



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Deep feature fusion classification network (DFFCNet): Towards accurate diagnosis of COVID-19 using chest X-rays images

Jingyao Liu^{a,b}, Wanchun Sun^a, Xuehua Zhao^c, Jiashi Zhao^{a,*}, Zhengang Jiang^{a,*}

^a School of Computer Science and Technology, Changchun University of Science and Technology, Changchun, Jilin 130022, China

^b School of Computer and Information Engineering, Chuzhou University, Chuzhou 239000, China

^c School of Digital Media, Shenzhen Institute of Information Technology, Shenzhen 518172, China

ARTICLE INFO

Keywords:

COVID-19
Deep learning
Classification
Feature fusion
SVM

ABSTRACT

The widespread of highly infectious disease, i.e., COVID-19, raises serious concerns regarding public health, and poses significant threats to the economy and society. In this study, an efficient method based on deep learning, deep feature fusion classification network (DFFCNet), is proposed to improve the overall diagnosis accuracy of the disease. The method is divided into two modules, deep feature fusion module (DFFM) and multi-disease classification module (MDCM). DFFM combines the advantages of different networks for feature fusion and MDCM uses support vector machine (SVM) as a classifier to improve the classification performance. Meanwhile, the spatial attention (SA) module and the channel attention (CA) module are introduced into the network to improve the feature extraction capability of the network. In addition, the multiple-way data augmentation (MDA) is performed on the images of chest X-ray images (CXRs), to improve the diversity of samples. Similarly, the utilized Grad-CAM++ is to make the features more intuitive, and the deep learning model more interpretable. On testing of a collection of publicly available datasets, results from experimentation reveal that the proposed method achieves 99.89% accuracy in a triple classification of COVID-19, pneumonia, and health X-ray images, there by outperforming the eight state-of-the-art classification techniques.

1. Introduction

COVID-19 has already caused over 5.7 million casualties and infected more than 396 million people, as of February 8, 2022 [1]. Since its appearance in December 2019, it has spread throughout the globe, which has forced countries to take drastic measures, including closing borders, canceling flights and quarantining people in countries with related cases, and containing the virus spread appears to be a challenging task [2]. Owing to the critical health risks associated with it, COVID-19 was declared by the World Health Organization (WHO) as an international public health emergency and pandemic on 30/01/2020 and 11/03/2020, respectively. And new mutated strains continue to emerge.

The common symptoms of COVID-19 include fever, cough, shortness of breath, and pneumonia [3], and it affects the human heart, brain, liver and many other organs and requires prompt detection and treatment. It relies primarily on real-time reverse transcription polymerase chain reaction (PCR) for its determination, however, this method takes a longer time to detect. X-ray imaging is a cheaper, faster, and readily

available method, where the body gets exposed to a much smaller amount of harmful radiation compared to CT [4]. Chest X-ray imaging (CXRs) is widely used as an assistive diagnostic tool in COVID-19 screening, and it is reported to have high potential prognostic capabilities [5]. However, the diagnosis throughput of human experts is not comparable with that of machines, while early symptoms are difficult to spot and may be overlooked by human experts [6]. Therefore, there exists an urgent need to develop a smarter and more accurate algorithm for assisting to detect diseases automatically (e.g. COVID-19).

In recent years, there has been an increasing amount of research on the application of artificial intelligence to disease diagnosis. For instance, A. Esteva et al. [7] trained a CNN on fine-grained skin cancer images, and obtained the results, which were generally consistent with the expert judgment. In another work, M. Guo et al. [8] used a deep learning approach to classify thyroid images with better results, achieving an accuracy of 83.88%. Likewise, Lu et al. [9] introduced wavelet transform and extreme learning machine techniques to predict the healthy or abnormal brain MRI pictures, with an accuracy of 97.04%. Others are the early prognostication of Alzheimer's disease

* Corresponding authors.

E-mail addresses: zhaojiashi@cust.edu.cn (J. Zhao), jiangzhengang@cust.edu.cn (Z. Jiang).

<https://doi.org/10.1016/j.bspc.2022.103677>

Received 31 August 2021; Received in revised form 22 March 2022; Accepted 9 April 2022

Available online 13 April 2022

1746-8094/© 2022 Elsevier Ltd. All rights reserved.

dementia [10], diagnosis of brain hemorrhage [11], detection of diabetic retinopathy [12], identification of arrhythmia [13], and classification of various types of cancer (e.g. breast [14], brain [15], and prostate [16]). With the wide application of deep learning, its research in the diagnosis of lung diseases is also increasing. For instance, Chen et al. [17] proposed the DualCheXNet which used ResNet and DenseNet to the extract features, and used weighting of multiple classifiers for classification of fourteen diseases of the lung. Gundel et al. [18] proposed a location-aware dense network to improve the accuracy of thoracic disease classification by using high-resolution image data and spatial information of the lesion to accurately find the lesion.

Most of the COVID-19 studies use a single network to extract features, but different networks acquire features in different ways and therefore focus on different regions, and fusing multiple networks can make features richer. Therefore, we fuse the features extracted by EfficientNetV2 and ResNet. Deep learning has a strong advantage in feature extraction, while SVM classification is a proven machine learning method; we combine the two to achieve better classification results. Although our network has a strong learning ability, it is poorly interpretable. People cannot perceive small changes in grayscale, but they can perceive color changes better, so we introduce a color visualization method that can present the results of network learning well.

The main contributions of this work are as follows:

- We proposed a deep feature fusion classification network (DFFCNet), and introduced two modules: deep feature fusion module (DFFM) and multi-disease classification module (MDCM).
- EfficientNetV2 was introduced as the backbone network to fuse with the features extracted by ResNet101. The spatial attention (SA) module and the channel attention (CA) module are introduced into the network.
- We used four multiple-way data augmentation (MDA) ways to enhance the training set. To simplify the interpretation of proposed deep learning model, a color visualization approach is employed via the Grad-CAM++ technique.
- We used the same dataset for our experiments. Compared to the 8 state-of-the-art diagnosis methods for COVID-19, experimental results from this work illustrate very good results achieved through the DFFCNet.

The structure of this paper is organized as below. In Section 2, we summarize the current state of research on COVID-19. In Section 3, introduces the dataset, the involved deep learning methods and the proposed new model. In Section 4, we describe the experimental steps and results. In Section 5, finally concludes this paper.

2. Related works

The sudden appearance of COVID-19 has led many researchers to propose various artificial intelligence methods to study it. These artificial intelligence methods are divided into three categories. First, deep learning networks such as DenseNet, AlexNet, ResNet and Xception are used for disease diagnosis, and transfer learning can be used to reduce the network parameters. Second, weakly supervised learning or unsupervised methods are used to solve the problem of small labeled samples, and machine learning such as clustering or support vector machines are used as classifiers in order to improve the recognition accuracy. Third, methods such as U-Net are used to segment the lesions.

Some researchers use transfer learning for training because it allows higher initial performance of the network, faster training rate, and better convergence of the obtained model, which can reduce the network parameters and make the network small. For instance, Narayan et al. [19] used transfer learning to pre-train Inception (Xception) parameters first on a large dataset and then applied them to the COVID-19 dataset to automatically diagnose diseases. Majeed et al. [20] proposed a new network, named CNN-X, which had fewer parameters and suited for

smaller datasets. Maghdid et al. [21] used transfer learning to introduce AlexNet and proposed a simple CNN network. Experiments were performed on collected X-ray and CT images with an accuracy of 98%. Katsamenis et al. [22] proposed a simple CNN, which used transfer learning to introduce ResNet-50, changed the last fully connected layer. Pre-trained on ImageNet and achieved better classification results. Montalbo [23] introduced DenseNet as the backbone network through transfer learning, optimized migration learning by freezing some layers and adding a new layer to improve performance, a method called Fused-DenseNet-Tiny, who was able to achieve 97.99% classification accuracy.

