the energy of professional physicians.

Computer vision approaches have significant improvements in many fields including medical image analysis. At present, most of the methods used for medical diagnosis are in the form of two-dimensional (2D) image analysis. However, extracting features from 2D images could lead to a low accuracy of computer diagnosis, because 2D images do not contain the spatial structure characteristics of the object (depth information). In order to improve the accuracy of medical diagnosis by computer, it is not enough to use the feature extraction technology of 2D images only. Therefore, the successful extraction of texture features from 3D data may contribute in the performance improvement of computerized medical diagnosis.

In this paper, we propose a method that extracts the features of 3D medical image data for COPD detection. Meanwhile, this method is also used to realize the automatic diagnosis of COPD patients with high accuracy.

2 RELATED WORK

There are currently two methods to classify 3D data: one method is to directly use 3D volume data based on a 3D Convolutional Neural Network (3DCNN) to carry out the problem. Another is to use multi-view 2D images to classify on a 2D convolutional neural network.

2.1 3D CNN

Alakwaa et al. (Alakwaa et al., 2017) proposed the method of using 3DCNN to classify 3D data. They detected data points of interest based on the U-Net model (Ronneberger et al., 2015) and used for classification. Similarly, Zhu et al. (Zhu et al., 2018) proposed a deep three-dimensional Dual-Path network (DPN). This method divided the classification network into two parts; the pathological feature detection and classification and used the 3D Fast R-CNN and U-Net to perform the classification. Dey et al. (Dey et al., 2018) proposed 3D Multi-Output DenseNet (MoDenseNet) that used two 3D images with different input scales to cover local and global images of pulmonary nodules. However, a major problem with 3D CNN is high computation cost. Compared with 2D CNN, it requires to calculate one more dimension, which leads to convolutions with high computational costs. Shen et al. (Shen et al., 2017) proposed a new Multi-Crop Convolutional Neural Network (MC-CNN) with a new pooling method to replace the traditional maximum pooling layer. Their aim is to extract multi-scale features from the feature map extracted from the convolution layer. This method relatively decreases the computation required time to obtain the multi-scale features. However, the biggest disadvantage of using 3DCNN is that it requires a huge amount of computation, which stretches the computational power of computers to its limits. This shortcoming directly leads to the research results which may not be useful in to the medical field. Secondly, the disadvantage of 3DCNN is that not all parts of 3D data have contributing features. This results in a waste of computing resources if a full convolution operation is deployed. Therefore, employing the Multi-View 2D approach may greatly decrease the computational time without compromising the classification performance.

2.2 Multi-View 2D

Hatt et al. (Hatt et al., 2018) intercepted 8 slices of the same size from the 3D volume data along the spindle direction of the high-resolution CT image. The eight slices were combined into one montage image. Finally, it is used as the input of a five-layer convolutional neural network. They demonstrate that combining multiple slices of 3D data into a single and compact shape descriptor has higher recognition performance than a single image recognition architecture. However, using only eight multi-angle images may not capture all the features of a 3D medical image. This may be one of the reasons for the low classification accuracy. Similarly, Liu and Kang (Liu and Kang, 2017) proposed a Multi-View Convolutional Neural Network (MV-CNN). The whole lung CT images were cropped into multi-view and multi-scale images. These images are stacked together into the neural network. However, this method only considered the scale invariance of the multi-perspective images, and ignored the effect of rotation and translation on the results. Gomez-Donoso et al. (Gomez-Donoso et al., 2017) presented the LochaNet method, which loads 3D volume data and calculates the center point of each axis. Then the slices with model size of 5% were obtained on the XY, XZ and YZ planes respectively. They are projected onto a plane, producing a 500×500 pixel image.

However, the main disadvantage of these methods is that they could not effectively extract the features of 3D volume data. The possible reasons may be that the number of network layers is too small or the main features of 3D volume data cannot be well represented by the intercepted images.

In this paper, we proposed a 15-direction multi-view deep convolutional neural network model for the diagnosis and classification of COPD patients. We first divide the 3D data of each lung into 15 views, which is shown in Figure 1. Then the 15 views are input into the multi-input anti-aliased ResNet18 (He et al., 2016) Pre-Training Model. Finally, back propagation is used to reduce the classification error so as to achieve the purpose of accurate classification. Different from the previous network, the multi-view we extracted is more comprehensive. In addition, we train the neural network by adding rotation, translation and multi-scale data enhancement methods.

