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.

Biomedical Signal Processing and Control 73 (2022) 103415

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Automatic detection of multiple types of pneumonia: Open dataset and a multi-scale attention network

Pak Kin Wong [a,1], Tao Yan [b,a,1,*], Huaqiao Wang [c,1], In Neng Chan [a], Jiangtao Wang [d], Yang Li [c], Hao Ren [d,*], Chi Hong Wong [e]

[a] Department of Electromechanical Engineering, University of Macau, Taipa 999078, Macau
[b] School of Mechanical Engineering, Hubei University of Arts and Science, Xiangyang 441053, China
[c] Xiangyang No.1 People's Hospital, Hubei University of Medicine, Xiangyang 441000, China
[d] Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science, Xiangyang 441021, China
[e] Faculty of Medicine, Macau University of Science and Technology, Taipa 999078, Macau

ARTICLE INFO

Keywords:
COVID-19
Pneumonia identification
Multi-scale convolution neural network
Attention mechanism
Chest computed tomography

ABSTRACT

The quick and precise identification of COVID-19 pneumonia, non-COVID-19 viral pneumonia, bacterial pneumonia, mycoplasma pneumonia, and normal lung on chest CT images play a crucial role in timely quarantine and medical treatment. However, manual identification is subject to potential misinterpretations and time-consumption issues owing the visual similarities of pneumonia lesions. In this study, we propose a novel multi-scale attention network (MSANet) based on a bag of advanced deep learning techniques for the automatic classification of COVID-19 and multiple types of pneumonia. The proposed method can automatically pay attention to discriminative information and multi-scale features of pneumonia lesions for better classification. The experimental results show that the proposed MSANet can achieve an overall precision of 97.31%, recall of 96.18%, F1-score of 96.71%, accuracy of 97.46%, and macro-average area under the receiver operating characteristic curve (AUC) of 0.9981 to distinguish between multiple classes of pneumonia. These promising results indicate that the proposed method can significantly assist physicians and radiologists in medical diagnosis. The dataset is publicly available at https://doi.org/10.17632/rf8x3wp6ss.1.

1. Introduction

Pneumonia is a common form of lung infection, which can cause serious mortality and morbidity worldwide, especially among children and elderly people [1]. There are more than 30 different causes of pneumonia, and the main types of pneumonia are bacterial pneumonia (BP), viral pneumonia (VP), mycoplasma pneumonia (MP), and other types of pneumonia according to the causative pathogens [2]. The outbreak of COVID-19 pneumonia, which is a type of VP, poses a real threat to all countries and leads to innumerable casualties [3]. The quick and precise identification of the type of pneumonia is imperative, which can guide clinicians in medication and patient management; for example, BP requires an emergency referral for immediate antibiotic treatment, whereas VP requires supportive care [4,5].

Currently, chest computed tomography (CT) is an important non-invasive and effective method of diagnosing multiple types of pneumonia [6,7]. To confirm COVID-19, virus-specific reverse-transcriptase polymerase chain reaction (RT-PCR) is regarded as the gold standard. However, RT-PCR has strict requirements for the laboratory, which may delay the accurate diagnosis of suspected patients [8]. In addition, the SARS-CoV-2 virus of COVID-19 pneumonia can mutate over time, which may cause the RT-PCR to fail to detect COVID-19 patients [9]. Thus, CT examination, which has the advantages of reusability, simple operation, and a high positive rate, is used in many countries to further evaluate COVID-19 and other types of pneumonia.

Visual variations in different pneumonia lesions can be easily observed on chest CT images. Fig. 1 shows some typical CT image slices of multiple types of pneumonia patients, that is, COVID-19, non-COVID-19 VP, BP, and MP. In Fig. 1, lesions of different pneumonia appear at varying scales, shapes, and locations. For example, abnormal findings of ground-glass opacities in patients with COVID-19 are usually multifocal, bilateral, and peripheral, which differs from the diffuse distribution of

* Corresponding authors.
E-mail addresses: yantao@hbuas.edu.cn (T. Yan), angel_angel2345@163.com (H. Ren).
[1] Pak Kin Wong, Tao Yan, and Huaqiao Wang contributed equally to this work.

https://doi.org/10.1016/j.bspc.2021.103415
Received 11 July 2021; Received in revised form 31 October 2021; Accepted 28 November 2021
Available online 9 December 2021

the other non-COVID-19 VP. BP characteristically produces focal segmental (bronchopneumonia) or lobar pulmonary opacities (lobar pneumonia). For patients suffering from MP, CT usually shows centrilobular nodules (tree-in-bud appearance), and bronchial wall thickening is also seen. These different visual features provide a theoretical basis for the classification of different types of pneumonia. Although CT imaging is a good method for the screening of multiple types of pneumonia, CT images of different types of pneumonia also have intra-lesion variances and inter-lesion similarities. These fine-grained characteristics result in a low rate of interobserver concordance in manual diagnosis [6–8,10,11]. In addition, a CT scan contains over 300 image slices, and manual diagnosis is a time-consuming task and sometimes non-reproducible. To solve this problem, a potential solution is to develop an automatic pneumonia detection system using advanced deep learning techniques, which are commonly used for medical diagnosis.

The current deep learning techniques, especially convolutional neural networks (CNNs), have achieved much success in the detection of pneumonia from chest CT images since COVID-19 emerged in early December 2019. Suri et al. [12] and Khanday et al. [13] reviewed artificial intelligence (AI) algorithms used in diagnosing COVID-19 and found that most of the AI algorithms are based on CNNs. In this section, we introduce some representative examples. Li et al. [14] built a diagnostic system that combines several 2D CNN models to classify COVID-19, community-acquired pneumonia (CAP), and non-pneumonia. The system was trained using 1186 CT scans with 132,583 CT images and achieved 90% sensitivity and 96% specificity. This study paved the way for the diagnosis of COVID-19 based on CNNs from chest CT. Zhang et al. [15] applied several CNN models to develop an intelligent system that can identify COVID-19 and differentiate it from common pneumonia and normal lung. Their system was trained using 6752 CT scans from 3777 patients, and achieved satisfactory diagnostic performance. Bai et al. [16] reported that the use of a CNN-based intelligent system could improve the performance of radiologists in distinguishing COVID-19 from CAP on chest CT. Ardakani et al. [17] applied 10 state-of-the-art CNNs to distinguish COVID-19 from non-COVID-19 groups and found that ResNet-101 performed the best with an area under the receiver operating characteristic (ROC) curve (AUC) of 0.994. Ouyang et al. [18] used a dual-sampling attention network to diagnose COVID-19 and CAP. The network was trained using 4,982 CT scans of 3,645 patients and acquired a sensitivity of 86.9%. Rahimzadeh et al. [19] distributed a large dataset of CT images and presented a completely automated approach to detect COVID-19 with high accuracy and speed. Gilanie et al. [20] developed a CNN model to detect COVID-19 from both chest X-ray and CT images with an average accuracy of 96.68 %, specificity of 95.65 %, and sensitivity of 96.24 %. In our previous work [21], we developed a CNN model that can effectively distinguish COVID-19 from other common pneumonia.