At the beginning of the outbreak of COVID-19, due to the lack of samples, many researchers proposed many models suitable for small amount of data from this perspective. For example, Aradhya, et al. [24] proposed a new model for one-time learning based on the idea of clustering, introducing two classifiers GRNN (Generalized Regression Neural Network) and PNN (Probabilistic Neural Network). Voulodimos et al. [25] proposed a new online learning model for COVID-19 based on U-Net network, called few-shot driven U-Net. It can learn features of small datasets and accurately segment COVID-19 lesion regions in CT images. Chen et al. [26] developed an end-to-end trainable deep few-shot learning framework in the shortage of annotated COVID-19 CT images in order to save computational costs. It can expand one image into multiple images to accurately diagnose diseases. Yang et al. [27] proposed a new semi-supervised learning network based on less labeled images, which can be applied to new datasets with better generalization performance based on the disease features learned on a limited dataset.

In the process of learning disease features, deep learning often learns some features that are not related to the disease, and the obtained model has poor generalization ability. In order to get accurate lesion regions for their study, the segmentation task is necessary, and many researchers have proposed many methods for this problem. Among them, Voulodimos et al. [28] proposed a lightweight segmentation model using U-Net and FCN (Fully Convolutional Neural Networks). The model can be trained without GPU, meaning that it can be run on a PC (personal computer) without parallel computing capabilities. Chen et al. [29] improved U-Net by adding an attention mechanism, and the 10-fold cross-validation results showed a 10% improvement in segmentation performance compared to the traditional U-Net. Saeeidzadeh et al. [30] proposed a new model with a new regularization term in U-Net, and the segmentation performance was improved by 2%, and this model is called TV-Unet. Zhou et al. [31] added the spatial attention module and the channel attention module to U-Net, effective feature relations can be obtained. Meanwhile, the dice loss was changed to the focal tversky loss, the obtained model takes only 0.29 s to segment a CT. Chen et al. [32] proposed an unsupervised segmentation network with synthetic data and limited labeled data, which can guide the segmentation network to perform cross-domain learning and improve the segmentation performance. Liu et al. [33] used transfer learning twice for accurate segmentation of COVID-19 lesions, and proposed nCoVSegNet. Due to the small amount of labeled data, the model parameters were first trained on ImageNet for the first time; the pulmonary nodules image lesion features were similar to COVID-19, so the second time was trained on a dataset with labeled pulmonary nodules to further refine the parameters and find similar lesion areas. Finally, the CT images of COVID-19 were segmented again, and better results were achieved.

As mentioned above, most of the current studies on COVID-19 use single networks for learning, but different networks extract different features, so there is a great need to develop a new method for combining multiple networks for learning.

3. Dataset and methodology

For better understanding, Table 5 in Appendix A list the abbreviations. Moreover, the detailed methodology is described below.

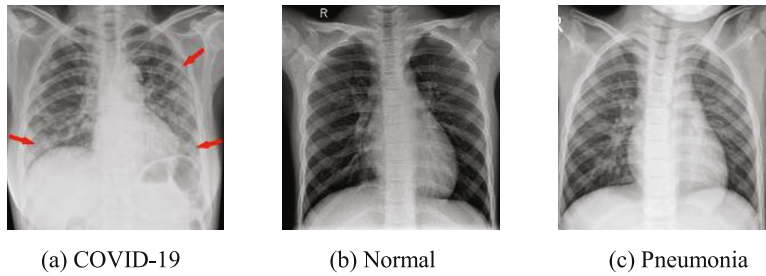


Fig. 1. Sample images of CXRs. (a) COVID-19. (b) Normal. (c) Pneumonia.

Table 1
Data distribution in the model.

Dataset	COVID-19	Normal	Pneumonia	Total
Training (70%)	897	2289	3260	6446
Testing (20%)	256	654	931	1841
Validation (10%)	128	327	466	921
Total (100%)	1281	3270	4657	9208

3.1. Improvement I: MDA on training set

3.1.1. Original dataset

Sait et al. [34] collected 15 publicly available COVID-19 datasets and removed the duplicates to form a new dataset, which is the one used in this work. The dataset contains 1281 COVID-19 X-rays, 1656 viral-pneumonia X-rays, 3270 Normal X-rays, and 3001 bacterial-pneumonia X-rays. We combined viral pneumonia and bacterial pneumonia in a single category. Fig. 1 shows three samples from the dataset. And Fig. 1(a) shows the lesion sites of COVID-19, which we have marked with red arrows. The images in this dataset vary in size and are not labeled for the severity. The main lesion characteristics are described below. The location of infection in COVID-19 is mainly in the bilateral subpleural, whereas in common pneumonia the location of infection is along the trachea, bronchi and blood vessels. The nature of the lesion in

COVID-19 is predominantly ground-glass opacities, whereas the main feature of common pneumonia is a solid shadow.

3.1.2. Dataset preprocessing

The dataset contain COVID-19 X-ray, pneumonia X-ray and healthy X-ray images. We resize the set O of original images to a uniform size of 224×224 , and obtain a new image set R , as shown in Eq. (1).

$$R = \text{Resize}(O, [224, 224]) = \{o_1, o_2, o_3, \dots, o_n\} \quad (1)$$

The abstraction of images in deep learning networks changes from input to convolutional layer, pooling layer, and the last layer of feature map to fully connected layer. Among them, the feature map can be 3×3 , 5×5 , 7×7 , etc. Among these sizes, if the size is too small, then the information is easily lost, and if the size is too large, the abstraction level of information is not high enough and the computation is more, so the size of 7×7 is the most suitable. The input of the image must be $7 \times$ (exponential power of 2) and the size of the dataset images are around 300, so $224 = 7 \times 32$ is the most suitable.

3.1.3. Data augmentation

Through random hold-out (RHO) method, the dataset was randomly divided into three subsets: the testing set (X : 20%), the training set (Y : 70%), and the validation set (Z : 10%). The relevant information is listed in Table 1. Furthermore, to mitigate any potential over fitting, MDA [6]

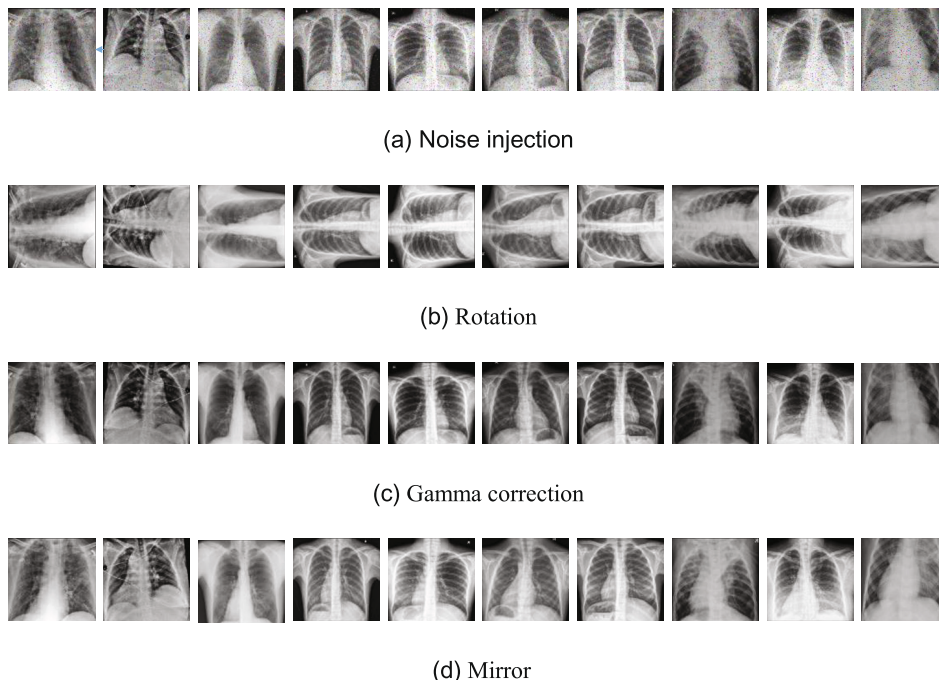


Fig. 2. Four multiple-way data augmentation applied to training set. (a) Noise injection. (b) Rotation. (c) Gamma correction. (d) Mirror.