3 METHOD

The multi-View DCNN algorithm is mainly composed of three steps: The first step is to extract images from 15 perspectives from three-dimensional data. The second step is to enhance the data of these 15 views respectively. The final step is to construct 15 Multi-View DCNN (MV-DCNN) models to extract and classify the features.

3.1 Multi-View Image Extraction

First of all, the experimental data are clipped into cubes of $64 \times 64 \times 64$ size. The reason the size of the data is processed to $64 \times 64 \times 64$ is not only that the characteristic information can be retained in large quantities, but also that the computing power of the computer is not burdened. Although the larger the size of the data, the more features it posses. The large amount of data may cause the computer to run out of memory or compute slowly. Next, we extract 15 slices from the 3D experimental data. The location of multi-view image extraction can be seen in the Figure 1. The blue cube represents the cube data with the size of $64 \times 64 \times 64$. The red part indicates the location of the 2D slice extracted from the 3D data. This process provides us 15 slices (View 1-15) in each cube. The purpose of this is to enrich the depth characteristic information of the direction, so as to make the extracted 3D data features more comprehensive.

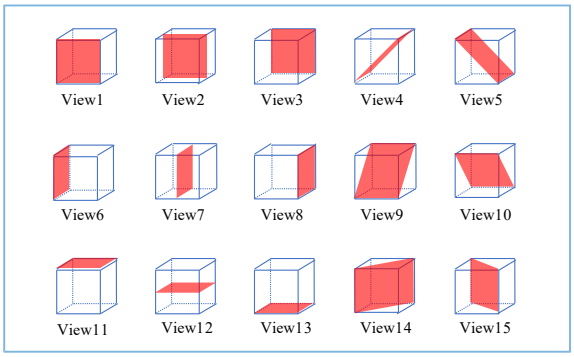

Figure 1: Principle diagram of multi-view image extraction.

3.2 Data Agumentation

One of the reasons computer vision has not been widely adopted in the medical field is the small number of data sets used for training. Due to the small size of the datasets, a neural network model usually suffers from the problem of overfitting. The result of overfitting problem is that the training set may have high accuracy. However, the accuracy of the test set is much lower than that of the training set. This is because small data sets are insufficient to describe the true distribution of the problem. Therefore, we solve the overfitting problem by adding data enhancement. Each data is enhanced by means of subsampling, rotation, translation, Gaussian blur, and noise addition. The specific methods are as follows:

- **Noise**: Gaussian noise is added to slices (View 1-15).
- **Rotation**: The function (**torchvision.transforms.RandomAffine**) of PyTorch is used to rotate the slices (View 1-15) randomly.
- **Translation**: The function (**torchvision.transforms.RandomAffine**) of PyTorch is used to translate the slices (View 1-15) randomly.
- **Blur**: The function (**ImageFilter.GaussianBlur**) of PIL is used to translate the slices (View 1-15) randomly.
- **Subsampling**: Randomly select a 32×32 region from the entire slice and tile it into a 64×64 slice.

The above operation is performed 50 times in each training. The volume of data has been expanded 200 times. The advantage of this is that it not only increases the amount of data, but also improves the robustness of the data in terms of translation, blur, and noise.

3.3 MV-DCNN Model

The whole feature extraction network is based on the Anti-aliased ResNet18 (Zhang, 2019). We adopt 15 Anti-aliased ResNet18 models and the Classification layer to construct the MV-DCNN. The specific network structure diagram is shown in the Figure 2. Their network structure and principles are explained in details in the following sections.

3.3.1 Anti-aliased ResNet18

Anti-aliased ResNet 18 is a shift-invariant network structure, which is combination of the ResNet 18 network architecture and BlurPool. Its structure can be seen in Figures 2 (c) and (d). At present, most of the networks in the convolutional networks is non shift-invariant. That is, a small translation of the input image can greatly change the output of the neural network, thus affecting the experimental results. The anti-aliasing CNN model can enhance the robustness of the network to image translation and improve the classification accuracy. The details of principle of ResNet18 (He et al., 2016) and BlurPool are described in the following sections.

