



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.



# Progressive attention integration-based multi-scale efficient network for medical imaging analysis with application to COVID-19 diagnosis<sup>☆</sup>

Tingyi Xie<sup>a</sup>, Zidong Wang<sup>b</sup>, Han Li<sup>c</sup>, Peishu Wu<sup>c</sup>, Huixiang Huang<sup>a</sup>, Hongyi Zhang<sup>a</sup>, Fuad E. Alsaadi<sup>d</sup>, Nianyin Zeng<sup>c,\*</sup>

<sup>a</sup> School of Opto-electronic and Communication Engineering, Xiamen University of Technology, Xiamen 361024, China

<sup>b</sup> Department of Computer Science, Brunel University London, Uxbridge UB8 3PH, UK

<sup>c</sup> Department of Instrumental and Electrical Engineering, Xiamen University, Fujian 361005, China

<sup>d</sup> Communication Systems and Networks Research Group, Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia

## ARTICLE INFO

### Keywords:

Imperfect data  
Artificial intelligence  
Attention mechanism  
Medical imaging analysis  
Progressive learning

## ABSTRACT

In this paper, a novel deep learning-based medical imaging analysis framework is developed, which aims to deal with the insufficient feature learning caused by the imperfect property of imaging data. Named as multi-scale efficient network (MEN), the proposed method integrates different attention mechanisms to realize sufficient extraction of both detailed features and semantic information in a progressive learning manner. In particular, a fused-attention block is designed to extract fine-grained details from the input, where the squeeze-excitation (SE) attention mechanism is applied to make the model focus on potential lesion areas. A multi-scale low information loss (MSLIL)-attention block is proposed to compensate for potential global information loss and enhance the semantic correlations among features, where the efficient channel attention (ECA) mechanism is adopted. The proposed MEN is comprehensively evaluated on two COVID-19 diagnostic tasks, and the results show that as compared with some other advanced deep learning models, the proposed method is competitive in accurate COVID-19 recognition, which yields the best accuracy of 98.68% and 98.85%, respectively, and exhibits satisfactory generalization ability as well.

## 1. Introduction

In modern medicine, various medical imaging techniques have been developing prosperously, such as the computer tomography, X-ray and magnetic resonance imaging [1–3], etc. In comparison to the traditional detection methods, applying medical imaging techniques has the advantages of short detection time, convenient operations and relatively low costs [4,5]. Hence, they have been widely applied in many important diagnostic scenes, which accordingly requires an effective analysis on the obtained imaging data.

In general, the medical imaging data always present the characteristics of large amount, huge dimension and highly non-linearity, which make it an extremely challenging task to efficiently extract the rich semantic information of some serious diseases. Owing to the

fast development of artificial intelligence (AI), the deep learning (DL)-based models have exhibited inspiring performance in analyzing data generated by various medical imaging techniques [3,6,7], which can realize the end-to-end learning from the great amount of data and has aroused great research interests [8–10].

It is noticeable that performance of the DL-based models can be severely affected by the imperfect medical imaging data, for example, in the chest X-ray images, the organic entity may merely occupy 50% of space, whereas the rest of background part contains much useless and redundant information, which causes the waste of computational resources and leads to inefficient feature extraction [11]. Similarity in imaging results of different diseases also impedes the accurate recognition of relevant conditions [12].

<sup>☆</sup> This work was supported in part by the Natural Science Foundation of China under Grant 62073271, the Fundamental Research Funds for the Central Universities of China under Grant 20720220076, the Natural Science Foundation of China under Grant 2020J01296, the Science and Technology Research Program of Chongqing Municipal Education Commission of China under Grant KJQN202001319, the Startup Project for High-level Talents of Xiamen University of Technology of China, the Scientific Research Climbing Program of Xiamen University of Technology of China, and the Xiamen Medical and Health Guidance Project of China under Grant 3502Z20224ZD1010.

\* Corresponding author.

E-mail addresses: [zidong.wang@brunel.ac.uk](mailto:zidong.wang@brunel.ac.uk) (Z. Wang), [zny@xmu.edu.cn](mailto:zny@xmu.edu.cn) (N. Zeng).

To deal with the imperfect property of data so as to improve the learning ability of the DL-based models, many efforts have been carried out, where the application of attention mechanism is a hot research topic. In [13], a novel feature attention super-resolution network has been proposed to take full use of the rich anatomical information in magnetic resonance images, which can effectively predict the missing high-resolution details in magnetic resonance images so as to realize accurate disease diagnosis. To overcome the data redundancy, a multimodal spatial attention module has been proposed in [14], which automatically emphasizes the important spatial regions and suppresses the normal ones.

It is worth mentioning that while applying attention mechanism, global information loss is commonly encountered due to the too much focus on the specific local areas, which declines the recognition accuracy. Moreover, to promote accurate diagnosis, the extracted features are required to contain both sufficient details and rich semantic information, and it is promising to take full advantages of various attention mechanisms to realize efficient feature learning. In addition, how to balance the model complexity and recognition accuracy when designing corresponding feature refinement components also deserves attention.

Based on above discussions, in this paper, a novel medical imaging analysis framework is developed, which dedicates to overcoming the weak feature representation resulted from the imperfect property of the imaging data. Particularly, in the developed multi-scale efficient network (MEN), a fused-attention and a multi-scale low information loss (MSLIL)-attention block are proposed to form an information refinement module so as to realize a from-coarse-to-fine feature extraction in a progressive learning manner. To be specific, in the fused-attention block, the squeeze-excitation (SE) attention mechanism is applied to highlight important channels and suppress the useless information. In the MSLIL-attention block, the efficient channel attention (ECA) mechanism is applied to capture global dependencies and, by doing so, the potential information loss can be compensated to some extents. As a result, not only detailed features can be efficiently extracted, the semantic correlations among different features are enhanced as well. Finally, the proposed MEN is applied to the important COVID-19 diagnostic tasks for performance evaluation.

The main contributions of this paper are outlined as follows.

- (1) A novel DL-based medical imaging analysis model MEN is developed, which efficiently extracts the features in a progressive learning manner.
- (2) The proposed fused- and MSLIL-attention blocks can effectively coordinate the focus on lesion areas and the compensation of global information loss.
- (3) Combinations of diverse attention mechanisms realizes sufficient learning of both details features and semantic information.

The remainder of this paper is organized as follows. Related work is reviewed in Section 2. The proposed MEN is elaborated in Section 3. Results and discussions are presented in Section 4. Finally, the conclusions are drawn in Section 5.