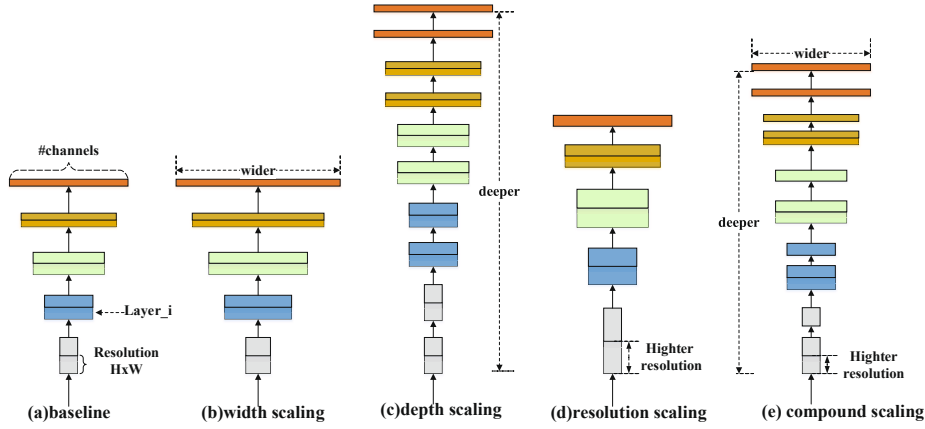


Fig. 3. Diagrammatic representation of an EfficientNet architecture.

technology is utilized in this work. We used four ways to enhance the training set

$$R \xrightarrow{RHO} \{X, Y, Z\} \quad (2)$$

and the relevant sizes related to these subsets satisfy the following equation.

$$|R| = |X| + |Y| + |Z| = |\{x_1, \dots, x_j\}| + |\{y_1, \dots, y_i\}| + |\{z_1, \dots, z_q\}| \quad (3)$$

where $|\cdot|$ refers to the cardinality of a set, i is the number of training set images, j is the number of testing set images, q is the number of validation set images.

Assuming that there are k_{MDC} MDA technique ($k_{MDC} = \{k_1, k_2, k_3, k_4\}$), in this paper, k_{MDC} including noise injection, rotation, gamma correction and mirror), and n_{MDA} images are generated using each MDA technique, and finally for all MDAs, $k_{MDC} \times n_{MDA}$ images are generated. The following four MDA are mainly used in this study:

① Noise injection (N_I)

Gaussian noise was injected into all the images of a training set, thereby generating many new noisy images.

$$\overrightarrow{y^{k_1}(i)} = N_I[y(i)] = [y_1^{k_1}(1), \dots, y_{n_{MDA}}^{k_1}(i)] \quad (4)$$

② Rotation (Ro)

The rotation angle $\theta^{Ro} = 90^\circ$ was applied to the images:

$$\overrightarrow{y^{k_2}(i)} = Ro[y(i)] = [y_1^{k_2}(1, \theta^{Ro}), \dots, y_{n_{MDA}}^{k_2}(i, \theta^{Ro})] \quad (5)$$

③ Gamma correction (G_C)

The gamma correction factor $r^{G-C} = 1.5$ was used to produce new images as follows:

$$\overrightarrow{y^{k_3}(i)} = G_C[y(i)] = [y_1^{k_3}(1, r^{G-C}), \dots, y_{n_{MDA}}^{k_3}(i, r^{G-C})] \quad (6)$$

④ Mirror (Mir)

$$\overrightarrow{y^{k_4}(i)} = Mir[y(i)] = [y_1^{k_4}(1), \dots, y_{n_{MDA}}^{k_4}(i)] \quad (7)$$

$$\overrightarrow{y^k(i)} = concat[\overrightarrow{y^{k_1}(i)}, \overrightarrow{y^{k_2}(i)}, \overrightarrow{y^{k_3}(i)}, \overrightarrow{y^{k_4}(i)}] \quad (8)$$

where $\overrightarrow{y^k(i)}$ means the data augmentation is concatenation of four MDA results..

$$y(i) \xrightarrow{MDA} concat[y(i), \overrightarrow{y^k(i)}] \quad (9)$$

where $y(i)$ means the training set consists of the original and augmentation images.

As shown in Fig. 2, we used four ways to enhance the training set. We

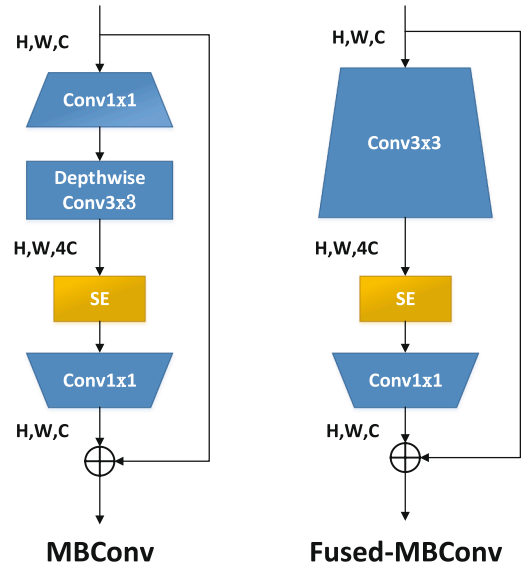


Fig. 4. Structure of MBConv and Fused-MBConv.

can observe that one image will become 5 images.

3.2. Improvement II: Backbone network of EfficientNetV2

In classification problems, to achieve better results, methods that increase the network depth, expand the input image size, and increase the network width are commonly used. However, simply increasing the depth of the network limits the accuracy improvement, as it can easily lead to gradient explosion or gradient disappearance. Besides, the storage requirements increase with an increase in network depth. Additionally, if we simply increase the width of the model, this will allow the model to learn more details. However, if the model is not deep enough, deeper features are not easily learned. Moreover, increasing the resolution of an input image enables the model to acquire more features, but increases the computational cost and reduces the training speed. Accordingly, EfficientNet [35] combines the above-mentioned three trade-off cases to achieve the best result, as demonstrated in Fig. 3. Fig. 3 (a) shows the basic network, while Fig. 3(b-d) improves the performance in terms of increasing the network width, depth, and resolution of the input image, respectively. Finally, Fig. 3(e) illustrates the main idea of EfficientNet, which is to integrate the above three elements to improve the network.

In this paper, we used EfficientNetV2 [36] as the backbone network, which is approximately ten times faster than EfficientNet in training,

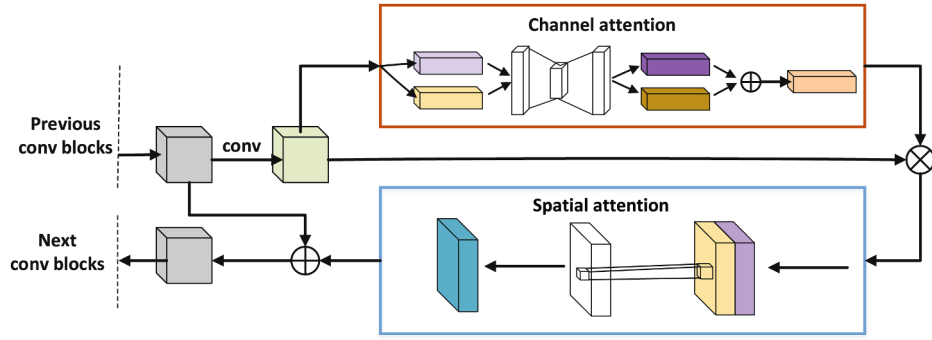


Fig. 5. Structure of ResBlock + CBAM.