3.3.2 BlurPool

Antialiasing CNN model is an improvement of ResNet model. Because ResNet is sensitive to shift. Thus, BlurPool (Zhang, 2019) is used instead of Maxpool to reduce aliasing effects. The principle is shown in the Figure 3. Different from the traditional pooling layer, the BlurPool layer uses Gaussian kernel to convolute with the image. Since the size of stride is set

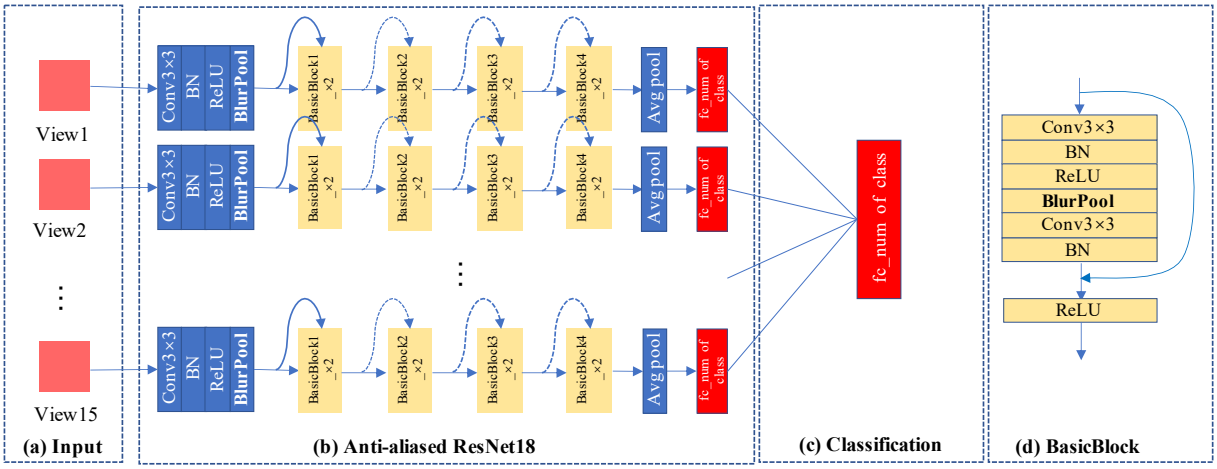

Figure 2: The network structure of MV-DCNN.

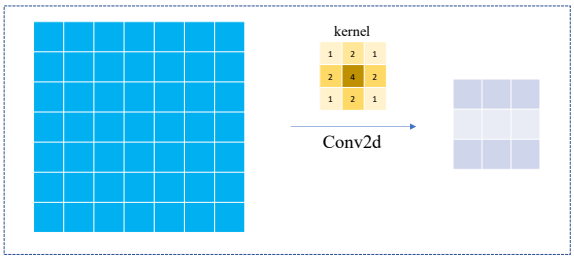

Figure 3: The principe of BlurPool.

to 2, the size after the convolution is the same as that of the traditional pooling layer. However, the difference is that the use of Gaussian convolution kernel can reduce the effects of aliasing, so as to reduce the impact of image shift on the output. Not only does Anti-aliased ResNet 18 replace the pooling layer, it also adds the BlurPool layer to the residual block.

3.3.3 Classification Layer

Each anti-aliased ResNet18 model has two outputs. We add a classification layer after the output of the 15 models. This classification layer consists of a full connection layer. The input is 30 channels and the ouput is 2 channels. The function of adding a classification layer is to vote on 15 models to get the optimal classification result.

Name of Category	Number
Fourier	15
Geometric	25
Interpolated	30
Mixed texture	25

Table 1: The number of the classes

4 Method evaluation

4.1 Dataset

The performance of this model is evaluated on the RFAI dataset. The database of RFAI is one of the synthetic texture datasets that can be used to evaluate the effectiveness of 3D texture feature classification methods (Paulhac et al., 2009). The 3D volume data of this database is composed of 64 2D images.