Although the above studies have shown that CNNs can achieve impressive diagnostic results, we find that there are still gaps between research and practical applications in pneumonia detection from chest CT images, which are listed below:

(1). Most current studies focus on only one or two specific types of pneumonia, that is, the differentiation of COVID-19 from CAP or normal lung. Consequently, less effort has been made to detect multiple types of pneumonia. For real clinical scenarios, multiple types of pneumonia can be found during CT screening; therefore, a solution for multiclass pneumonia identification is beneficial for clinical applications.

(2). Few studies have considered vital scale information and attention mechanisms to deal with the size and location of pneumonia lesions from the clinical facts that the infection characteristics of COVID-19 and other pneumonia can vary significantly in scale and location depending on the condition of the patients. For instance, in the early stage of COVID-19 infection, lung lesions such as ground-glass opacities on chest CT may be small, subpleural, and peripheral, which need to be analyzed more carefully and require more time. In contrast, lesions such as pulmonary consolidation in the late stage can be easily observed on a coarse scale [6–9]. These radiographic features also appear in other cases of pneumonia [12,13]. Therefore, paying attention to the

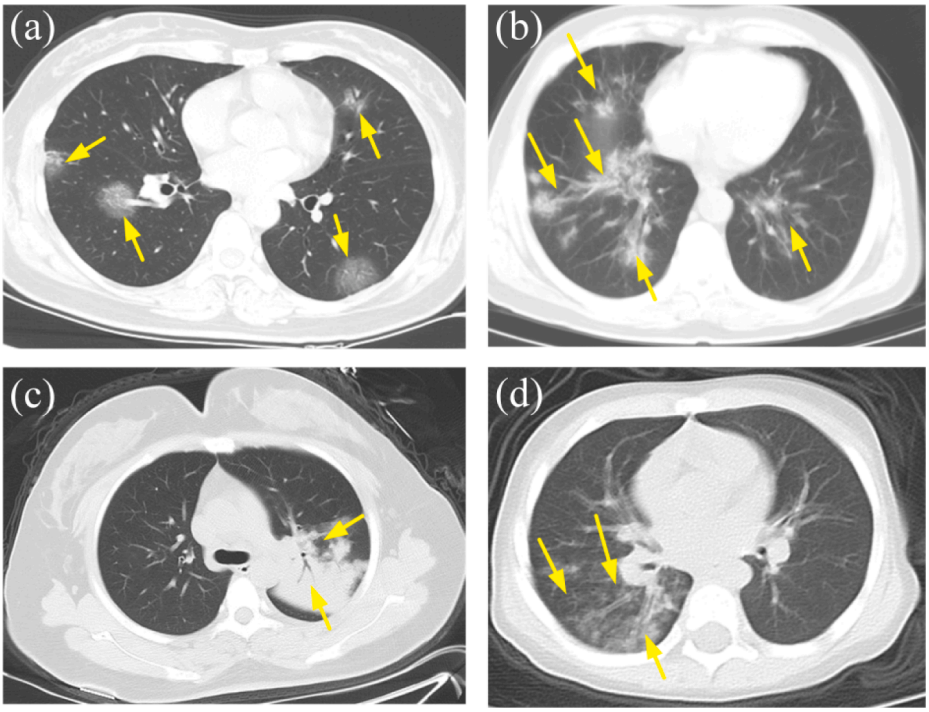

Fig. 1. Representative CT images of different types of pneumonia (The yellow arrows indicate pneumonia lesions in the right and/or left lobes): (a) COVID-19, (b) non-COVID-19 VP, (c) BP, (d) MP.

scale and location characteristics of lesions in CT images is beneficial for the pneumonia classification task.

(3). Clinical data are usually highly imbalanced with varied category distributions, for example, normal samples are more prevalent than COVID-19 infections. This imbalanced data distribution presents a challenge to most classification techniques because they are designed based on the assumption that each category has the same number of samples [22]. As a result, previous classification techniques may make the training ineffective and bias the classification results to the majority class.

This study is motivated by the aforementioned challenges. For the first problem, we collected and constructed a multicenter and multiclass chest CT dataset that contains four types of pneumonia, that is, COVID-19, non-COVID-19 VP, BP, MP, and normal lung. The dataset is named the COVID-19 & community-acquired pneumonia (CCAP) dataset and has been published for further research. To the best of our knowledge, this is the first multiclass CT open dataset for pneumonia. Compared to other open datasets, CCAP has more categories and has been cross-checked. COVID-19 patients were diagnosed as positive using RT-PCR. Other patients were laboratory-confirmed with no-COVID-19 VP, BP, and MP. Two experienced radiologists re-checked all the selected CT scans and images. For the second problem, we developed a multi-scale attention network (MSANet) that generates more discriminative features for better classification of multiclass pneumonia. More specifically, the key part of MSANet is a multi-scale CNN with three spatial attention blocks that focus on discriminative scales and fine-grained features. For the third problem, we apply the multiclass focal loss to address the category imbalance problem.

The remainder of this paper is organized as follows. Section 2 describes the preparation of the dataset. Section 3 describes the proposed methodology. In Section 4, the experiments are conducted, and the performance results are presented. Finally, Sections 5 and 6 present the discussion of the results and conclusions, respectively.

2. Dataset

This retrospective study was approved by the institutional review boards of Xiangyang Central Hospital (XCH) and Xiangyang No.1 People's Hospital. Because this was a retrospective study, the requirement for written informed consent was waived.