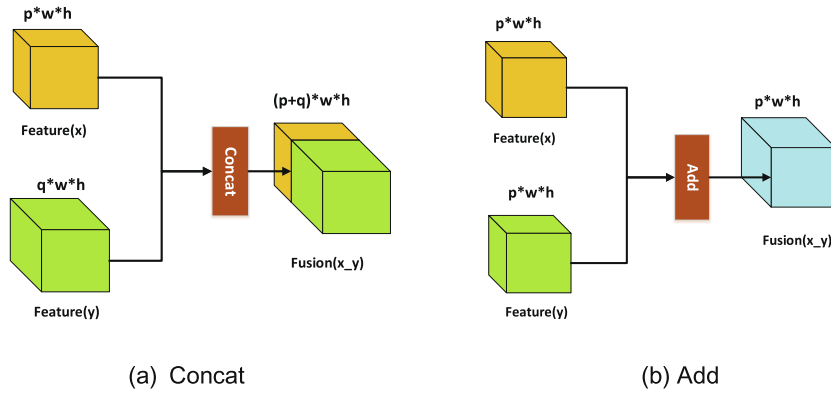


Fig. 6. Two feature fusion methods (a) Concat. (b) Add.

and has a better performance. Likewise, Fused-MBConv corresponds to the key section of EfficientNetV2, which replaces the 1×1 boosted convolution and 3×3 depth-wise convolution in MBConv with a normal 3×3 convolution to improve the training speed, as demonstrated in Fig. 4. Meanwhile, EfficientNetV2 adopts a progressive learning strategy, where the overall training process is divided into four stages. Each stage has stronger regularization for faster convergence, fewer parameters, and very high accuracy. The corresponding training speed is 5 to 11 times faster for the same computational resources.

3.3. Improvement III: Network adds Convolutional Block attention module

CBAM (Convolutional Block Attention Module) [37] is a lightweight module that has two sub-modules: the channel attention (CA) module and the spatial attention (SA) module. In the CA module, the average pool is used to summarize the feature information, the maximum pool is used to get the information of unique objects, and finally the channel relationship of features is used to find the desired feature description and generate the CA graph. In the SA module, which is mainly a complement to the CA module, it can use the spatial relationship between features to determine the location of information and get the SA map. Finally, the CA is arranged in series with the SA, which can improve the representation capability of CNN.

Residual Neural Network (ResNet) [38] was first introduced by K. He et al. Compared to traditional networks, ResNet has fewer parameters (e. g., VGG), better classification, flexible structure. In this paper, we use EfficientNetV2 and ResNet101 to extract features in parallel. The SE module already exists in EfficientNetV2, and we add CBAM to ResNet101 so that both networks can extract features accurately. Fig. 5 shows the exact placement of the modules when integrated into the ResBlock, with the spatial attention module inside the blue border and the CA module inside the red border. We apply CBAM on the

convolution output of each block.

3.4. Improvement IV: Feature fusion

The two commonly used feature-level fusion (FLF) methods are concat and add. Add method corresponds to an increase in information amount for the features describing the image; however, the dimensions describing the image do not increase, as show in Fig. 6(a). On the other hand, concat method refers to a merger of the number of channels, i.e., the number of channels describing the image increases, while relevant information for each feature stays constant. If the dimensions of the two input features x and y are p and q , the dimension of the output feature z is $p + q$, as show in Fig. 6(b). The relevant mathematical expressions are given in Eq. (5) and Eq. (6). In this study, we used concat for FLF, as show in Fig. 6(a). The number of Fusion(x,y) channels refers to the sum of Feature(x) and Feature(y) channels.

$$F_{ff} = \text{concat}(f_E, f_R) \quad (10)$$

$$F_{ff} = \text{add}(f_E, f_R) \quad (11)$$

where f_E is the feature extracted by EfficientnetV2, f_R is the feature extracted by ResNet, and F_{ff} is the fused features set.

3.5. Improvement V: SVM as the classifier

SVM (Support Vector Machine) [39] majorly solves the data classification problem in pattern recognition, and describes the data as points in space and maps them into one or more hyperplanes, constructed by kernel functions. The core idea is to find the separation interface between different categories so that the samples of two categories fall on both sides of the face and as far away from the separation interface as possible. This assists in separating the two different categories quickly. Eqs.(12) and (13) represent formulas for a line or hyper plane,

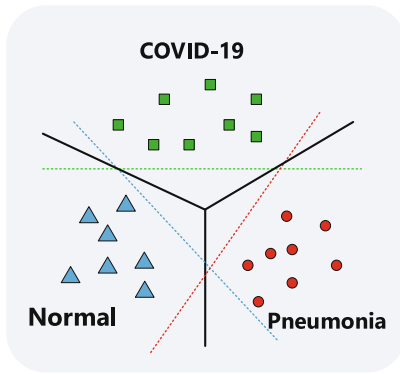


Fig. 7. LibSVM implements triple classification.

respectively. The traditional SVM only performs binary classification, while the LibSVM [40] program is small, has few parameters, flexible in use, and can perform multi-way classification with a good generalization. The LibSVM is the core of MDCM, as elaborated in Fig. 7.

$$\vec{w} \vec{x}_f + b = 0 \tag{12}$$

$$y_f = mx_f + b \tag{13}$$

where w denotes the normal vector of the hyperplane, which determines the direction of the hyperplane. b is the displacement term, which determines the distance between the hyperplane and the origin. x_f is the training sample and y_f is the output of the training example.

3.6. Proposed approach

In this paper, we propose a deep feature fusion classification network called DFFCNet, which comprises three major stages. For the first stage, the dataset is preprocessed and the training set is enhanced with MDA using four methods. In the second stage, feature learning is performed using EfficientNetV2 and ResNet101, where the CBAM module is added to ResNet101 to enhance its feature extraction capability. The third stage involves the classification of the fused features using SVM, which allows multi-disease efficient classification. Fig. 8 depicts the overall framework. To elaborate further, a pseudo code for DFFCNet algorithm is given in Algorithm 1.

Algorithm 1. Pseudo code of our DFFCNet algorithm.

- Phase I: Preprocessing X → Z**
- Step 1 Input: Original Image Set O.
 - Step2 Resizing: Resize the image to [224, 224], get dataset R. See Eq. (1).
 - Step 3 Y_{RHO} : testing set (X), training set (Y) and validate set (Z). See Eq. (2).
 - Step 4 MDA(Y):N,I, RO, G,C and Mir to augment training set (Y).
- Phase II: DFFM**
- Step 5 Read one raw Pre-trained model EfficientNetV2 and ResNet.
 - Step 6 Obtaining MBConv and Fused-MBConv Networks from EfficientNetV2 → M1.
 - Step 7 Adding CA and SA to ResNet → ResNet (CBAM)
 - Step 8 Obtaining residual Networks from ResNet (CBAM) → M2.
 - Step 9 Concat (M1, M2).
 - Step12 Generate DFFM.
- Phase III: MDCM**
- Step13 Get the fusion feature from DFFM → F_{ff} . See Eq. (10).
 - Step14 Create data labels based on feature values.
 - Step15 Normalize the feature values.
 - Step16 Construct MDCM by radial basis and SVM cross-validation.

(continued on next page)

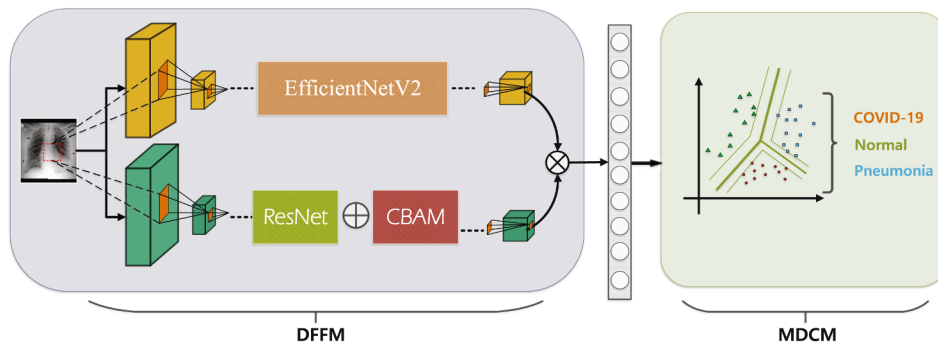


Fig. 8. Structure of the proposed DFFCNet.