The database has a total of four ways to generate textures. The first method is to use Fourier transform method to synthesize volume texture, as shown in the Figure 4 (a). The second method is to synthesize the volume texture using the 3D geometry, as shown in the Figure 4 (b). The third method is to interpolate two or more texture images to synthesize volume texture, as shown in the Figure 4 (c). The final method is to combine the first three methods to synthesize the volume texture, as shown in the Figure 4 (d). The number of classes contained in each category is shown in the Table 1. According to the

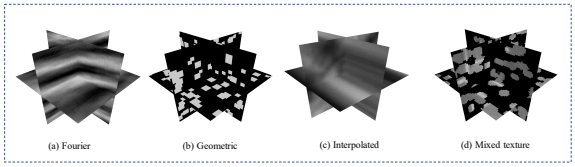

Figure 4: Examples from four categorizes.

different types of distortion, each class is divided into 4 categories: normal texture (a), Gaussian fuzzy texture (b), subsampled texture (c), and Gaussian noise texture (d). The schematic diagram is shown in the Figure 5. Each of the four texture types contains 10 volumetric images of $64 \times 64 \times 64$ size.

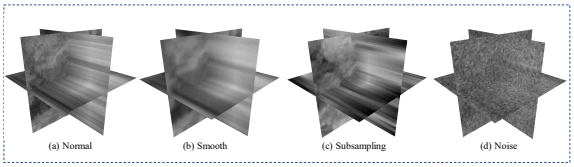

Figure 5: Example of volumetric textures synthesized by four methods. These four samples are from the texture of the stone in the interpolation class.

4.2 Implementation

The models of this experiment are all written with the open source Pytorch language framework. All models in this experiment are run on a Google Colab cloud server with NVIDIA Tesla 100 and 25 GB of RAM. The batch size of all models is 64. The filter size of Blurpool is set to 3. The initial learning rate is 0.001, and the learning rate decay is carried out once every ten times of training, with the decay rate of 0.5. The lower limit of attenuation rate is 0.00001. In the training, we use Adam optimization algorithm to optimize each model, and its parameters are default values. The loss function of the model adopts cross entropy. The pre-training model was used for training. Consequently, during the training we freeze the first seven layers of the model. All experiments in this paper have been carried out for 10 times. All parameter settings are shown in the Table 2.

Our proposed MV-DCNN model is tested on four types of RFAI datasets respectively. The parameters of each dataset are fine-tuned and the performance of the model is evaluated by its accuracy. Accuracy shows the classification performance of the model in each category. The ratio of training set and test set in this data set is 1:1. Parameter settings for each type of dataset are described below.

The data for each type of dataset is cropped by the center into 32×32 images, because rotation causes edge distortion. Each data is performed 50 times for data enhancement, all of which are added to the training set.

Interpolation: In this data set, we perform 5 ways of data enhancement in the training set. The first is random translation. The horizontal and vertical translation ranges are (-0.1*image width, 0.1*image width). The second is to add random Gaussian noise to the whole picture. The third is random Gaussian blur, whose standard deviation varies in a range of (0.8,1). The fourth and fifth are rotated randomly twice. The rotation ranges are (-120, 120) and (-90, 90), respectively. The final classification number is set to 30.

Fourier: In this data set, we perform data enhancement for the training set in six ways. Random rotation is added in the first and the second, and the rotation angle ranges are (-120, 120) and (-90, 90), respectively. The third is random subsampling. The remaining three are random translation, random Gaussian blur, and random Gaussian noise. They have the same parameter setting as the Interpolation. The number of categories is set to 15.

Geometric: In this data set, we add 5 ways of data enhancement to the training set. In the first and the second type respectively add the random rotation. The rotation angle range is (-45, 45) and (-90, 90). The remaining three data augmentation are random translation, random Gaussian blur, and random Gaussian noise. They have the same parameter setting as the Interpolation. The final classification number is set to 25.

Mixed Texture: All parameter settings are the same as in Geometric.

5 Result and Discussion

We evaluate all four categories of the datasets and compare them with the state of the art methods. In addition, this experiment considers all subsets of each category, including normal subset, rotation subset, noise subset, smoothing subset and subsampling subset. We take the normal subset of each category as the training set and test the remaining subset of the corresponding data set. The results are shown in Table 3. The accuracy of correct classification of the data in the table indicates that the MV-DCNN model is superior to the other two methods in terms of robustness to these three kinds of disturbances. Geometric regularity in this class was somewhat lower than in $GMRF_{42,1}^{3Dri}$ methods. However, under the interference of Gaussian blur and subsampling, the performance of MV-DCNN model is significantly better than the other two methods. The proposed method has good adaptability to each kind of data set.