From January 1 to May 1, 2020, 206 patients, confirmed COVID-19 positive using RT-PCR were selected, including 104 cases in XCH and 102 cases in Xiangyang No.1 People's Hospital. Even though a patient had more than one CT scan at disease progression, only one typical CT scan with abnormal findings was selected in this study. In addition, 566 patients (60 non-COVID-19 VP, 160 BP, 90 MP, and 256 normal lungs) and their chest CT scans with laboratory-confirmed symptoms from the two participating hospitals between January 1, 2017, and May 1, 2020, were collected. The CT scanners used in the two participating hospitals included Optima CT520 from GE Healthcare, Brilliance iCT (128), and Brilliance CT (64) from Philips Healthcare. Furthermore, CT scans were performed with a peak voltage of 120 kVp with an automatic tube current (50 ~ 300mAs). CT scans were performed on all the patients in a thin section.

Typically, a 3D CNN is used to process patient-level 3D CT scans. However, the retrospective selection of multiclass pneumonia requires numerous CT scans and laboratory and clinical information, which is difficult and time-consuming; thus, the number of collected CT scans is limited. As the 3D CNN contains more network parameters than 2D CNN and requires numerous labeled CT scans to train the parameters, the limited CT data can easily lead to strong overfitting for the data-hungry 3D CNN [18,23]. In addition, previous studies have shown that 3D CNNs have high hardware requirements and computational costs (e.g., GPUs), which leads to inflexibility in applying them to clinical applications [24,25]. Therefore, in this study, we used the 2D CNN framework to make image-level classification, which is also widely used in CT imaging [14–17,19–21,26–28].

To facilitate the above objective, all collected CT scans underwent a preprocessing step and an image-level selection before algorithm training. The entire CT scans were first preprocessed by setting lung window parameters (window width = 1,500 Hu, window level = –600 Hu) to increase the internal contrast of the lung. Subsequently, the entire image slices were extracted to 512 × 512 pixels and standardized by mapping the pixel values from 0 to 255. Approximately 20% of CT image slices without pulmonary parenchyma at the beginning and at the end of one CT scan were removed. Slices containing the pulmonary parenchyma and lesions were selected. All selected image slices were confirmed by two experienced radiologists. Finally, the selected CT scans and the corresponding image slices were randomly split into 60% for building the training dataset, 20% for the building validation set, and 20% for the building test dataset. The division of the enrolled scans and image slices in the CCAP dataset is shown in Table 1.

We also compared our CCAP dataset with other publicly available datasets (Table 2). The SARS-CoV-2 CT scan dataset is a 2D binary dataset collected from the hospitals of Sao Paulo, Brazil [26]. It only contains 2482 CT images, of which 1252 images are from 60 patients infected with COVID-19 and 1230 images are from 60 patients who were not infected with COVID-19 but had other pulmonary. The large COVID-19 CT scan slice dataset is another publicly available dataset proposed by Maftouni et al. [27]. Seven scattered CT datasets were collected and merged to build the dataset. It contains 7593 COVID-19 images, 2618 CAP images, and 6893 normal images. However, some of these images are extracted from research papers on COVID-19 with different resolutions, which may influence the training performance of the deep learning models. The integrative CT images and clinical features of the COVID-19 (iCTCF) dataset is an integrated chest CT image resource for the public [28]. To construct the iCTCF dataset and train the image-level CNN models, 4 radiologists from two hospitals manually labeled 19,685 CT image slices in JPEG format. The China Consortium of Chest CT Image Investigation (CC-CCII) dataset is an open-source chest CT image dataset that encompasses 3 classes of COVID-19, CAP, and normal lung [15]. It is currently one of the largest CT datasets for COVID-19 diagnosis, which contains 617,775 slices of CT images from 6752 scans of 3777 patients. However, the CC-CCII dataset contains some errors (e.g., damaged and disordered image slices, repeated and noisy image slices) that would have negative impacts on the deep learning models. In this study, only image slices with pneumonia lesions were selected and compared. The COVID-CTset dataset is a large chest CT dataset containing 63,849 images from 95 COVID-19 and 282 normal patients [19], and the images of this dataset are 16-bit grayscale in TIFF format. Table 2 shows that our CCAP dataset has more pneumonia categories, which can be used to develop multiclass diagnostic models with higher clinical value. In addition, the CCAP dataset was labeled by two experienced radiologists according to the lesion features and laboratory-confirmed symptoms, which can help develop more accurate diagnostic models. The accuracies in Table 2 are discussed in section 4.4.

3. Methodology

The structure of MSANet, as shown in Fig. 2, consists of four modules; a lung segmentation module, a spatial pyramid decomposition (SPD) module, a multi-scale feature extraction (MSFE) module, and a classification module. Specifically, the lung segmentation module receives CT images and removes noise or irrelevant information to obtain pure lung areas. The SPD module was applied to generate multi-scale inputs with different levels of contextual information. The MSFE module contains three CNN learners with three spatial attention blocks to focus on discriminable multi-scale and fine-grained features for better categorization. A classification module was used to obtain the final prediction. In summary, a raw CT image was first passed through the lung segmentation module to obtain pure lung areas. The SPD module

Table 1
Characteristics of enrolled patients and images.

Class	Training set (~60%)		Validation set (~20%)		Test set (~20%)	
	No. of patients	No. of images	No. of patients	No. of images	No. of patients	No. of images
COVID-19	123	6585	41	2067	42	2035
non-COVID-19 VP	36	2320	12	853	12	844
BP	96	4415	32	1707	32	1644
MP	54	1792	18	867	18	784
Normal	156	7298	50	2265	50	2103
Total	465	22,410	153	7759	154	7410

Table 2
Comparison of CCAP dataset and other open-source datasets.

Paper	2D/3D dataset	Dataset statistics			Classification statics**					Format	Accuracy (%)†	Accuracy (%)‡
		No. of patients	No. of scans	No. of slices	COVID-19	non-COVID-19						
						CAP*			Normal			
						VP	BP	MP				
[26]	2D	120	–	2482	1252	1230				PNG	97.38	97.61
[27]	2D	1130	–	17,104	7593	2618			6893	PNG	95.31	96.18
[28]	2D	104	–	19,685	4001	9979				JPEG	92.18	95.91
[15]	2D/3D	3777	6752	617,775	21,872	36,894			–	PNG	–	92.31
[19]	2D/3D	377	377	63,849	2282	–			9776	TIFF	98.49	98.76
CCAP	2D/3D	772	772	37,579	10,687	4017	7766	3443	11,666	JPG	–	97.46