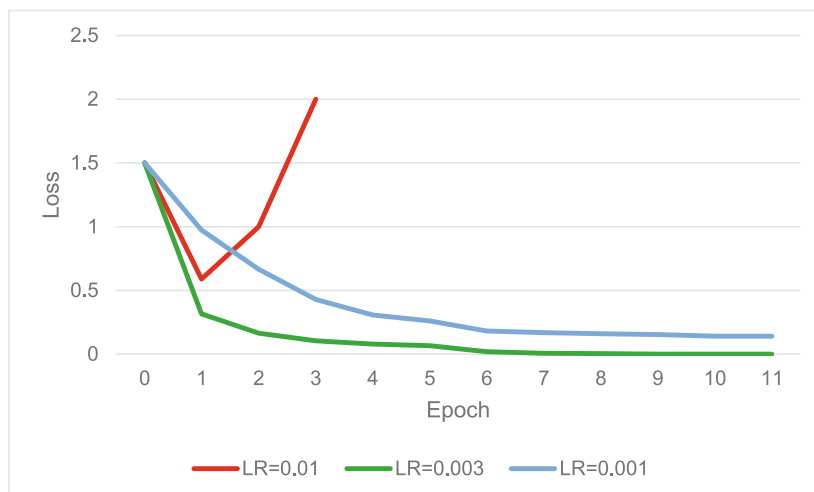


Fig. 9. Relationship between learning rate and loss.

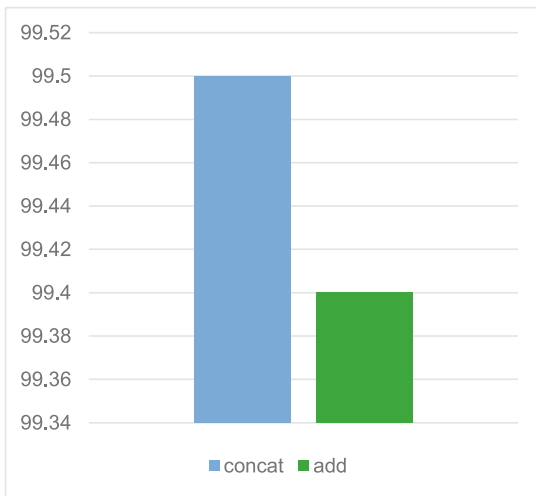


Fig. 10. Comparison of two fusion methods.

(continued)

- Step17 Classification results were obtained by MDCM.
- Step 18 Test confusion matrix, calculate indicators.
- Step 19 Output: The model DFFCNet and its performances.

4. Experiments and results

4.1. The experiment platform

The experiments are performed in a Linux environment, using an NVIDIA DGX Station deep learning workstation with a 32 GB Tesla V100 graphics card to run the experiments. Python language is used to implement the overall code, i.e., data pre-processing and algorithm implementation. Libraries such as Numpy and the deep learning toolbox Pytorch are used. The learning rate (LR), batch size (BS), epochs, optimizer, and dropout rate (DR) made up the tuned hyper-parameters of the model. The values produced the optimal experimental results: batch size, epochs, DR, and initial LR are set to 8, 30, 0.4 and 0.003. When the loss is reduced, the LR is reduced to the original value of 0.1. The main factor affecting the results is LR, which is usually set about three times higher or lower, so we choose 0.003, 0.001 and 0.01 for the experiment. The trend of Epoch and Loss is shown in Fig. 9, where the lower the loss, the better the network performance. When LR is 0.01, the loss does not decrease, but increases, as shown in the red line. The 10th epoch has converged when LR is 0.001 and 0.003. The loss is not minimized when LR is 0.001, as shown in the blue line. The loss is minimized when LR is 0.003, as shown in the green line. So neither LR less than 0.003 nor greater than 0.003 can achieve the best results, so we set LR to 0.003.

4.2. Experiment to determine the feature-fusion methods

The commonly used FLF methods are concat and add. To prove that concat is the best, we conducted an experimental comparison, we fuse the features extracted by EfficientNetV2 and ResNet101 using concat and add respectively, as shown in Fig. 10. The accuracy of fusion using the add method was 99.40%, and the accuracy of fusion using the concat method reached 99.51%. So we used the concat fusion in this paper.

4.3. Ablation study of DFFCNet

To determine the effect of each improvement, we performed an ablation study, as shown in Table 2, the experiment shows the results on the testing set. First we use the backbone network EfficientNetV2 for

Table 2
Ablation study (%).

backbone network	Feature fusion	CBAM	SVM	Acc (X)
✓				97.23
✓	✓			99.51
✓	✓	✓		99.73
✓	✓	✓	✓	99.89

classification and get the accuracy of 97.23%. EfficientNetV2 obtained the accuracy of 99.51% after feature fusion with ResNet101. Next, after adding CBAM to ResNet, the accuracy is 99.73%. Finally, after we replace the classifier with SVM, the accuracy is 99.89%. It can be seen that and every improvement is effect for DFFCNet, especially for feature fusion.

To demonstrate the training process of DFFCNet, we add Fig. 11. The changes of test accuracy and training loss are shown as the epoch increases, where the horizontal axis is epoch, training loss corresponds to the left vertical axis and test accuracy corresponds to the right vertical axis. As can be seen from Fig. 11, DFFCNet converges around the 10th epoch, while the training loss reaches 0.001 and the accuracy is already close to 100%, which has a good performance.

4.4. Experimental results

4.4.1. Classification performance

In order to evaluate the performance of the proposed DFFCNet method, we used various metrics in the validation set to determine, namely accuracy ($Acc(Z)$), precision ($Pre(Z)$), sensitivity ($Sen(Z)$), specificity ($Spe(Z)$), recall ($Rec(Z)$), and F1-score ($F1-sc(Z)$). The corresponding equations are expressed below.

$$Acc(Z) = \frac{TP(Z) + TN(Z)}{TP(Z) + TN(Z) + FP(Z) + PN(Z)} \quad (14)$$

$$Pre(Z) = \frac{TP(Z)}{TP(Z) + FP(Z)} \quad (15)$$

$$Sen(Z) = \frac{TP(Z)}{TP(Z) + FN(Z)} \quad (16)$$

$$Spe(Z) = \frac{TN(Z)}{TN(Z) + FP(Z)} \quad (17)$$

$$Rec(Z) = \frac{TP(Z)}{TP(Z) + FN(Z)} \quad (18)$$

$$F1-sc(Z) = \frac{2*Pre(Z)*Rec(Z)}{Pre(Z) + Rec(Z)} \quad (19)$$

Accordingly, Table 3 demonstrates the overall performance of DFFCNet for the validation set of 921 CXRs. The $Acc(Z)$, $Pre(Z)$, $Rec(Z)$, $Sen(Z)$, $Spe(Z)$ and $F1-sc(Z)$ are 99.9%, 100%, 99.2%, 99.2%, 100% and 99.6% for COVID-19 X-ray images. The $Acc(Z)$, $Pre(Z)$, $Rec(Z)$, $Sen(Z)$, $Spe(Z)$ and $F1-sc(Z)$ are 99.9%, 99.7%, 100%, 100%, 99.8% and 99.8% for pneumonia X-ray images. The $Acc(Z)$, $Pre(Z)$, $Rec(Z)$, $Sen(Z)$, $Spe(Z)$ and $F1-sc(Z)$ are 99.8%, 99.8%, 99.8%, 99.8%, 99.8% and 99.8% for normal X-ray images. In conclusion, the performance in terms of accuracy, sensitivity, recall and F1-score on the validation set is good, so the DFFCNet proposed in this paper is effective.