6 Application to COPD detection

In this experiment, the method proposed here is employed COPD detection. The COPD dataset in this paper is derived from 32 subjects. Thirteen subjects are diagnosed with COPD, and the remaining 19 subjects are healthy. The size of the HRCT volume image is $512 \times 512 \times 512$. Each sample is associated with a mask image of the entire lung to keep only the lung and discard the surrounding tissues. The CT data used in this work are acquired as a part of a study into the application of imaging to the characterization of the phenotypes of COPD. The written informed consent was given and signed by all subjects. The study was

Categories of Data Enhancement	Random Rotation1	Random Rotation2	Random Gaussian Blur	Random Translation
Fourier	(-120, 120)	(-90, 90)	(0.8, 1.0)	(0.1, 0.1)
Geometric	(-45, 45)	(-90, 90)	(0.8, 1.0)	(0.1, 0.1)
Interpolated	(-120, 120)	(-90, 90)	(0.8, 1.0)	(0.1, 0.1)
Mixed texture	(-45, 45)	(-90, 90)	(0.8, 1.0)	(0.1, 0.1)
COPD	-	-	(0.95, 1.0)	(0.01, 0.01)

Table 2: Experimental data parameter setting.

Synthetic texture dataset						
Descriptor	Fourier			Geometric		
	Noise	Smooth	Subsampling	Noise	Smooth	Subsampling
MV-DCNN	**100**	**96.5**	**56.7**	96.0	**84.2**	**33.8**
$GMRF_{42,1}^{3Dri}$	100	72.0	47.0	**100**	82.8	28.4
3DRiesz	100	87.0	45.0	96.0	42.0	21.0

(a) Result of Fourier and Geometric.

Synthetic texture dataset						
Descriptor	Interpolated			Mixed texture		
	Noise	Smooth	Subsampling	Noise	Smooth	Subsampling
MV-DCNN	**94.8**	**99.5**	**61.8**	**100**	**96.4**	**32.2**
$GMRF_{42,1}^{3Dri}$	93	78.59	49.16	99.6	92	26.8
3DRiesz	-	-	-	-	-	-

(b) Result of Interpolated and Mixed texture.

Table 3: Classification accuracies [%] of our method and some other methods using RFAI datasets.

approved by the Southampton and South West Hampshire local research ethics committee (LREC number: 09/H0502/91) and the University Hospital Southampton Foundation Trust Research and Development Department. The study was conducted in the Southampton NIHR Respiratory Biomedical Research Unit.
Because the amount of data is too small and the individual data is too large. Each whole lung HRCT image is divided into 16 volumes of $64 \times 64 \times 64$ size. The collection location of each data is randomly selected from the following figures 6 (a) and (b).

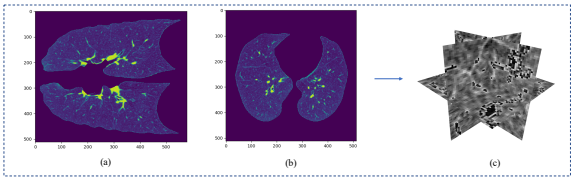

Figure 6: Figure (a) shows the YZ plane of the lung. Figure (b) shows the XY plane of the lung. Figure (c) shows a sample of randomly segmented lung data.

6.1 The result of MV-DCNN on COPD dataset

After the pre-processing step is performed, we feed the data into our proposed method. the ratio of training set and test set is randomly assigned to train and test at a ratio of 8:2. The model is trained for 10 times, and the training set and test set are randomly assigned at a ratio of 8:2. Then the average of the accuracy is calculated. Three methods of data enhancement are added to the COPD dataset: random translation (horizontal and vertical translation range (-0.01*width, 0.01*width)), random Gaussian noise, and random Gaussian blur. These three enhancements are run 30 times per data. The number of categories is set to 2, healthy and COPD. We compare the performance of the MV-DCNN model with texture-based methods deep learning-based methods.