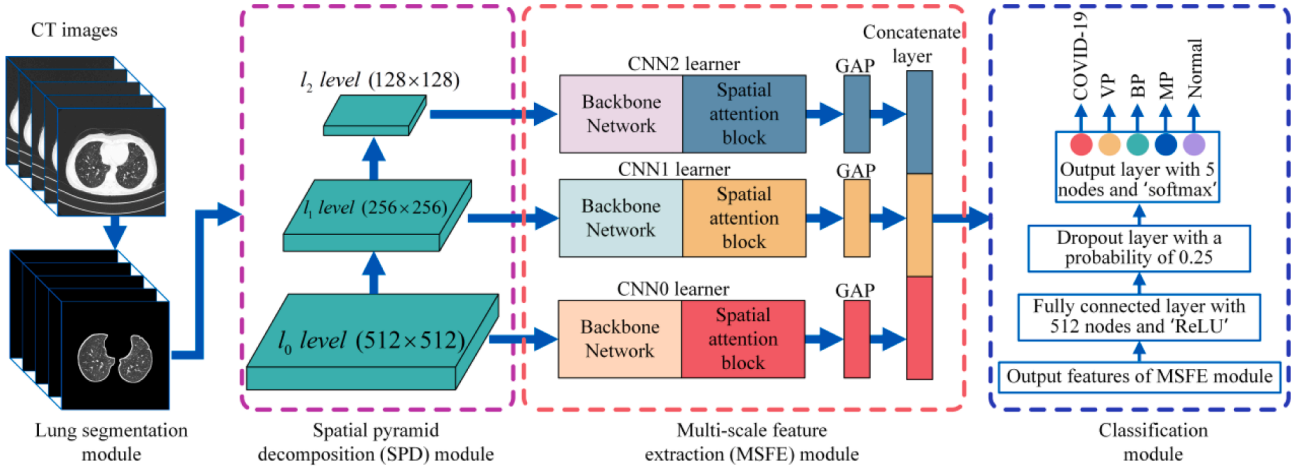

Fig. 2. Architecture of proposed MSANet.

was then applied to produce three images with different scales. Subsequently, the MSFE module was used to capture discriminable features from multi-scale images. Finally, the classification module generated the predicted results.

3.1. Lung segmentation module

Other than lung areas, a lung CT image contains different human tissues (e.g., bone, kidney), part of the CT equipment, and the black background. If we directly feed a raw image for classification, irrelevant information can make the model inaccurate and unreliable. Thus, we annotated 206 CT images and used them to train a standard DeepLabV3 model [29], which can quickly and precisely segment the lung area from the CT images and remove irrelevant parts with a mean intersection-over-union (MIoU) score of 0.9793 on our segmentation test set. An MIoU score close to 1 indicates an excellent segmentation model. Fig. 3 shows typical CT images segmented by the lung segmentation module.

3.2. Spatial pyramid decomposition (SPD) module

The strategy of multi-scale observation is inspired by previous studies in other applications [30,31] and the clinical fact that pneumonia lesions exhibit key radiographic characteristics at different scales. If there are both small and large pneumonia lesions in a CT image, the detection of small lesions (low contrast) usually requires a higher resolution or image scale. In contrast, the detection of large lesions requires a smaller resolution or an image scale. Therefore, SPD was applied to produce multi-scale views of the CT images.

SPD can offer a flexible, convenient, multi-resolution format that emulates multi-scale image processing in the human visual system [32]; it is widely used in the medical imaging community. Suppose a CT image g is represented by a 2-D array; this image is at the zero level ($l = 0$) of the pyramid, and the different levels of the pyramid are calculated as:

$$g_l(m,n) = \sum_{p=-2}^{2}\sum_{q=-2}^{2} w(p,q)\, g_{l-1}(2m+p, 2n+q) \quad (1)$$