## 2. Related work

Since the proposed model is applied into the COVID-19 detection task, in this section, some related recent studies are reviewed in terms of the performance improvement on the DL-based model.

Regarding to the optimization on model training, in [15], a framework called COVID-ResNet has been proposed for COVID-19 detection, where during training, the image size is progressively adjusted to enhance the generalization ability. In [16], an improved snapshot ensemble technique has been proposed for COVID-19 chest X-ray images classification, which replace the original probability average by a weighted one, and experiments have shown the effectiveness of such optimization.

To cope with the over-fitting problem caused by limited chest X-ray image in COVID-19 category, some light-weight neural networks

have been designed. In [17], a novel framework DarkCovidNet has been proposed for COVID-19 detection, where both the number of hidden layers and filters are reduced. Experimental results have shown that the proposed method is a reliable light-weight model that effectively removes redundant information, which obtains accuracy of 98.08% and 87.02% in binary classification and multi-classification case, respectively. A 24-layer convolutional neural network (CNN) has been proposed in [18], which is based on the first five hidden layers of the VGG [19] architecture and is pre-trained to ensure the accuracy. In [20], the EfficientNetB4 network has been improved by adding a GPA module and dropout layer to reduce the parameters. Although experiments have shown that above strategy can effectively simplify the model structure, in complicated multi-classification tasks, these light-weight models may suffer from insufficient learning, resulting in low classification accuracy.

To realize sufficient feature extraction from images, a multi-kernel-size spatial-channel attention mechanism has been proposed in [21], which applies convolutional kernels in different sizes combined with channel and spatial attention to promote accurate recognition of the COVID-19. In [22], the DenseNet has been combined with GAT [23] model for COVID-19 detection, where multiple independent attention mechanisms are deployed, and it is proven that the multi-head mechanism can distinguish the attention on multiple features between the central and neighboring nodes. Depth-wise dilated convolution module has been applied in [24] to enlarge the receptive field, which allows the model capturing rich global information of the chest X-ray image. In [25], a PEPX module has been designed, which not only reduces the network complexity, but also obtains channel information in different dimensions.

In this study, by applying the progressive learning approach, multi-scale feature extraction and multi-modal residual fusion methods are integrated to enhance the model performance without significantly increasing the computation complexity.

## 3. Methodology

In this section, the proposed multi-scale efficient network is presented, whose overall framework is shown in Fig. 1, including modules of data preprocessing, information refinement and classifier.

### 3.1. Overall framework of MEN

According to Fig. 1, three main modules are contained in the proposed MEN, where sufficient feature extraction and fusion are performed to realize accurate classification. Firstly, images in three categories are input to the data preprocessing module, which are further expanded with a series of data augmentation operations, including normalization, center cropping, modal splicing and data partitioning. In particular, to enrich the channel information of images in pneumonia class,  $3 \times 3$  convolution, layer normalization and Silu activation function [26] are applied to transfer the RGB images to 24-channel enhanced vectors.

Then, attention mechanism is applied in the subsequent information refinement module, which is mainly realized by the designed fused- and MSLIL-attention blocks. In particular, the former is used to extract fine-grained information and the latter is responsible for enhancing semantic correlations by capturing global dependencies, which realizes a progressive learning and more implementation details of above two attention blocks are provided in the following subsections.

Finally, classification is accomplished in the classifier, where  $1 \times 1$  convolution, adaptive average pooling and a fully-connected (FC) layer are applied to process the extracted features.

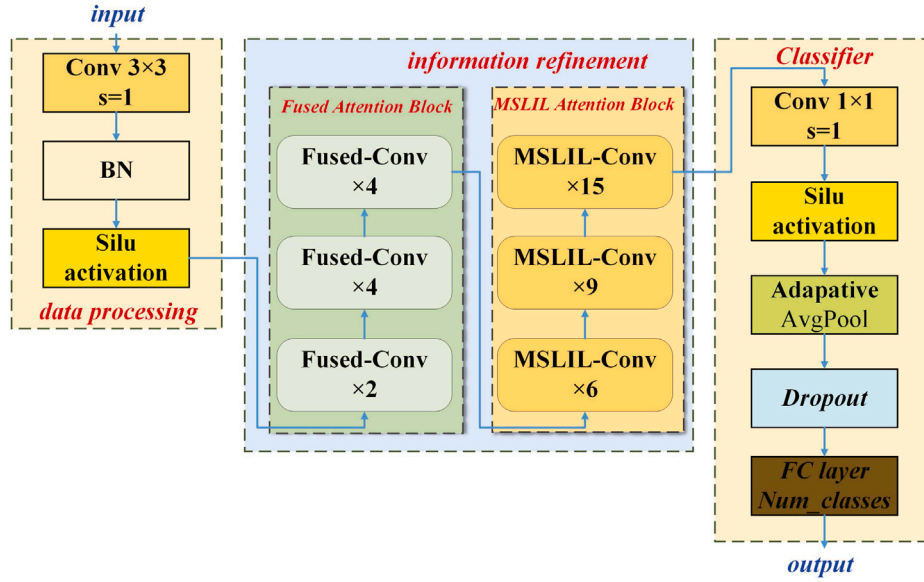


Fig. 1. Framework of the proposed multi-scale efficient network.

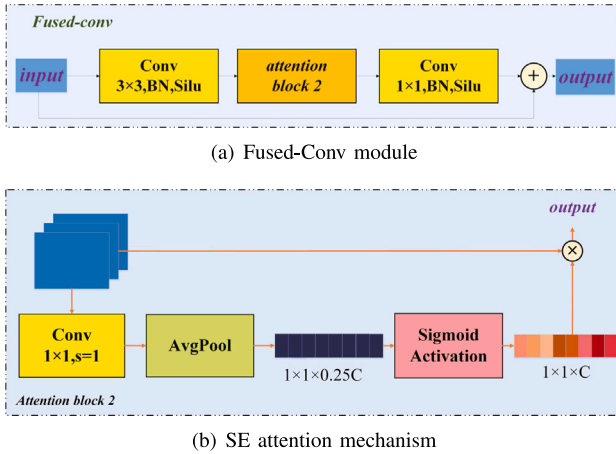


Fig. 2. Architecture of components in the fused-attention block.