The classification performance is reported as classification accuracy, sensitivity, and specificity.
The results of MV-DCNN tests on the COPD dataset are shown in the Table 4. According to the results

Confusion Matrix	True	False
Positive	42	2
Negative	59	0

(a) Confusion Matrix

Result (%)	Accuracy	Specificity	Sensitivity
MV-DCNN	98.05	96.72	100

(b) Results

Table 4: The data results in the table are one out of ten random sample classification training sessions. The confounding matrix, accuracy, specificity, and sensitivity of the test results are recorded in the table.

	Method	Classification accuracy [%]
A	**MV-DCNN**	**97.7**
B	$GMRF_{42,1}^{3Dri}$	90.25
C	LPH	75.0
D	LBP	78.12
E	3D GLCM	75
F	Gabor filters	78.12
G	SIFT	75
H	Intensity features method	68.75
I	Density-based method	71.88

Table 5: Classification accuracies [%] of our method and some other methods using COPD datasets.

	Method	Classification accuracy [%]
A	**MV-DCNN**	**97.7**
B	Voting CNN	88.2
C	ResNet-50	87.5
D	AlexNet	85.94
E	VoxResNet 3D	74.3

Table 6: Comparison of our method with deep learning-based methods on the COPD dataset.

in Table 4 (b), the specificity of this model is 96.2%. This suggests that the model has the potential to diagnose healthy individuals as COPD. However, the probability of such an error is only 3.8%. The sensitivity of the model to the diagnosis of COPD has reached 100%. This means that the model can diagnose COPD accurately. Therefore, from the test results, our model can be used in the COPD detection of medical institutions.

6.2 Comparison to texture based methods

This section compares our approach with five methods based on texture feature extraction. They are respectively $GMRF_{42,1}^{3Dri}$ (Almakady et al., 2020) , the local parameter histogram (LPH) (Dharmagunawardhana et al., 2016), 3D Gray Level Co-occurrence Matrix (3DGLCM) (Han et al., 2015), Gabor filters (Manjunath and Ma, 1996) and the SIFT (Xu et al., 2012) method. The result of these methods are reported in the literature. $GMRF_{42,1}^{3Dri}$ and SIFT are both rotation and scale invariant descriptors for texture classification. For each slice, they detect a number of key points unrelated to rotation. Then the 128-dimensional eigenvectors are constructed with these key points and the classification task is performed. $GMRF_{42,1}^{3Dri}$ has better rotation invariance from the method B and method G in Table 5. However, method A in Table 5 proposed here is significantly superior to these five methods, with an accuracy rate of 97.7%. This suggests that the manual extraction method may still be missing some information. This results in poor overall distinguish ability despite good performance on some invariance. On the contrary, our method has a strong comprehensive ability and can well extract the pathological characteristics of COPD.

6.3 Comparison with Deep-Learning-Based methods

In this section, we compare our method with four methods based on deep neural network. They are ResNet 3D (He et al., 2016), AlexNet (Krizhevsky et al., 2012), VoxResNet 3D (Zhu et al., 2018), and Voting CNN (Du et al., 2020). All experimental results are shown in the Table 6. The experimental results show that our method has the highest accuracy and is significantly higher than the second one. Methods C, D, and E in the Table 6 directly convolvs 3D data. There are two possible reasons for their inaccuracy. One is that such operations require the sacrifice of data size to satisfy the computing power of the computer. As a result, the features carried by the data is not enough, which makes it impossible to extract contributing features for classification. Another reason is that they use less data and do not enhance the data, which can lead to over-fitting of the model. Method B in Table 6 also uses the method of multi-view. However, it may be that the author only considers Air Tree and the perspective is not comprehensive enough, leading to low accuracy. In conclusion, MV-DCNN model has a high accuracy rate and is higher than the current state-of-art methods. These results suggest that our proposed method can assist physicians in the medical diagnosis of COPD and can be applied to the diagnosis of off-line medical institutions.

7 CONCLUSIONS

This paper proposes the MV-DCNN model COPD diagnosis. In this paper, the concrete implementation method of MV-DCNN is introduced in details. Finally, we introduce the whole experiment process and performance of the model. Based on deep neural network, MV-DCNN classifies 3D texture data by using 15 multi-view slices of 3D data. Then, the accuracy of classification is successfully improved through various methods of data enhancement and transfer learning. The accuracy of our MV-DCNN is 97.7% for the diagnosis of COPD, which has good classification performance compared with the state of the art method in this paper.