where g_l is the image obtained at scale l and m and n are pixel co-

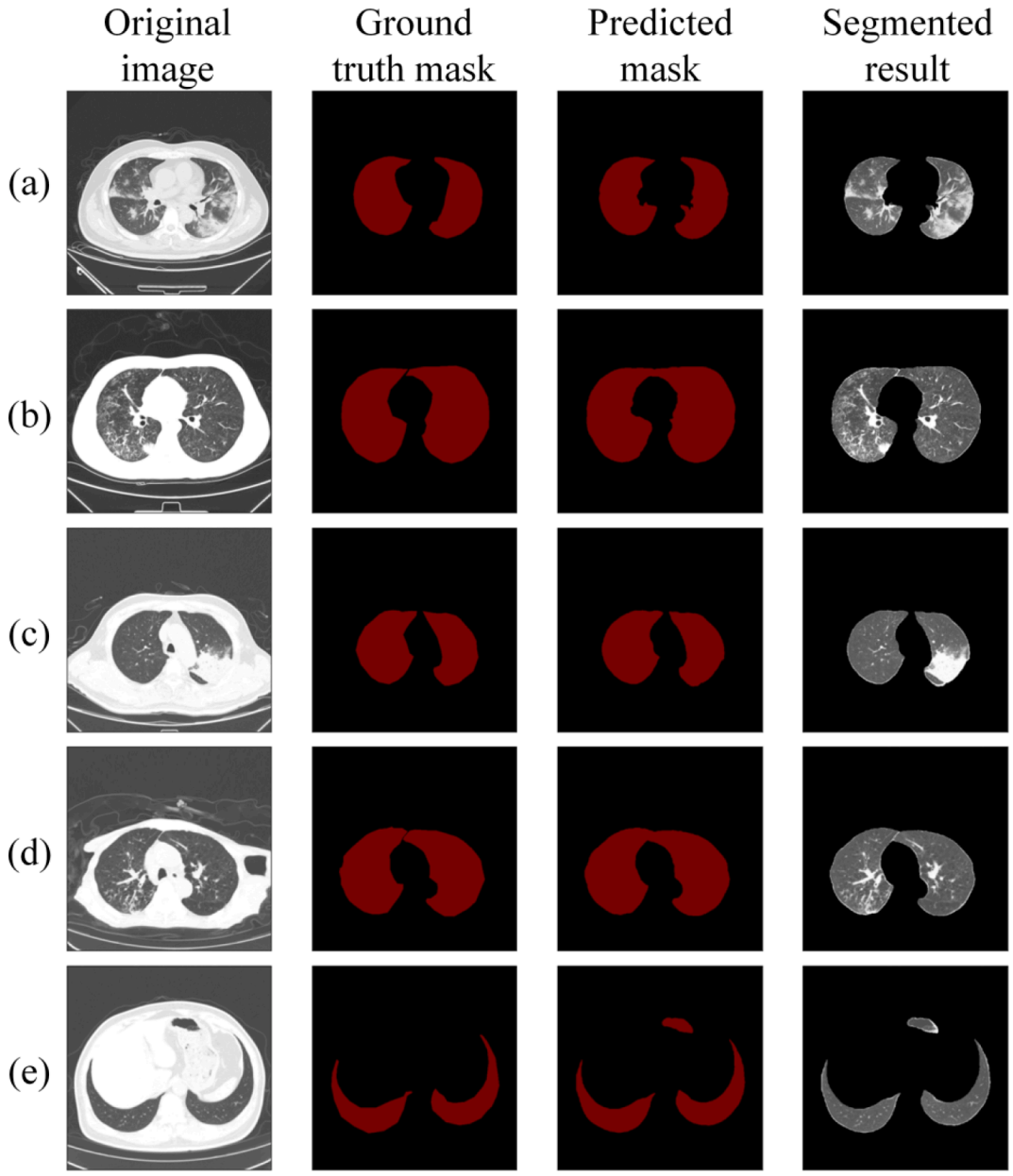

Fig. 3. Typical CT images segmented by lung segmentation module: (a) COVID-19, (b) non-COVID-19 VP, (c) BP, (d) MP, (e) Normal.

ordinates. We set the separable kernel $w(p,q) = w(p)w(q)$ with $w = [(1/4) - (a/4), 1/4, a, 1/4, (1/4) - (a/4)]$ and $a = 0.375$ in this study. The SPD produces image pyramids with 3 levels (i.e., l_0=512 × 512 pixels, l_1=256 × 256 pixels, andl_2 = 128 × 128 pixels). If the scale is more than 3, the image resolution will become too small to capture the necessary diagnostic information [20]. Through the pyramid decomposition of a CT image, we obtain images with different scales or resolutions, which can enhance the accuracy of lesion detection [32].

3.3. Multi-scale feature extraction (MSFE) module

To extract the important features related to pneumonia lesions, we constructed an MSFE module that contains three single-scale CNN learners. The SPD module decomposes a CT image at different scales, then the multi-scale images are fed to the corresponding CNN learners to extract scale-specific information (i.e., the image at l_0 level is fed into the CNN0 learner, at l_1 level is fed into the CNN1 learner, and at l_2 level is fed into the CNN2 learner). After each CNN learner, a global average-pooling (GAP) layer, which can enforce correspondences between feature maps and categories, and avoid overfitting [33], is used to replace the traditional fully connected layers in CNN. Finally, a concatenate layer is applied to combine the outputs of GAP layers along a specified dimension for better feature representation [34]. Essentially, feeding multi-scale images to the MSFE module is equivalent to ensemble learning, which enables a model to recognize lesions across a broad range of scales for better classification. Satisfactory diagnostic results can be achieved without numerous labeled CT images. Extensive studies show that the performance of the overall classifier can be improved if the CNNs in the ensemble model have different learning preferences [35]. The MSFE module with varying input scales met this requirement.

The three CNN learners have the same structure and different input sizes. Each CNN learner contains a backbone network and spatial attention block. The backbone network of CNN learners plays a vital role in determining the performance of pneumonia lesion classification. In this study, several state-of-the-art CNN architectures were compared,

and the best one was selected as the backbone network. The strategy of using the attention mechanism is the clinical fact that multiple types of pneumonia exhibit inter-lesion similarities and huge intra-lesion differences, which make it difficult for CNNs to find discriminative features. Instead of analyzing the entire CT image, experienced radiologists generally assess the diagnosis results by viewing the salient areas. This attention mechanism is developed throughout human evolution, which gradually improves the efficiency and precision of visual information processing [36]. The attention mechanism was successfully applied to CNNs to analyze generic images [37,38] and medical images [39,40]. Inspired by the attention mechanism and imitating the radiologists' diagnostic process, a novel spatial attention block is proposed, which can locate the discriminative regions from the input feature maps to achieve powerful classification models. In addition, it can be combined with any CNN backbone to construct a CNN learner.

Fig. 4 shows the proposed spatial attention. A feature map $F \in \mathbb{R}^{C\times H\times W}$ is given as the output of the backbone network in one CNN learner, where $H \times W$ is the size of the feature map and C denotes the channels. To calculate the spatial attention features $F_{SA} \in \mathbb{R}^{C\times H\times W}$, we first apply a 1×1 convolutional layer with parameters W^1_{SB} on F to yield per-channel attentive maps. A rectified linear unit (ReLU) function was used to increase the nonlinear modeling ability. Thus, the second 1×1 convolutional layer with parameters W^2_{SB} was used to squeeze the attentive maps. Subsequently, the sigmoid function was applied to normalize the attentive map. Finally, element-wise multiplication was performed to obtain the final spatial attention features. The process can be formulated as:

$$\begin{aligned} F_{SB} &= \sigma(Conv2(\delta(Conv2(F)))) \otimes F \\ &= \sigma\left(W^2_{SB}\left(\delta\left(W^1_{SB}F\right)\right)\right) \otimes F \end{aligned} \tag{2}$$

where $W^1_{SB} \in \mathbb{R}^{C\times C\times 1\times 1}$ and $W^2_{SB} \in \mathbb{R}^{1\times 1}$. σ and δ denote the sigmoid and ReLU functions, respectively. *Conv*2 represents the 1×1 convolutional layer, and $\otimes$ denotes element-wise multiplication. In summary, high-level feature maps are first passed through a 1×1 convolutional layer and a ReLU function. Subsequently, the second 1×1 convolutional layer and a sigmoid function are applied to produce the attentive map. Finally, element-wise multiplication is performed on the original feature maps and the attentive map to obtain the final spatial attention features. With the spatial attention block, the representational power of the CNN learners and the MSFE module is enhanced by adaptively focusing on salient parts and reducing the influence of redundant information.

3.4. Classification module

The classification module is a three-layer, fully connected network. The first layer is fully connected with 512 nodes, together with a ReLU function, which is used to improve the capability of nonlinear modeling. The second layer is a dropout layer with a probability of 0.25. It is used in increasing the generalization and control of overfitting [41]. Finally, a new fully-connected layer with 5 output nodes together with a 'softmax' function is appended to generate 5 continuous numbers between 0 and 1, which indicates the probability of each category; the sum of the probabilities of all outputs equals 1.