### 3.2. Fused-attention block

As is shown in Fig. 1, the fused-attention block is composed of several repeated fused-convolution (Fused-Conv) modules, whose architecture is shown in Fig. 2. Notice that the input images of Fused-Conv have large size but few channels, as a result, extracting detailed information is focused in this module by embedding in the squeeze-excitation (SE) attention mechanism [27]. At first, the global average pooling is performed to realize feature compression in spatial dimension. Next, importance of each channel is predicted on the compressed features by an FC layer, which is reflected by weight coefficients  $\omega$ , and the principle of obtaining  $\omega$  can be described as:

$$s = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W I(i, j) \quad (1)$$

$$\omega = \sigma(W_2 \delta(W_1 s))$$

where  $H$  and  $W$  are number of channels,  $I$  is the initial input and  $s$  is the compressed one.  $W_1$  and  $W_2$  are two matrices,  $\sigma$  and  $\delta$  denotes activation functions in the FC layer. At last, the output is obtained by channel-wise multiplication of the weight coefficient and the initial input.

It is noticeable that the SE attention mechanism is beneficial for highlighting the lesion areas in the images so as to determine the focus on important semantic information in case of high-resolution images. At the end of Fused-conv module, the  $1 \times 1$  convolution is used to adjust the number of channels, which enables the fusion of context information.

### 3.3. MSLIL-attention block

To enhance the extraction of semantic information by introducing useful features, a novel MSLIL-attention block is developed in the proposed MEN, which consists of a series of MSLIL convolution (MSLIL-Conv) modules.

As is shown in Fig. 3(a), a multi-branch structure is deployed in the MSLIL-Conv module, where two dilated-convolutions [28] and a group-convolution are set in parallel. Particularly, the dilated-convolution can make compensation for the lost receptive fields by providing rich global information, and the  $3 \times 3$  group-convolution can further reduce computational burdens. By adjusting the hyper-parameters of convolution operators, the outputs of above three branches can be integrated to enter the efficient channel attention (ECA) mechanism [29]. In addition, the integrated multi-branch output is further multiplied with the output of above ECA mechanism, which can remain partial original features so as to reduce the loss of essential information.

Framework of the ECA mechanism is illustrated in Fig. 3(b), which is equipped with a residual structure to extract the superior semantic information. Different from the SE mechanism, the learning of channel importance is performed directly on the input via 1D convolutions. In this regard, information loss caused by dimension-reduction operations in FC layers can be effectively avoided. Moreover, appropriate cross-channel interaction is also realized to enhance the correlations of each channel information. In ECA mechanism, the weights of the 1D convolution are interleaved with each other in a cross-channel manner, which present in groups and for each group, the number of weight depends on the size of convolution kernel  $k$ . The weight of feature  $y_i$  is calculated by considering only the interaction between  $y_i$  and its  $k$  adjacent channels, which is expressed as:

$$\omega_i = \sigma \left( \sum_{j=1}^k w^j y_i^j \right), y_i^j \in \Omega_i^k \quad (2)$$

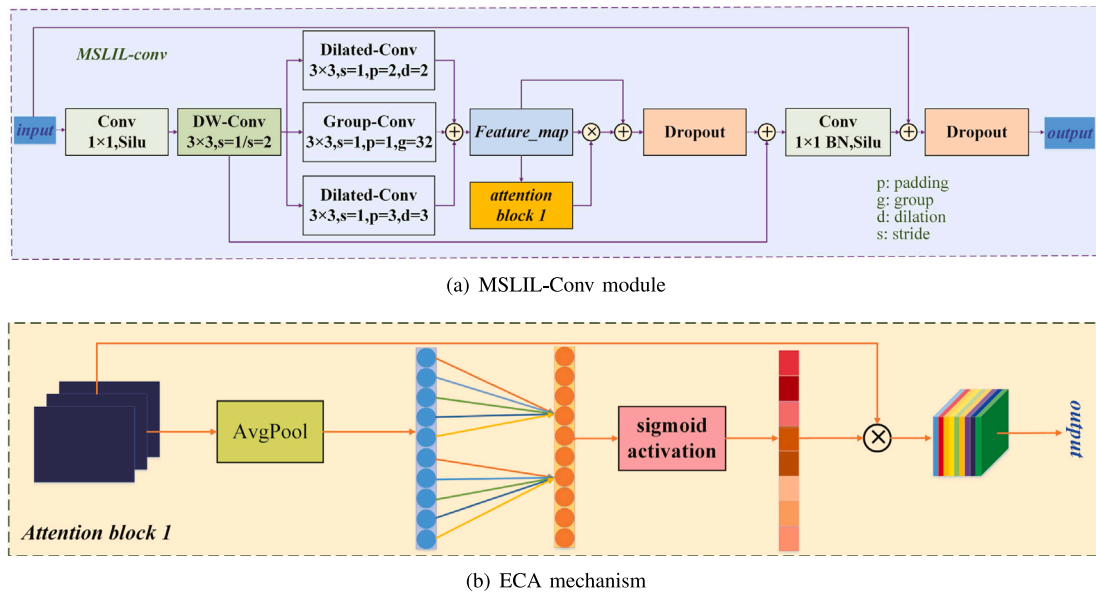


Fig. 3. Architecture of components in the MSLIL-attention block.

where  $\Omega_i^k$  is the set of  $k$  adjacent channels of  $y_i$  and  $w$  is the weight of  $y_i$ . Note that above Eq. (2) can be implemented by a fast 1D convolution with kernel size of  $k$  by:

$$\omega = \sigma(\text{CID}_k(\mathbf{y})) \quad (3)$$

where CID refers to the 1D convolution, and the obtained weights are shared in all  $C$  channels. In addition, considering that the channel number  $C$  in the proposed MEN is the power of 2, following non-linear map  $\phi$  is applied to characterize the relationship between  $C$  and  $k$  by:

$$C = \phi(k) = 2^{\gamma \times k - b} \quad (4)$$

where  $\gamma$  and  $b$  are two coefficients, and in this study, they are set as  $\gamma = 2$  and  $b = 1$ , respectively. Finally, given a channel dimension  $C$ , the kernel size  $k$  can be adaptively obtained as:

$$k = \phi^{-1}(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (5)$$

where  $\lfloor \cdot \rfloor_{\text{odd}}$  outputs the nearest odd number.

In addition, two dropout layers are deployed in the MSLIL-attention block, which randomly discard the 20% of features so as to effectively reduce the computational complexity and prevent the overfitting.

## 4. Results and discussions

In this section, the proposed MEN is applied to the COVID-19 detection tasks for performance evaluation, where two public datasets of chest X-ray images are adopted, and the results are compared with other popular deep learning models.