REFERENCES

Alakwaa, W., Nassef, M., and Badr, A. (2017). Lung cancer detection and classification with 3d convolutional neural network (3d-cnn). *Lung Cancer*, 8(8):409.

Almakady, Y., Mahmoodi, S., Conway, J., and Bennett, M. (2020). Rotation invariant features based on three dimensional gaussian markov random fields for volumetric texture classification. *Computer Vision and Image Understanding*, 194:102931.

Dey, R., Lu, Z., and Hong, Y. (2018). Diagnostic classification of lung nodules using 3d neural networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 774–778. IEEE.

Dharmagunawardhana, C., Mahmoodi, S., Bennett, M., and Niranjan, M. (2016). Rotation invariant texture descriptors based on gaussian markov random fields for classification. *Pattern Recognition Letters*, 69:15–21.

Dirksen, A., Holstein-Rathlou, N.-H., Madsen, F., Skovgaard, L. T., Ulrik, C. S., Heckscher, T., and Kok-Jensen, A. (1998). Long-range correlations of serial fev1 measurements in emphysematous patients and normal subjects. *Journal of applied physiology*, 85(1):259–265.

Du, R., Qi, S., Feng, J., Xia, S., Kang, Y., Qian, W., and Yao, Y.-D. (2020). Identification of copd from multi-view snapshots of 3d lung airway tree via deep cnn. *IEEE Access*, 8:38907–38919.

Gomez-Donoso, F., Garcia-Garcia, A., Garcia-Rodriguez, J., Orts-Escolano, S., and Cazorla, M. (2017). Lonchanet: A sliced-based cnn architecture for real-time 3d object recognition. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 412–418. IEEE.

Han, F., Wang, H., Zhang, G., Han, H., Song, B., Li, L., Moore, W., Lu, H., Zhao, H., and Liang, Z. (2015). Texture feature analysis for computer-aided diagnosis on pulmonary nodules. *Journal of digital imaging*, 28(1):99–115.

Han, M. K., Agusti, A., Calverley, P. M., Celli, B. R., Criner, G., Curtis, J. L., Fabbri, L. M., Goldin, J. G., Jones, P. W., MacNee, W., et al. (2010). Chronic obstructive pulmonary disease phenotypes: the future of copd. *American journal of respiratory and critical care medicine*, 182(5):598–604.

Hatt, C., Galban, C., Labaki, W., Kazerooni, E., Lynch, D., and Han, M. (2018). Convolutional neural network based copd and emphysema classifications are predictive of lung cancer diagnosis. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 302–309. Springer.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Lapperre, T. S., Sont, J. K., van Schadewijk, A., Gosman, M. M., Postma, D. S., Bajema, I. M., Timens, W., Mauad, T., Hiemstra, P. S., Group, G. S., et al. (2007). Smoking cessation and bronchial epithelial remodelling in copd: a cross-sectional study. *Respiratory research*, 8(1):85.

Liu, K. and Kang, G. (2017). Multiview convolutional neural networks for lung nodule classification. *International Journal of Imaging Systems and Technology*, 27(1):12–22.

Manjunath, B. S. and Ma, W.-Y. (1996). Texture features for browsing and retrieval of image data. *IEEE Transactions on pattern analysis and machine intelligence*, 18(8):837–842.

Mannino, D. M. and Buist, A. S. (2007). Global burden of copd: risk factors, prevalence, and future trends. *The Lancet*, 370(9589):765–773.

Paulhac, L., Makris, P., Ramel, J.-Y., et al. (2009). A solid texture database for segmentation and classification experiments. In *VISAPP (2)*, pages 135–141.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Shen, W., Zhou, M., Yang, F., Yu, D., Dong, D., Yang, C., Zang, Y., and Tian, J. (2017). Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognition*, 61:663–673.

Xu, Y., Huang, S., Ji, H., and FermüLler, C. (2012). Scale-space texture description on sift-like textons. *Computer Vision and Image Understanding*, 116(9):999–1013.

Zhang, R. (2019). Making convolutional networks shift-invariant again. In *ICML*.

Zhu, W., Liu, C., Fan, W., and Xie, X. (2018). Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 673–681. IEEE.