3.5. Multiclass focal loss

Although the parameters of the MSANet can be easily trained end-to-end by back-propagating the gradients of the classification loss, we still need to pay attention to choosing an appropriate loss function to address the category imbalance problem. As shown in Table 1, our CCAP dataset was imbalanced. The imbalance of the dataset biases the classification results to the majority class, resulting in poor detection of minority ones [42]. To overcome this problem, a multiclass focal loss function is used to train the CNNs. The focal loss was originally designed to handle class imbalance for binary classification in object detection tasks, and we modified and extended it to handle multiclass image classification problems in this study [43].

The focal loss function makes the loss indirectly focuses on challenging classes and down-weights well-classified examples, which is more computationally efficient in addressing the imbalanced issue [43]. To introduce multiclass focal loss, we first introduce the cross-entropy loss function [44] for multiclass classification. The conventional cross-entropy loss ($\mathscr{L}_{\mathrm{CE}}$) is given as:

$$\mathscr{L}_{\mathrm{CE}} = -\sum_{i=1}^{M} \mathrm{t_i} log(\mathrm{y_i}) \tag{3}$$

where M is the number of classes and y_i represents the predicted probability. t_i is the real probability distribution, $t_i = 1$ when i belongs to the true label; otherwise it is 0. $\mathscr{L}_{\mathrm{CE}}$ provides equal weights for classification errors of all classes, which will lead to incorrect classification of the minority class. Our dataset has 5 categories, and each category is imbalanced; thus, we use the multiclass focal loss to handle this imbalanced problem, which can be formulated as:

$$\mathscr{L}_{\mathrm{MFL}} = -\sum_{i=1}^{M} (1-\mathrm{y_i})^{\gamma} \mathrm{t_i} log(\mathrm{y_i}) \tag{4}$$

where $(1-y_i)^{\gamma}$ is a modulating factor with a tunable focusing parameter $\gamma \geqslant 0$ for the cross-entropy loss. Intuitively, hard samples are those with large errors, and the model classifies samples with high probability. When the modulation factor is applied, the loss contribution from challenging samples is increased, this is how the multiclass focal loss handles the imbalanced issue. When $\gamma = 0$, the multiclass focal loss is equivalent to the cross-entropy loss. In this study, we set $\gamma = 2$ according to the experimental results in [43], which performs the best in handling the category imbalance problem.

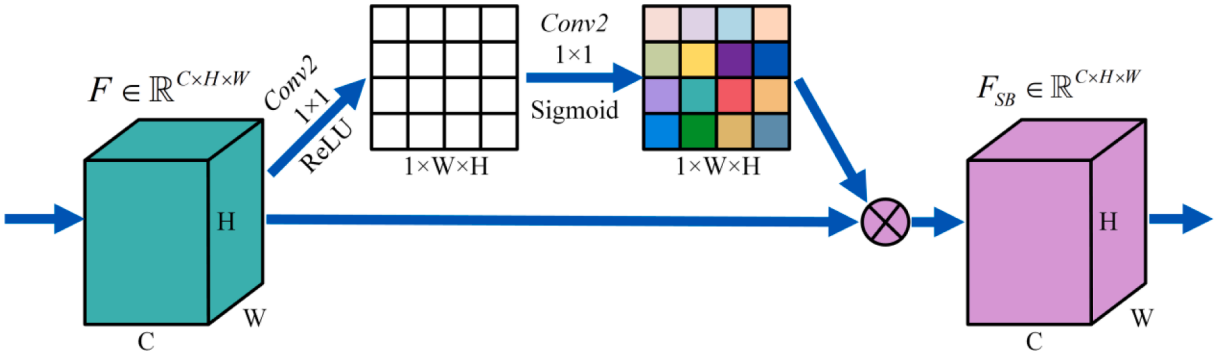

Fig. 4. Architecture of proposed spatial attention block.

4. Experiments and results

4.1. Evaluation metrics

To evaluate the performance of the proposed methods, according to the 5-class confusion matrix, which is used to check any confusion between two classes [45], the values of accuracy, recall, precision, and F1-score are employed. The formulas of these evaluation metrics are shown in Table 3, where TP, FP, TN, and FN refer to the numbers of true positives, false positives, true negatives and false negatives, respectively. Recall measures the proportion of positives correctly predicted, precision evaluates the precision of a model in predicting positive labels, and F1-score is the harmonic average of precision and recall [46]. In addition, the macro-average AUC was used to assess the overall performance. When there is an uneven class distribution, the AUC is a more appropriate statistic. In reality, the value of AUC fluctuates between 0.5 and 1, with a value near 1 indicating a good classifier [43].

4.2. Effect of multiclass focal loss towards classification

The classification results obtained by the conventional cross-entropy loss and multiclass focal loss are compared to demonstrate their effectiveness. For quick verification, we designed a single-scale classifier as the baseline model. The baseline model is constructed using a CNN1 learner with ResNet101 as the backbone; subsequently, a GAP layer is added. After the GAP layer, the classification module in MSANet followed. The weights of ResNet101 were initialized using the weights pre-trained on the ImageNet dataset [47]. The input size is 256 × 256, and the evaluation indexes are precision, recall, and F1-score of each category. A total of 7410 test CT images, including 2035 COVID-19 images, 844 non-COVID-19 VP images, 1644 BP images, 784 MP images, and 2103 normal images were used for evaluation. The training process was executed based on a mini-batch size of 8 using the Adam optimizer with an initial learning rate of 0.0003 for 20 epochs. We used the Keras (https://keras.io/) library on the Tensorflow backend to develop and run the CNNs on an Intel i7 CPU with a GeForce RTX 2080Ti GPU personal computer.

In our CCAP dataset, the number of non-COVID-19 VPs and MPs was relatively small compared with other categories. Table 4 shows that the multiclass focal loss improves the class-wise precision of non-COVID-19 VP from 63.39% to 66.44%, recall from 38.98% to 68.96%, F1-score from 48.28% to 67.67%, and AUC from 0.9204 to 0.9591 as compared with the conventional cross-entropy loss. For MP and other types of pneumonia, most evaluation metrics can achieve a certain degree of improvement when using multiclass focal loss. Hence, the multiclass focal loss can improve the diagnostic performance of the minority classes without influencing the diagnostic performance of the majority classes; thus, the method of multiclass focal loss is applied to train the models in the following section, and all models are trained using the same hyperparameters to provide a fair comparison.