### 4.1. Experimental environment

#### 4.1.1. Datasets

In this study, the proposed MEN is evaluated on two public COVID-19 chest X-ray image datasets (<https://aistudio.baidu.com/aistudio/datasetdetail/163046>) [30], which are denoted as D1 and D2, respectively. In each dataset, three categories of images are contained, including COVID-19, normal and viral pneumonia. It is noticeable that the samples in COVID-19 and viral pneumonia category are highly similar, which brings difficulties in accurate recognition. The data partition for model training and testing is based on the ratio of 8:2 as shown in Fig. 4, and moreover, some data augmentation operations like random flip and center crop are performed on the training set, which further enrich the samples to promote robust feature learning.

Table 1  
Hyperparameter settings.

Parameters	Settings
Training epochs	200
Batch size	16
Optimizer	SGD
Learning rate	0.01
Learning rate decay	Cosine decay

#### 4.1.2. Evaluation metrics

For the above mentioned three-category classification tasks, confusion matrix is adopted for performance evaluation, based on which five metrics of accuracy, precision, recall, specificity, and F1 score are calculated as follows.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (9)$$

$$\text{F1 score} = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (10)$$

where TP and FP refer to the true and false positive classification results, TN and FN are the true and false negative ones, respectively.

#### 4.1.3. Experimental settings

In Table 1, the model configurations when training the proposed MEN are presented, and the comparison models employed in this study are parameterized according to the corresponding literature, including ResNet101 [31], MobileNetV2 [32], DensNet201 [33], EfficientNetV2\_s [34], and EfficientNet\_b0 [35]. All experiments are carried out on the deep learning framework Pytorch, and the operating system is Windows 10 with an NVIDIA GTX 2060 single GPU.

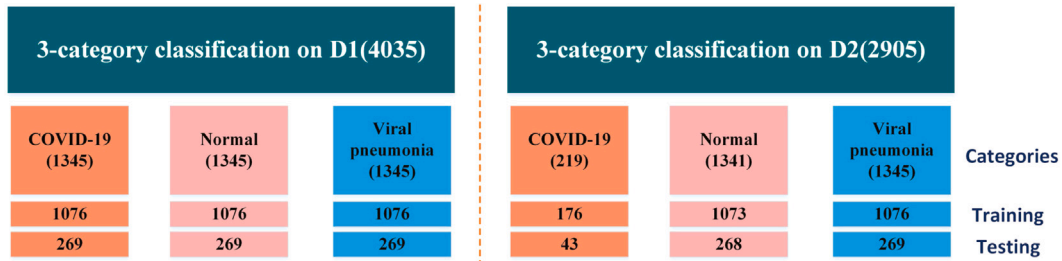


Fig. 4. Data partition for model training and testing.

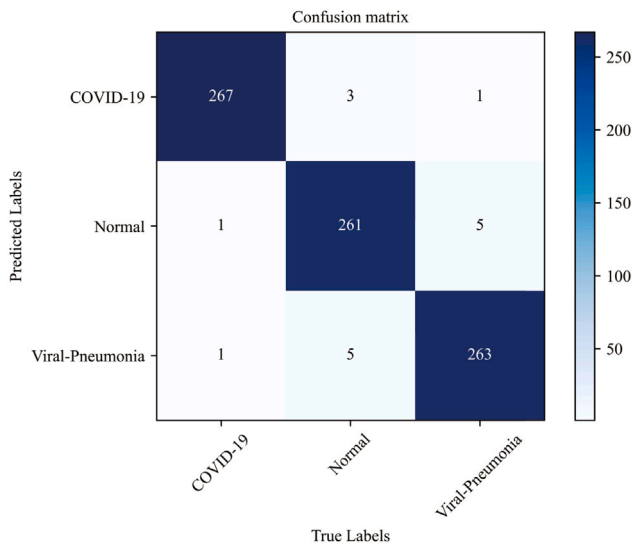


Fig. 5. Confusion matrix on the D1 dataset.

Table 2

Percentage of Ground-truth and Prediction results based on confusion matrix on D1.

Ground-truth	Prediction results		
	COVID-19	Normal	Viral-Pneumonia
COVID-19	99.27%	0.372%	0.372%
Normal	1.115%	97.03%	1.860%
Viral-Pneumonia	0.372%	1.860%	97.77%

Table 3

Percentage of Ground-truth and Prediction results based on confusion matrix on D2.

Ground-truth	Prediction results		
	COVID-19	Normal	Viral-Pneumonia
COVID-19	97.67%	2.33%	0%
Normal	0%	99.25%	0.746%
Viral-Pneumonia	0%	2.60%	97.40%

take proportion of 0.37%, which shows that the proposed MEN can effectively distinguish the highly similar images in COVID-19 and viral pneumonia category. For the other two categories, the misclassification rates are both around 2.6%, which implies that the proposed MEN presents a satisfactory classification performance on dataset D1.

Similarly, the evaluation results on dataset D2 are displayed in Fig. 6 and Table 3, respectively. It is noticeable that the data are quite limited in D2, especially the COVID-19 samples, which makes it tough for sufficient feature extraction. While according to the results, only one image in COVID-19 category is misclassified as the normal one and none of samples in the other two categories are recognized as the COVID-19 sample. Moreover, even the lowest accuracy occurring in viral pneumonia image identification reaches 0.974, which demonstrates that the proposed MEN does extract features with strong representation from the limited data, thereby exhibiting outstanding feature extraction ability.

Based on the above confusion matrices, five indicators in Eq. (6)–(10) are calculated, and corresponding evaluation results on D1 and D2 are reported in Table 4 and Table 5, respectively. Notice that the presented results are in class-wise, and an intuitive presentation is also shown in Figs. 7 and 8.

As is shown in Table 4, all of the evaluation metrics exceed 97% on dataset D1, where the recall, specificity and accuracy obtained in COVID-19 class obtain the highest value of 99.26%, which demonstrates that the proposed MEN is competent in accurate recognition of the COVID-19 samples. In addition, accuracy obtained in the other two categories also reaches 0.9827 and 0.9851, respectively, which shows the competitiveness of our MEN in accurate classification.