4.4.2. Confusion matrix

To illustrate the classification of data from the validation set, a confusion matrix [41] is employed in this work. For each class $C = 1, 2, 3$ (1: COVID-19, 2: Pneumonia, 3: Normal), we set that class tag to “positive” and the other two classes to “negative”. Likewise, Fig. 12 presents a schematic confusion matrix for the three categories. The True positives

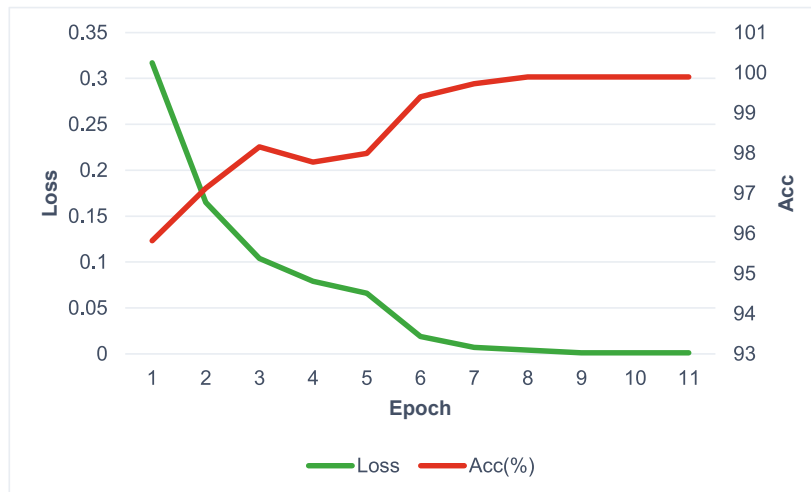


Fig. 11. Test accuracy and training loss of DFFCNet.

Table 3

The classification of DFFCNet networks after two kinds of validation (%).

Class	Acc (Z)	Pre (Z)	Rec (Z)	Sen (Z)	Spe (Z)	F ₁ -sc(Z)
COVID-19	99.9	100	99.2	99.2	100	99.6
Normal	99.9	99.7	100	100	99.8	99.8
Pneumonia	99.8	99.8	99.8	99.8	99.8	99.8

(TP), true negatives (TN), false positives (FP), and false negatives (FN) are used to identify the diagnosis of CXRs by the model. The TP indicates a positive outcome for both real category of the sample and the recognition result. Similarly, the FN highlights a positive real category of the sample, however, it is identified as negative by the model. Alternatively, FP refers to a negative real category of the sample recognized as a positive by the model. Finally, TN indicates a negative category for both real category of the sample and the recognition result. Fig. 13 show the confusion matrix of the DFFCNet model proposed in this paper.

4.4.3. Explainable deep learning using Grad-CAM++

To enhance the intuitive nature of the features and the interpretability of deep learning model, a technique called gradient-weighted class activation mapping plus (Grad-CAM++), proposed by A. Chattopadhyay et al. [42], is adopted in this work. Initially, an image is normalized, and the trained neural network model parameters are loaded. Next, the feature map of the target layer is extracted, and the gradient information of the target class on the feature map is recorded. Next, the heat map is obtained through the weighted summation operation of all the feature maps of a target layer. Eventually, using linear interpolation, the heat map is reduced to the size same as that of original image. The obtained heat map is then superimposed on the original

image to complete the visualization operation. The visualization results for the features generated after the last convolution layer of DFFCNet, using the Grad-CAM++ method are shown in Fig. 14. Among them, the COVID-19 and Pneumonia images possess more obvious features whereas the Normal images have no lesion features.

4.4.4. Comparison with state-of-the-art approaches

To demonstrate the effectiveness of the DFFCNet method, we compared it with eight state-of-the-art methods: ECOVNet [43], Fused-DenseNet-Tiny [23], BCNN_SVM [44], COVNet [45], InceptionV3 [46], DTL-V19 [47], ResNet152V2 [48], and VGG16 [49]. All methods used the unified dataset and MDA preprocessing methods, experiments were performed on the testing set. Table 4 illustrates the relevant comparison results. It can be seen that, among all methods, the proposed DFFCNet

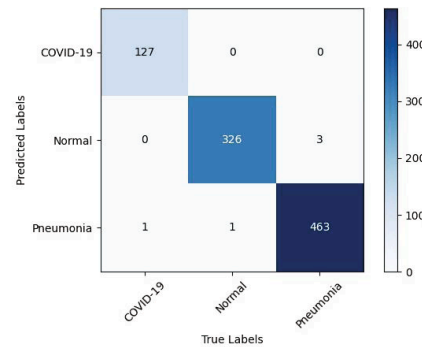


Fig. 13. Classification results of the DFFCNet visualized with a confusion matrix.

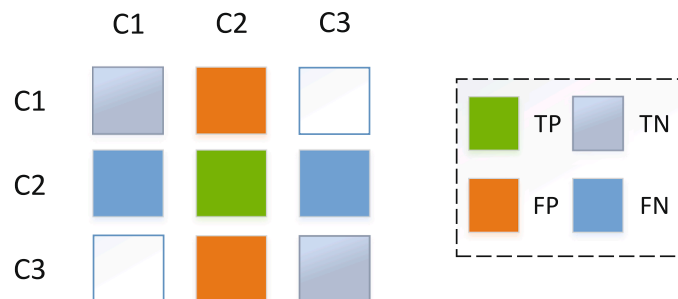


Fig. 12. Confusion matrix of multiple class conditions.

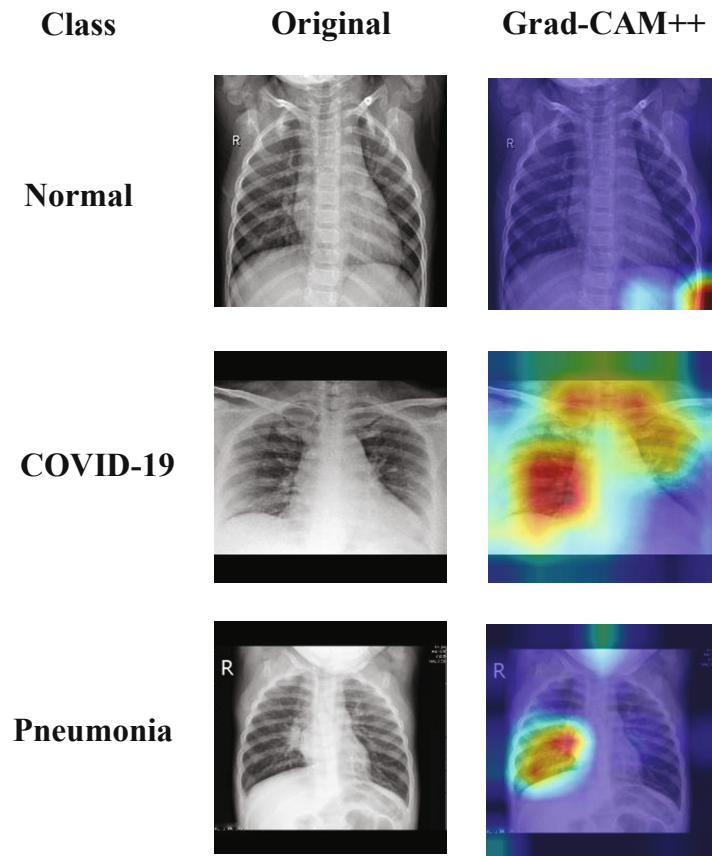


Fig. 14. Grad-CAM++ of the DFFCNet.

Table 4
Performance comparison of the proposed DFFCNet with other studies (%).

Method	Sen (z)	Pre (z)	F1-sc (z)	Acc (z)
ECOVNet [43]	97.53	98.15	97.84	97.72
Fused-DenseNet-Tiny [23]	98.15	98.38	98.26	97.99
BCNN_SVM [44]	96.53	98.06	97.26	97.39
COVNet [45]	95.12	94.34	94.65	95.11
InceptionV3 [46]	98.23	98.31	98.26	97.99
DTL-V19 [47]	95.15	95.66	95.40	95.33
ResNet152V2 [48]	98.09	98.25	98.17	97.88
VGG16 [49]	96.94	97.06	96.97	96.58
DFFCNet (this work)	99.60	99.79	99.70	99.89

achieved the best results. Moreover, accuracy achieved 99.89%. The high accuracy was mainly achieved through the feature fusion and attention mechanism coordination. The use of the newly-proposed EfficientNetV2 as the backbone network and SVM as classifier, the effectiveness of which is demonstrated through the experimental results. In addition, the MDA prevents overfitting of the model, thus improving its performance.