4.3. Selection of backbone networks

Although the major contribution of the proposed MSANet is the use of a multi-scale strategy and attention mechanism, the backbone network in CNN learners also significantly influences the performance of the pneumonia categorization. Thus, in addition to ResNet101 as the backbone discussed in Section 4.2, several state-of-the-art CNN architectures were also compared. In this experiment, we alternately used pre-trained Xception [48], VGG16 [49], InceptionV3 [50], MobileNetV2 [51], DenseNet121 [52], and EfficientNetB0 [53] to replace ResNet101 in the baseline model. Table 5 shows that EfficientNetB0 as the backbone achieves the best accuracy and the average precision, recall, F1-score, accuracy and macro-average AUC are 92.03%, 90.07%, 90.86%, 93.44%, and 0.9887, respectively. Therefore, in this study, EfficientNetB0 was selected as the backbone network to build the CNN learners and MSANet.

Table 3
Evaluation metrics.

Metrics	Calculation equations
Accuracy	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
Recall	$Recall = \frac{TP}{TP + FN}$
Precision	$Precision = \frac{TP}{TP + FN}$
F1-score	$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$

Table 4
Effect of loss function on class-wise performance.

Method	class	Precision (%)	Recall (%)	F1-score (%)	AUC
CNN1 learner with $\mathscr{L}_{CE}$	COVID-19	94.29	91.74	93.00	0.9934
	non-COVID-19 VP	63.39	38.98	48.28	0.9204
	BP	75.23	93.31	83.30	0.9827
	MP	91.95	72.83	81.28	0.9842
	Normal	89.56	95.86	92.60	0.9945
CNN1 learner with $\mathscr{L}_{MFL}$	COVID-19	94.62	94.20	94.20	0.9962
	non-COVID-19 VP	66.44	68.96	67.67	0.9591
	BP	88.12	89.29	88.70	0.9898
	MP	86.52	72.83	79.09	0.9831
	Normal	90.33	93.72	92.00	0.9898

Note: CNN1 learner is ResNet101 as the backbone with an input size of 256 × 256. $\mathscr{L}_{CE}$ denotes conventional cross-entropy loss. $\mathscr{L}_{MFL}$ denotes multi-class focal loss.

4.4. Performance of proposed MSANet

Three single-scale CNN learners equipped with EfficientNetB0 as the backbone were designed to individually classify pneumonia, and they were compared with the proposed MSANet. In addition, to verify the effectiveness of the spatial attention block, we compared the performance of CNNs with and without spatial attention blocks. Furthermore, we analyzed the complexity of the models from the perspective of training time and test time. The performance scores, training times, and test times are listed in Table 6.

Table 6 reveals that CNN0 learner with input sizes of 512 × 512 achieved average precision, recall, F1-score, accuracy, macro-average AUC, training time, and test time of 94.78%, 92.80%, 93.64%, 95.51%, 0.9975, 26664 s, and 569 s, respectively. For CNN1 learner with an input size of 256 × 256, the average precision, recall, F1-score, accuracy, macro-average AUC, training time, and test time were 92.03%, 90.07%, 90.86%, 93.44%, 0.9887,7546 s, and 169 s, respectively. For CNN2 learner with an input size of 128 × 128, the average precision, recall, F1-score, accuracy, macro-average AUC, training time, and test time were 83.04%, 79.92%, 81.16%, 84.94%, 0.9671, 3806 s,

Table 5
Comparison of different backbone networks.

Backbone network	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)	macro-average AUC
Baseline (ResNet101)	85.20	83.80	84.37	87.84	0.9837
Xception	90.13	88.64	89.24	91.73	**0.9907**
VGG16	73.58	72.00	72.11	76.83	0.9264
InceptionV3	90.61	89.60	90.03	92.05	0.9849
MobileNetV2	88.08	87.24	87.44	90.97	0.9898
DenseNet121	91.39	88.67	89.71	92.54	**0.9907**
EfficientNetB0	**92.03**	**90.07**	**90.86**	**93.44**	0.9887

Note: Bold denotes the best.

Table 6
Details and average performance of the proposed methods on the test dataset.

Model	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)	macro-average AUC	Training time* (s)	Test time** (s)
CNN0	94.78	92.80	93.64	95.51	0.9975	26,664	569
CNN0 without SAB	93.67	92.89	93.21	95.53	0.9971	25,806	508
CNN1	92.03	90.07	90.86	93.44	0.9887	7546	169
CNN1 without SAB	91.87	88.83	90.08	93.06	0.9901	7040	161
CNN2	83.04	79.92	81.16	84.94	0.9671	3806	96
CNN2 without SAB	82.75	80.11	81.18	84.25	0.9636	**3542**	**91**
MSANet	**97.31**	**96.18**	**96.71**	**97.46**	**0.9981**	39,727	917
MSANet without SABs	95.28	92.58	93.69	95.60	0.9970	37,599	832

Note: SAB denotes spatial attention block. Bold means the best. *Training time refers to the total time spent for training 20 epochs on the training set and verification set in Table 1. **Test time represents the total time spent for judging 7410 test images in Table 1.

and 96 s, respectively. The experimental results demonstrate that the CNN0 learner with an input size of 512 × 512 obtains the best accuracy among the three single-scale CNN learners. This is because high-resolution images provide more detailed information and microscopic features of pneumonia lesions. High resolution also means that it takes more time to extract features. Compared with the CNN0 learner, the CNN1 learner showed lower performance and shorter training and test times, and the CNN2 learner achieved the shortest training and test times; however, its performance was the worst. This is because the resolution and details of the image decrease with an increase in the scale number. When the three single-scale CNNs were integrated to obtain the MSANet, the average precision, recall, F1-score, and accuracy of the MSANet were better than those of the single-scale CNN learners. However, the training and test times were all the longest. This is because the