In Table 5, evaluation metrics obtained by the proposed MEN on dataset D2 are reported, and it is found that both precision and specificity in the COVID-19 class reach 1.0. It is worth mentioning that as compared to dataset D1, the amount of data in D2 is fewer with unbalanced distribution, especially the COVID-19 samples are quite limited. Therefore, such result sufficiently prove the superiority of the proposed MEN, and the accuracy obtained on dataset D2 is even 0.57% higher than that on D1. Moreover, the F1 score on all categories

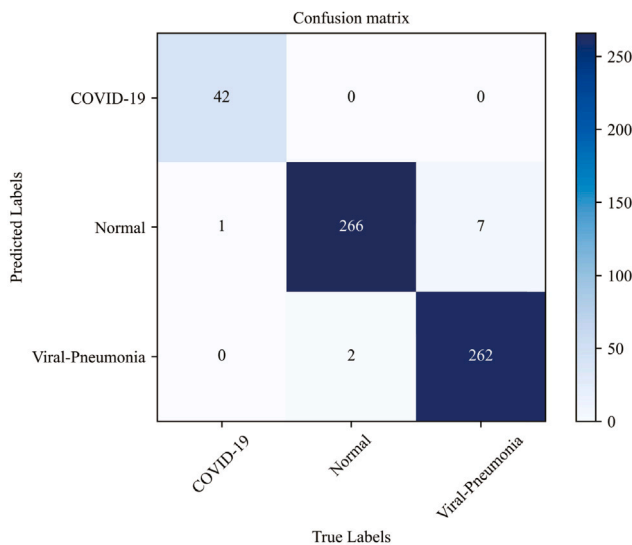


Fig. 6. Confusion matrix on the D2 dataset.

#### 4.2. Evaluation results

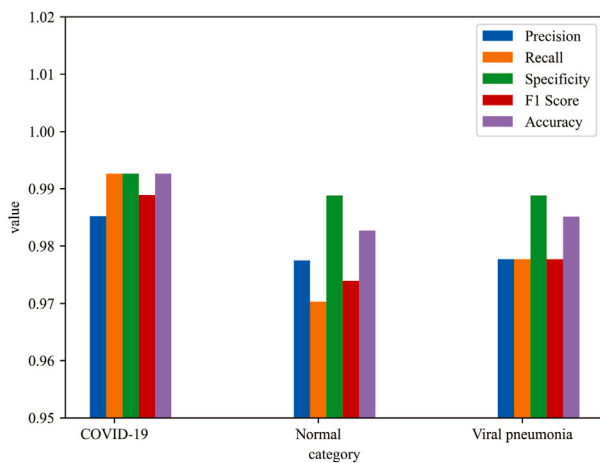
At first, the confusion matrix obtained by the proposed MEN on dataset D1 is illustrated in Fig. 5, and corresponding data are reported in Table 2 for a clear view. It is found that on the COVID-19 category, the highest classification accuracy of 99.27% is achieved, and the COVID-19 samples that are misdiagnosed as viral pneumonia only

**Table 4**  
Performance of the proposed MEN on dataset D1.

Metrics	Category		
	COVID-19	Normal	Viral-Pneumonia
Precision	<b>0.9852</b>	0.9775	0.9777
Recall	<b>0.9926</b>	0.9703	0.9777
Specificity	<b>0.9926</b>	0.9888	0.9888
F1 score	<b>0.9889</b>	0.9739	0.9777
Accuracy	<b>0.9926</b>	0.9827	0.9851

**Table 5**  
Performance of the proposed MEN on dataset D2.

Metrics	Category		
	COVID-19	Normal	Viral-Pneumonia
Precision	<b>1.0</b>	0.9708	0.9924
Recall	0.9767	<b>0.9925</b>	0.9740
Specificity	<b>1.0</b>	0.9744	0.9936
F1 score	<b>0.9884</b>	0.9816	0.9832
Accuracy	<b>0.9983</b>	0.9828	0.9845



**Fig. 7.** Performance of the proposed MEN on dataset D1.

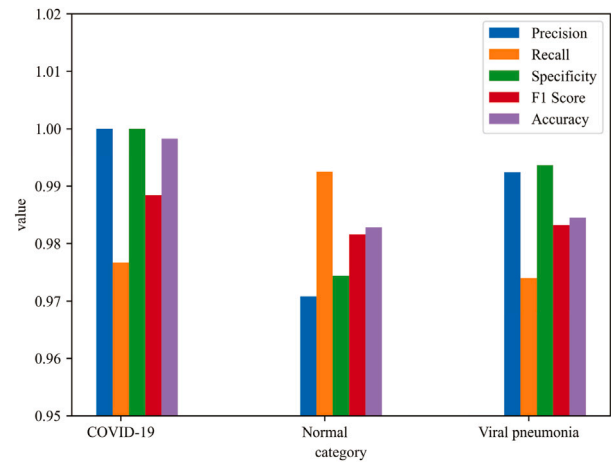
surpasses 0.98, which shows the excellent comprehensive performance of MEN by well balancing the precision and recall.

Based on above discussions, one can conclude that the proposed MEN is reliable in the important diagnostic task of recognizing COVID-19 based on the chest X-ray images. On two different diagnostic scenes, our MEN achieves satisfactory results, which also exhibits a well generalization ability.

#### 4.3. Comparisons with other advanced models

To further verify the competitiveness of our method in feature extraction, the proposed MEN is compared with other five advanced deep learning models, and the comparison results in terms of the overall classification performance are displayed in Table 6, Table 7 and Fig. 9. Particularly, in [16], an improved snapshot ensemble technique has been combined with a neural network, and data-specific augmentation operations have been investigated in [30]. Both of the above two recently proposed models have obtained satisfactory results on dataset D2, and as a result, in Table 7, they are also compared with the proposed MEN, where the results are directly cited from the original papers.

According to Table 6, the proposed MEN shows an improved performance on all five metrics as compared with other five advanced deep learning models, with each metric exceeding 98%. It may be due to the deployments of two different channel attention mechanisms in appropriate locations, which allows a feature extraction with rich



**Fig. 8.** Performance of the proposed MEN on dataset D2.

and accurate information. In the more difficult diagnostic task on dataset D2, it is also found from Table 7 that the proposed MEN outperforms other seven state-of-the-art models in terms of each metric. It is noticeable that as compared to the results on dataset D1, there are declines to different extents regarding to accuracy on D2 obtained by the Resnet101, Densenet201 and Efficientnetv2\_s model. On the contrary, accuracy obtained by Efficientnet\_b0, Mobilenetv2 and the proposed MEN is even higher on D2 than that on D1, which implies that a lighter structure may be more suitable for handling limited and imbalanced data. In the designed MSLIL-Conv module of MEN, not only the data are further enriched by the deployed multi-branch structure, two embedded dropout layers also effectively prevent the over-fitting phenomenon. Consequently, the proposed MEN can present stable performance in the case of insufficient and imbalanced data.