ECOVNet [43], Fused-DenseNet-Tiny [23], COVNet [45] and DTL-V19 [47] are the proposed methods for COVID-19 diseases. BCNN_SVM [44], InceptionV3 [46], ResNet152V2 [48] and VGG16 [49] are better classification networks proposed in recent years, and these methods are very representative. Compared to the other methods, the strategy proposed in this work is unique. BCNN_SVM [44] used a BCNN bilinear fusion of two deep learning networks, VGG16 and VGG19, to extract the features, and then used an SVM to classify for the presence of COVID-19. Since both fusion networks were VGG, the extracted features were similar and the relevant accuracy was lower than our DFFCNet. Fused-DenseNet-Tiny [23] used transfer learning to introduce DenseNet as a backbone network and optimized transfer learning to improve the

performance by freezing some layers and adding new ones. It used the same dataset as in this paper for experiments and the performance was inferior to DFFCNet. Additionally, DTL-V19 [47] used a deep transfer learning of VGG19 for COVID-19 classification. There were fewer training parameters. However, due to the limited network performance, the network had fewer layers and was prone to overfitting. ECOVNet [43] and COVNet [45] were trained for classification using EfficientNetB3 and ResNet50, respectively. These models were simple, lack the fusion of features, and their performance was inferior to our proposed network. Furthermore, we compared the proposed method with the currently popular classification networks, namely InceptionV3 [46], ResNet152V2 [48] and VGG16 [49], and results indicate that these networks were not as effective as DFFCNet. To better display the results, we have added Fig. 15.

5. Conclusion

Coping up with the sudden emergence of COVID-19 virus poses a primary challenge for the medical systems. Due to the lack of doctors and testing reagents, it is difficult to timely diagnose all the potential patients. Nevertheless, the application of AI, which can quickly assist to diagnose diseases through CXRs, saves a lot of time. Likewise, this paper proposed a deep feature fusion efficient classification network (DFFCNet). The proposed network enables an accurate diagnosis of COVID-19, health and pneumonia, especially the prediction accuracy of COVID-19 diseases reached 99.89%. To validate the performance of DFFCNet, we compared the experimental results of 8 state-of-the-art methods. DFFCNet achieved good results in terms of accuracy, precision, sensitivity, F1-score. This helps doctors to make faster and more accurate diagnosis of COVID-19, and thus, our method makes a significant contribution to society and hospitals.

Moreover, the proposed DFFCNet suffers from two disadvantages:

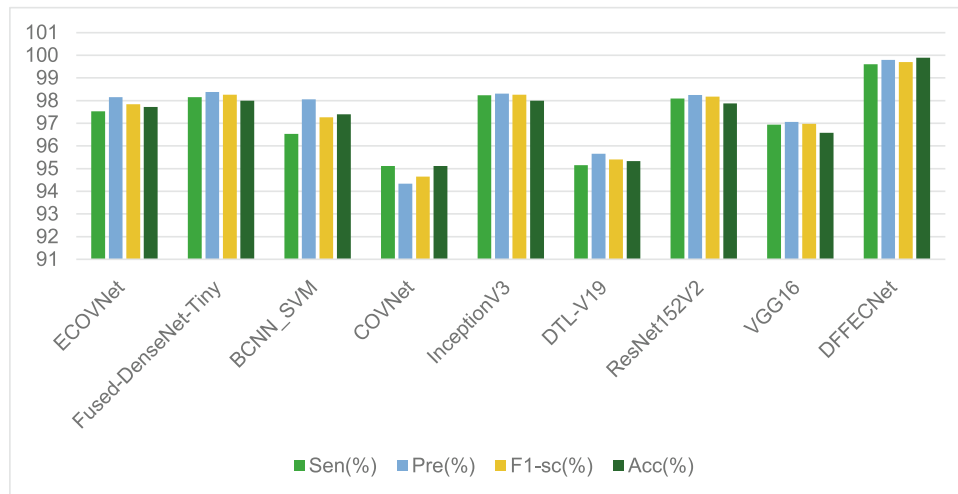


Fig. 15. Comparison of our method with 8 state-of-the-art approaches.

Table 5
Abbreviation list.

Abbreviation	Full name
BS	batch size
CA	channel attention
CXRs	chest X-ray images
CBAM	Convolutional Block Attention Module
DFFECNet	deep feature fusion classification network
DFFM	deep feature fusion module
DR	dropout rate
FLF	feature-level fusion
FN	false negatives
FP	false positives
G_C	gamma correction
Grad-CAM++	gradient-weighted class activation mapping plus
LR	learning rate
MDA	multiple-way data augmentation
MDCM	multi-disease classification module
Mir	mirror
N_I	noise injection
ResNet	Residual Neural Network
Ro	rotation
RHO	random hold-out
SA	spatial attention
SVM	Support Vector Machine
TN	true negatives
TP	true positives

(1) It does not make a judgment about the grade for COVID-19. (2) It cannot handle the datasets constructed via a mixing of CT and CXR. In our future work, we hope to solve the above problems.

6. Data availability

The data that support the findings of this study are openly at [<https://data.mendeley.com/datasets/9xkhgts2s6/1>], reference number. [34].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported by the Jilin Scientific and Technological

Development Program (No. 20200401078GX) and the Guangdong Natural Science Foundation (2021A1515011994).

Appendix A

References

- [1] World Health Organization, WHO coronavirus disease (COVID-19) Dashboard, Available, https://covid19.who.int/?gclid=Cj0KCQjwZtZ7BRDzARIsAGj_bK2ZxWRpJROEl97HGmSOx0_ydkVbc02Ka1FlcysGjEI7hnaLeR6xWhr4aAu57EALw_wcB, 2022.
- [2] O.A. Ataguba, J.E. Ataguba, Social determinants of health: the role of effective communication in the covid-19 pandemic in developing countries, *Global Health Action* 13 (1) (2020) 1788263.
- [3] A. Pormohammad, S. Ghorbani, A. Khatami, R. Farzi, B. Baradaran, D.L. Turner, R. J. Turner, N.C. Bahr, J. Idrovo, Comparison of confirmed COVID-19 with SARS and MERS cases—Clinical characteristics, laboratory findings, radiographic signs and outcomes: a systematic review and meta-analysis, *Rev. Med. Virol.* 30 (4) (2020), e2112.
- [4] D.J. Brenner, E.J. Hall, Computed tomography—an increasing source of radiation exposure, *N. Engl. J. Med.* 357 (22) (2007) 2277–2284.
- [5] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19, *IEEE Rev. Biomed. Eng.* 14 (2020) 4–15.
- [6] S.H. Wang, V.V. Govindaraj, J.M. Górriz, X. Zhang, Y.D. Zhang, Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network, *Inf. Fusion* 67 (2021) 208–229.
- [7] A. Esteve, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature* 542 (7639) (2017) 115–118.
- [8] M.H. Guo, Y.Z. Du, Classification of thyroid ultrasound standard plane images using ResNet-18 networks, *IEEE 13th Int. Conf. Anti-Counterfeiting* (2019) 324–328.
- [9] S. Lu, Z. Lu, J. Yang, M. Yang, S. Wang, A pathological brain detection system based on kernel based ELM, *Multimed. Tools Appl.* 77 (3) (2018) 3715–3728.
- [10] H. Li, M. Habes, D.A. Wolk, Y. Fan, Alzheimer’s Disease neuroimaging initiative, A deep learning model for early prediction of Alzheimer’s disease dementia based on hippocampal magnetic resonance imaging data, *Alzheimer’s Dementia* 15 (8) (2019) 1059–1070.
- [11] M.H. Grewal, M.M. Srivastava, P. Kumar, S. Varadarajan, Radnet: Radiologist level accuracy using deep learning for hemorrhage detection in ct scans, in: 2018 IEEE 15th International Symposium on Biomedical Imaging, 2018, pp. 281–284.
- [12] V. Gulshan, L. Peng, M. Coram, M.C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P.C. Nelson, J.L. Mega, D.R. Webster, Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, *JAMA* 316 (22) (2016) 2402.
- [13] P. Rajpurkar, A. Y. Hannun, M. Haghpahani, C. Bourn, A. Y. Ng, Cardiologist-level arrhythmia detection with convolutional neural networks, (2017) arXiv preprint arXiv: 1707.01836.
- [14] S. Karthik, R. Srinivasa Perumal, P.V.S.S.R. Chandra Mouli, Breast cancer classification using deep neural networks, in: S. Margaret Anouncia, U.K. Wiil (Eds.), *Knowledge Computing and Its Applications*, Springer Singapore, Singapore, 2018, pp. 227–241.