In addition, the receiver operator characteristic (ROC) curves obtained by above six models are illustrated in Fig. 10, which selects recall as the vertical coordinate and 1-specificity as the horizontal one. A property of the ROC curve is that its shape is able to remain unchanged as the distribution of positive and negative samples changes, which thereby reduces the interference caused by different test sets and provides an objective estimation of the model performance. Note that if the ROC curve is closed to the top left corner, that is, the area under curve (AUC) is large, then the model is deemed to have good performance. As can be seen in Fig. 10, the proposed MEN yields the best AUC values of 0.9979 on D1 and 0.9999 on D2, respectively, which shows the competitiveness of our method in extracting discriminative features from both positive and negative samples.

Moreover, a supplement dataset from Kaggle without training is adopted for performance evaluations of the six models, where the task is also to recognize the COVID-19 samples from images of three categories, and the ROC curves are presented in Fig. 11. Combining the results shown in Fig. 10, it is found that the AUC value obtained by Efficientnetv2\_s, which is the sub-optimal model on both datasets D1 and D2, is greatly reduced to 0.8691 on the supplement dataset. By contrast, the proposed MEN still maintains the highest AUC value of 0.9397, which shows the outstanding robustness and generalization ability of our method.

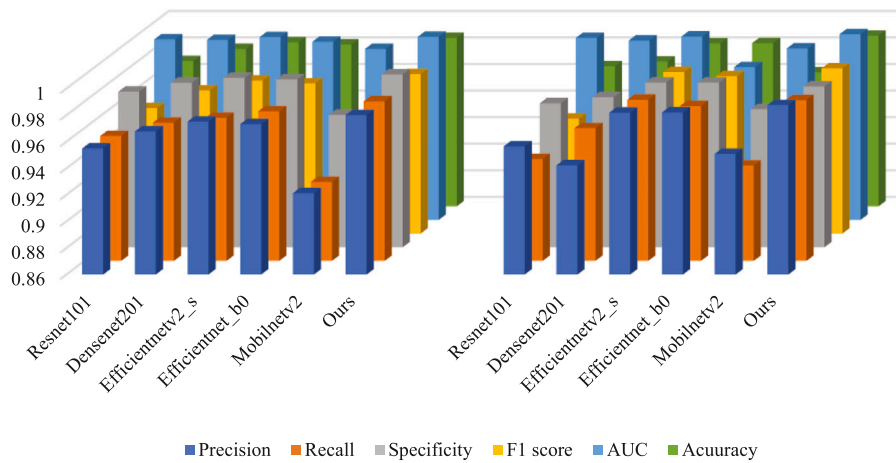
According to the above discussions, the proposed MEN is competent in recognizing COVID-19 based on chest X-ray images with considerable overall performance. Although satisfactory results are achieved on different diagnostic tasks, there are still some spaces for further improvements on our method. Firstly, the ability of extracting highly discriminative features can be enhanced by diverse convolution variants and multi-scale feature fusion methods [36,37]. Secondly, common spatial pattern [38] and multi-agent systems [39] can be applied to

**Table 6**  
Performance comparison of the proposed MEN and other five advanced models on dataset D1.

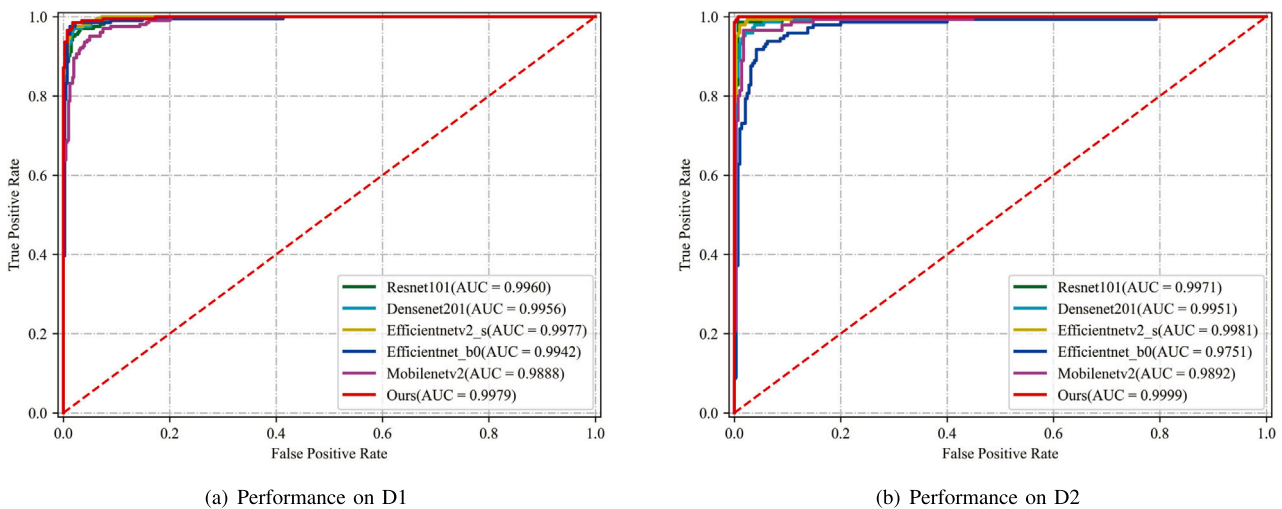
Databases	Metric	Models					
		Resnet101	Densenet201	Efficientnetv2_s	Efficientnet_b0	Mobilenetv2	Ours
D1	Precision	0.9550	0.9677	0.9752	0.9733	0.9213	<b>0.9801</b>
	Recall	0.9541	0.9641	0.9678	0.9727	0.9195	<b>0.9802</b>
	Specificity	0.9771	0.9839	0.9876	0.9864	0.9597	<b>0.9901</b>
	F1 score	0.9545	0.9678	0.9752	0.9730	0.9203	<b>0.9801</b>
	Accuracy	0.9694	0.9785	0.9835	0.9818	0.9463	<b>0.9868</b>

**Table 7**  
Performance comparison of the proposed MEN and other seven advanced models on dataset D2.

Databases	Metric	Models							
		Resnet101	Densenet201	Efficientnetv2_s	Efficientnet_b0	Mobilenetv2	[16]	[30]	Ours
D2	Precision	0.9565	0.9423	0.9819	0.9821	0.9509	0.9523	0.9795	<b>0.9877</b>
	Recall	0.9368	0.9600	0.9814	0.9767	0.9318	0.9563	0.9794	<b>0.9811</b>
	Specificity	0.9684	0.9729	0.9840	0.9840	0.9640	0.9716	0.9880	<b>0.9893</b>
	F1 score	0.9466	0.9512	0.9817	0.9785	0.9413	0.9542	0.9794	<b>0.9844</b>
	Accuracy	0.9655	0.9690	0.9827	0.9828	0.9609	0.9518	0.9794	<b>0.9885</b>



**Fig. 9.** Performance comparison of MEN and other five advanced models on datasets D1 (left) and D2 (right).



**Fig. 10.** ROC curves of the developed MEN and other comparison models on two datasets.

optimize the training process, and aiming at the training data, some filtering techniques can be adopted to alleviate the interference of noises [40–44]. In addition, some data enhancement techniques can

be considered as well [45]. Moreover, a number of optimization algorithms and systems can be used to realize potential better structural configurations so that the established network can exhibit superior



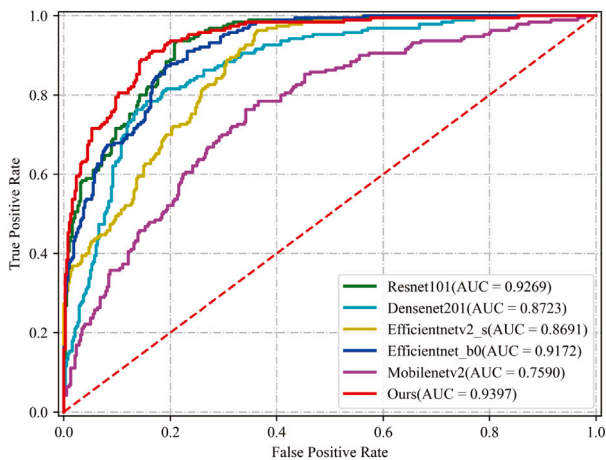


Fig. 11. ROC curves of six models on the supplement dataset.

performance [46–50]. Finally, it is also promising to apply the proposed MEN to more medical diagnostic scenes [51].

## 5. Conclusion

In this paper, a novel multi-scale efficient network has been developed, which is a competent medical imaging analysis framework with efficient feature extraction boosted by the two designed attention blocks. In a progressive learning manner, both fine-grained features and semantic information have been sufficiently obtained, which promotes the improvement of the detection accuracy. The proposed MEN has been comprehensively evaluated on two COVID-19 diagnosis tasks, and comparison results with some other advanced models show that the proposed method has presented overwhelming advantages in terms of all applied evaluation metrics, which demonstrates its effectiveness and competitiveness.

## 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.

## References

- [1] A. Munif, A. Bandar, Detection of COVID-19 using deep learning on X-ray images, *Intell. Autom. Soft Comput.* 29 (2021) 885–898.
- [2] G. Sarra, B. Marwa, A. Adel, K. Anis, H. Habib, Deep learning based detection of COVID-19 from chest X-ray images, *Multimedia Tools Appl.* 80 (2021) 31803–31820.
- [3] S. Cui, C. Liu, M. Chen, S. Xiong, Brain tumor semantic segmentation from MRI image using deep generative adversarial segmentation network, *J. Med. Imag. Health Inform.* 9 (9) (2019) 1913–1919.
- [4] D. Keidar, D. Yaron, E. Goldstein, et al., COVID-19 classification of X-ray images using deep neural networks, *Eur. Radiol.* 31 (2021) 9654–9663.
- [5] G. Rubin, C. Ryerson, L. Haramati, et al., The role of chest imaging inpatient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner society, *Radiology* 296 (2020) 172–180.
- [6] Y. Hong, B. Wei, Z. Han, X. Li, Y. Zheng, S. Li, MMCL-net: Spinal disease diagnosis in global mode using progressive multi-task joint learning, *Neurocomputing* 399 (2020) 307–316.
- [7] R. Rajagopal, R. Karthick, P. Meenalochini, T. Kalaichelvi, Deep convolutional spiking neural network optimized with arithmetic optimization algorithm for lung disease detection using chest X-ray images, *Biomed. Signal Process. Control* 79 (2) (2022).
- [8] H. Li, N. Zeng, P. Wu, C. Kathy, Cov-net: A computer-aided diagnosis method for recognizing COVID-19 from chest X-ray images via machine vision, *Expert Syst. Appl.* 207 (2022) 118029.
- [9] S. Tang, C. Wang, J. Nie, N. Kumar, Y. Zhang, Z. Xiong, A. Barnawi, EDL-COVID: Ensemble deep learning for COVID-19 case detection from chest X-ray images, *IEEE Trans. Ind. Inform.* 17 (2022) 6539–6549.