- [15] G.S. Tandel, M. Biswas, O.G. Kakde, A. Tiwari, H.S. Suri, M. Turk, J. Laird, C. Asare, A.A. Ankrah, N.N. Khanna, B.K. Madhusudhan, L. Saba, J.S. Suri, A review on a deep learning perspective in brain cancer classification, *Cancers* 11 (1) (2019) 111.
- [16] I. Arvidsson, N.C. Overgaard, F.E. Marginean, A. Krzyzanowska, A. Bjartell, K. Åström, A. Heyden, Generalization of prostate cancer classification for multiple sites using deep learning, in: 2018 IEEE 15th International Symposium on Biomedical Imaging, 2018, pp. 191–194.
- [17] B. Chen, J. Li, X. Guo, G. Lu, DualCheXNet: dual asymmetric feature learning for thoracic disease classification in chest X-rays, *Biomed. Signal Process. Control* 53 (2019), 101554.
- [18] S. Guendel, S. Grbic, B. Georgescu, S. Liu, A. Maier, D. Comaniciu, Learning to recognize abnormalities in chest x-rays with location-aware dense networks, *in: Inberoamerican Congress on Pattern Recognition*. Springer, 2018, pp. 757–765.
- [19] N. Narayan Das, N. Kumar, M. Kaur, V. Kumar, D. Singh, Automated deep transfer learning-based approach for detection of COVID-19 infection in chest X-rays, *IRBM* (2020).
- [20] T. Majeed, R. Rashid, D. Ali, A. Asaad, Issues associated with deploying CNN transfer learning to detect COVID-19 from chest X-rays, *Phys. Eng. Sci. Med.* 43 (4) (2020) 1289–1303.
- [21] H.S. Maghdid, A.T. Asaad, K.Z. Ghafoor, A.S. Sadiq, S. Mirjalili, M.K. Khan, Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms, *Multimodal Image Exploit. Learn.* 11734 (2021) 117340E.
- [22] I. Katsamenis, E. Protopapadakis, A. Voulodimos, A. Doulamis, N. Doulamis, Transfer learning for COVID-19 pneumonia detection and classification in chest X-ray images, *24th Pan-Hellenic Conf. Informatics* (2020) 170–174.
- [23] F.J.P. Montalbo, Diagnosing Covid-19 chest X-Rays with a lightweight truncated DenseNet with partial layer freezing and feature fusion, *Biomed. Signal Process. Control* 68 (2021), 102583.
- [24] V.N. Aradhya, M. Mahmud, D.S. Guru, B. Agarwal, M.S. Kaiser, One-shot cluster-based approach for the detection of COVID-19 from chest X-ray images, *Cogn. Comput.* 13 (4) (2021) 873–881.
- [25] A. Voulodimos, E. Protopapadakis, I. Katsamenis, A. Doulamis, N. Doulamis, A few-shot U-net deep learning model for COVID-19 infected area segmentation in CT images, *Sensors* 21 (6) (2021) 2215.
- [26] X. Chen, L. Yao, T. Zhou, J. Dong, Y. Zhang, Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images, *Pattern Recogn.* 113 (2021), 107826.
- [27] D. Yang, Z. Xu, W. Li, A. Myronenko, H.R. Roth, S. Harmon, D. Xu, Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan, *Med. Image Analysis* 70 (2021), 101992.
- [28] A. Voulodimos, E. Protopapadakis, I. Katsamenis, A. Doulamis, N. Doulamis, Deep learning models for COVID-19 infected area segmentation in CT images, *The 14th PErvasive Technologies Related to Assistive Environments Conference*, 2021, pp. 404–411.
- [29] X. Chen, L. Yao, Y. Zhang, Residual Attention U-Net for Automated Multi-Class Segmentation of COVID-19 Chest CT Images, (2020) arXiv preprint arXiv: 2004.05645.
- [30] N. Saeedizadeh, S. Minaee, R. Kafieh, S. Yazdani, M. Sonka, COVID TV-Net: Segmenting COVID-19 chest CT images using connectivity imposed Unet, *Comput. Methods Prog. Biomed. Update* 1 (2021), 100007.
- [31] T. Zhou, S. Canu, S. Ruan, Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism, *Int. J. Imaging Syst. Technol.* 31 (1) (2021) 16–27.
- [32] H. Chen, Y. Jiang, M. Loew, H. Ko, Unsupervised domain adaptation based COVID-19 CT infection segmentation network, *Appl. Intell.* (2021) 1–14.
- [33] J. Liu, B. Dong, S. Wang, H. Cui, D.P. Fan, J. Ma, G. Chen, COVID-19 lung infection segmentation with a novel two-stage cross-domain transfer learning framework, *Med. Image Anal.* 74 (2021), 102205.
- [34] U. Sait, K. G. Lal, S. Prajapati, R. Bhaumik, T. Kumar, S. Sanjana, K. Bhalla. Curated Dataset for COVID-19 Posterior-Anterior Chest Radiography Images (X-Rays), *Mendeley Data*, 1, 2020.
- [35] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, *Int. Conf. Mach. Learn.*, PMLR (2019) 6105–6114.
- [36] M. Tan, Q. Le, Efficientnetv2: Smaller models and faster training, (2021) arXiv preprint arXiv: 2104.00298.
- [37] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
- [38] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *2016 IEEE conference on computer vision and pattern recognition (CVPR)*, 2016, pp. 770–778.
- [39] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (3) (1995) 273–297.
- [40] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, *ACM Trans. Intell. Syst. Technol.* (TIST) 2 (3) (2011) 1–27.
- [41] D. Chicco, N. Tötsch, G. Jurman, The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation, *BioData Min.* 14 (1) (2021) 1–22.
- [42] A. Chattopadhyay, A. Sarkar, P. Howlader, V. N. Balasubramanian, Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks, *2018 IEEE winter conference on applications of computer vision (WACV)*, 2018, pp. 839–847.
- [43] A. Garg, S. Salehi, M. La Rocca, R. Garner, D. Duncan, Efficient and visualizable convolutional neural networks for COVID-19 classification using chest CT, *Expert Syst. Appl.* 195 (2022), 116540.
- [44] R. Mastouri, N. Khelifa, H. Neji, S. Hantous-Zannad, A bilinear convolutional neural network for lung nodules classification on CT images, *Int. J. Comput. Assist. Radiol. Surg.* 16 (1) (2021) 91–101.
- [45] H. Ko, et al., COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image: model development and validation, *Journal of medical Internet research* 22 (6) (2020), e19569.
- [46] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 2016, pp. 2818–2826.
- [47] H. Panwar, P.K. Gupta, M.K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, V. Singh, A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images, *Chaos, Solitons Fractals* 140 (2020), 110190.
- [48] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, *European conference on computer vision*, Springer, Cham, 2016, pp. 630–645.
- [49] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, (2014) arXiv preprint arXiv: 1409.1556.