- [10] M. Szankin, A. Kwasniewska, Can AI see bias in X-ray images? *Int. J. Netw. Dyn. Intell.* 1 (1) (2022) 48–64.
- [11] H. Monday, J. Li, G. Nneji, S. Nahar, M. HossinHappy, J. Jackson, C. Ejayi, COVID-19 diagnosis from chest x-ray images using a robust multi-resolution analysis siamese neural network with super-resolution convolutional neural network, *Diagnostics* 12 (3) (2022) 741.
- [12] W. Bahgat, H. Balaha, A. Yousry, M. Badawy, An optimized transfer learning-based approach for automatic diagnosis of COVID-19 from chest x-ray images, *PeerJ Comput. Sci.* 7 (2021) e555.
- [13] L. Wang, J. Du, H. Zhang, Z. He, Y. jia, Brain MR image super-resolution using 3D feature attention network, in: 2020 IEEE International Conference on Bioinformatics and Biomedicine, 2020, pp. 1151–1155.
- [14] X. Fu, L. Bi, A. Kumar, M. Fulham, J. Kim, Multimodal spatial attention module for targeting multimodal PET-CT lung tumor segmentation, *IEEE J. Biomed. Health Inf.* 25 (9) (2021) 3507–3516.
- [15] M. Farooq, A. Hafeez, COVID-ResNet: A deep learning framework for screening of COVID19 from radiographs, 2020, [arXiv:abs/2003.14395](https://arxiv.org/abs/2003.14395).
- [16] P. Samson, C. Annavarapu, Deep learning-based improved snapshot ensemble technique for COVID-19 chest X-ray classification, *Appl. Intell.* 51 (2021) 3104–3120.
- [17] T. Ozturk, M. Talo, E. Yildirim, U. Baloglu, O. Yildirim, U. Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, *Comput. Biol. Med.* 121 (2020) 103792.
- [18] H. Panwar, P. Gupta, M. Siddiqui, R. Morales-Menendez, V. Singh, Application of deep learning for fast detection of COVID-19 in X-rays using ncvnet, *Chaos Solitons Fractals* 138 (2020) 109944.
- [19] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: *International Conference on Learning Representations*, 2015, pp. 1–14.
- [20] G. Marques, D. Agarwal, I. de la Torre Díez, Automated medical diagnosis of COVID-19 through efficientnet convolutional neural network, *Appl. Soft. Comput.* 96 (2020) 106691.
- [21] Y. Fan, J. Liu, R. Yao, X. Yuan, COVID-19 detection from X-ray images using multi-kernel-size spatial-channel attention network, *Pattern Recognit.* 119 (2021) 108055.
- [22] J. Li, D. Zhang, Q. Liu, R. Bu, Q. Wei, COVID-gatnet: A deep learning framework for screening of COVID-19 from chest X-ray images, in: 2020 IEEE 6th International Conference on Computer and Communications, ICC, 2020, pp. 1897–1902.
- [23] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, 2017, [arXiv:1710.10903v3](https://arxiv.org/abs/1710.10903v3).
- [24] T. Mahmud, M. Rahman, S. Fattah, CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization, *Comput. Biol. Med.* 122 (2020) 103869.
- [25] L. Wang, Z. Lin, A. Wong, COVID-net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images, *Sci. Rep.* 10 (2020) 19549.
- [26] S. Elfving, E. Uchibe, K. Doya, Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, *Neural Netw.* 107 (2018) 3–11.
- [27] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2018, pp. 7132–7141.
- [28] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, 2016, [arXiv:1511.07122v3](https://arxiv.org/abs/1511.07122v3).
- [29] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-net: Efficient channel attention for deep convolutional neural networks, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 11531–11539.
- [30] M. Chowdhury, T. Rahman, A. Khandakar, et al., Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* 8 (2020) 132665–132676.
- [31] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
- [32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen, Mobilenetv2: inverted residuals and linear bottlenecks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2018, pp. 4510–4520.
- [33] G. Huang, Z. Liu, L. Van Der Maaten, K. Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 2261–2269.
- [34] M. Tan, Q. Le, Efficientnetv2: Smaller models and faster training, in: *International Conference on Machine Learning (ICML)*, Vol. 139, 2021, pp. 7102–7110.
- [35] M. Tan, Q. Le, Efficientnet: Rethink model scaling for convolutional neural networks, in: *International Conference on Machine Learning, ICML*, Vol. 97, 2019.
- [36] H. Li, P. Wu, Z. Wang, J. Mao, F. Alsaadi, N. Zeng, A generalized framework of feature learning enhanced convolutional neural network for pathology-image-oriented cancer diagnosis, *Comput. Biol. Med.* 151 (2022) 106265.

- [37] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, X. Liu, A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection, *IEEE Trans. Instrum. Meas.* 71 (2022) 3507014, 1-14.
- [38] N. Yu, R. Yang, M. Huang, Deep common spatial pattern based motor imagery classification with improved objective function, *Int. J. Netw. Dyn. Intell.* 1 (1) (2022) 73–84.
- [39] G. Bao, L. Ma, X. Yi, Recent advances on cooperative control of heterogeneous multi-agent systems subject to constraints: A survey, *Syst. Sci. Control Eng.* 10 (1) (2022) 539–551.
- [40] J. Xu, L. Sheng, M. Gao, Fault estimation for nonlinear systems with sensor gain degradation and stochastic protocol based on strong tracking filtering, *Syst. Sci. Control Eng.* 9 (2) (2021) 60–70.
- [41] J. Mao, Y. Sun, X. Yi, H. Liu, D. Ding, Recursive filtering of networked nonlinear systems: A survey, *Internat. J. Systems Sci.* 52 (6) (2021) 1110–1128.
- [42] Z. Li, J. Hu, J. Li, Distributed filtering for delayed nonlinear system with random sensor saturation: a dynamic event-triggered approach, *Syst. Sci. Control Eng.* 9 (1) (2021) 440–454.
- [43] F. Qu, X. Zhao, X. Wang, E. Tian, Probabilistic-constrained distributed fusion filtering for a class of time-varying systems over sensor networks: a torus-event-triggering mechanism, *Internat. J. Systems Sci.* 53 (6) (2022) 1288–1297.
- [44] H. Geng, H. Liu, L. Ma, X. Yi, Multi-sensor filtering fusion meets censored measurements under a constrained network environment: advances, challenges and prospects, *Internat. J. Systems Sci.* 52 (16) (2021) 3410–3436.
- [45] P. Lu, B. Song, L. Xu, Human face recognition based on convolutional neural network and augmented dataset, *Syst. Sci. Control Eng.* 9 (s2) (2021) 29–37.
- [46] Y. Su, H. Cai, J. Huang, The cooperative output regulation by the distributed observer approach, *Int. J. Netw. Dyn. Intell.* 1 (1) (2022) 20–35.
- [47] X. Wang, Y. Sun, D. Ding, Adaptive dynamic programming for networked control systems under communication constraints: a survey of trends and techniques, *Int. J. Netw. Dyn. Intell.* 1 (1) (2022) 85–98.
- [48] Q. Zhang, Y. Zhou, Recent advances in non-Gaussian stochastic systems control theory and its applications, *Int. J. Netw. Dyn. Intell.* 1 (1) (2022) 111–119.
- [49] B. Song, H. Miao, L. Xu, Path planning for coal mine robot via improved ant colony optimization algorithm, *Syst. Sci. Control Eng.* 9 (1) (2021) 283–289.
- [50] W. Li, Y. Niu, Z. Cao, Event-triggered sliding mode control for multi-agent systems subject to channel fading, *Internat. J. Systems Sci.* 53 (6) (2022) 1233–1244.
- [51] H. Li, P. Wu, N. Zeng, Y. Liu, F. Alsaadi, A survey on parameter identification, state estimation and data analytics for lateral flow immunoassay: from systems science perspective, *Internat. J. Systems Sci.* (2022) <http://dx.doi.org/10.1080/00207721.2022.2083262>.