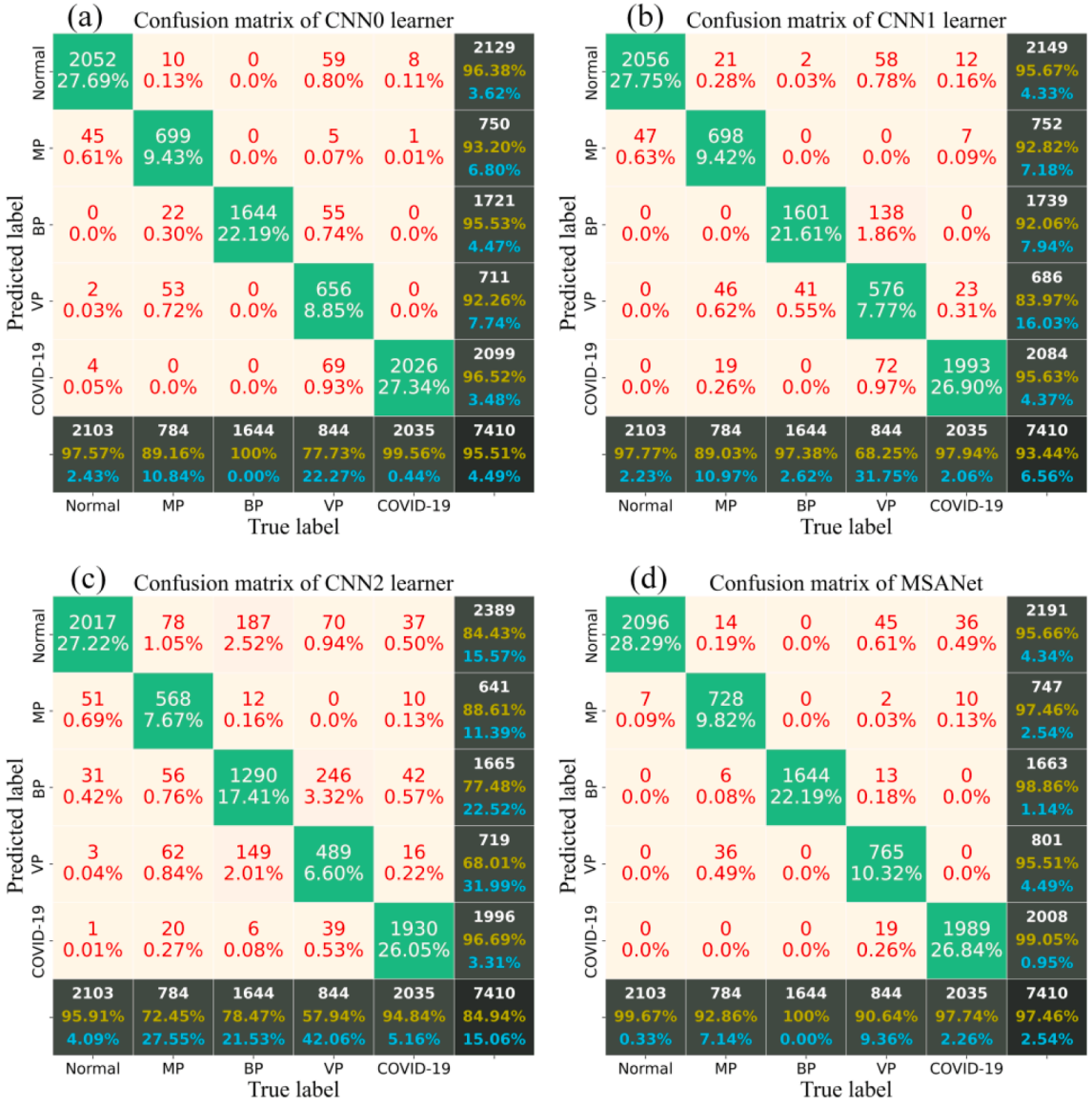

Fig. 5. Confusion matrices of different CNN configurations: (a) Confusion matrix of CNN0 learner, (b) Confusion matrix of CNN1 learner, (c) Confusion matrix of CNN2 learner, (d) Confusion matrix of MSANet.

proposed MSANet can capture the inter-scale variability of the pneumonia lesions, which results in better classification accuracy, whereas the ensemble of multiple single-scale CNNs in the MSANet also requires more computation and consumes more time. Table 6 also shows that when spatial attention blocks are removed from the three single-scale CNN learners, the values of the evaluation metrics, the training time, and the test time are all slightly decreased, which demonstrated that the spatial attention blocks can indeed enhance the feature extraction capability of the CNN, and it also proves the "No Free Lunch Theorem" [54]. In this study, although MSANet is more complex and requires consumes more computation than other single-scale CNN models, its diagnostic performance is the best. Because we need to accurately diagnose each pneumonia patient, especially during the epidemic, MSANet is applied to diagnose multiple types of pneumonia. In addition, the diagnosis time of MSANet is 0.12 s per image, whereas experienced radiologists require an average of 6 s per image, which shows that the proposed MSANet can significantly improve the diagnosis time for pneumonia identification.

Fig. 5(a)-(d) present the image-level confusion matrices of the three single-scale CNN learners and the proposed MSANet, respectively. For the three single-scale CNN learners, there are more cases for them to misclassify non-COVID-19 VP as COVID-19, this situation also appears in identifying MP as non-COVID-19 VP, this is because COVID-19, non-COVID-19 VP and MP share some common radiological characteristics which are difficult to distinguish. When combining multi-scale CNN learners into the proposed MSANet, these mistakes are significantly reduced, and there are only a few misclassifications. However, radiologists also have certain error rates [16]. Fig. 6 presents the ROC curves and AUCs of MSANet. In terms of AUC, the ideal value is 1. The AUC values also revealed that the proposed MSANet can accurately distinguish between various types of pneumonia.

Comparative experiments on other publicly available datasets were conducted to prove the efficiency of the proposed MSANet. Different datasets contain different categories: some are binary and some are ternary datasets. Accordingly, the accuracy, which is also applied to evaluate the deep learning models on all the related datasets [19,26–28], was selected as the evaluation index. All experiments were evaluated on 2D image-level slices in which the pulmonary parenchyma could be captured for judgment. The comparative results in Table 2 show that MSANet can achieve higher accuracy than the existing studies, especially on the iCTCF dataset, which again proves the effectiveness of MSANet. To further prove the generalization of the proposed MSANet, we evaluated its performance on unseen data collected from more patients and sources. The unseen dataset contains 5 COVID-19 scans (253 images), 3 VP scans (281 images), 4 BP scans (172 images), 3 MP scans (127 images), and 6 normal scans (323 images). All 903 non-COVID-19 CT images were acquired from patients who visited the Radiology Center of XCH from May 2021 to June 2021. The 253 COVID-19 images were selected from the CC-CCII dataset [15], as no COVID-19 cases have recently appeared in XCH. The selection criteria for images in the unseen dataset were the same as the training data. For MSANet at slice level analysis on the unseen dataset, the average precision, recall, F1-score, accuracy, macro-average AUC, and diagnostic time were 94.90%, 93.61%, 93.98%, 96.18%, 0.9921, and 143 s, respectively. The diagnostic results show that the proposed MSANet can also achieve satisfactory diagnostic performance on unseen data, which shows that the system can significantly assist physicians and radiologists in the decision-making process.

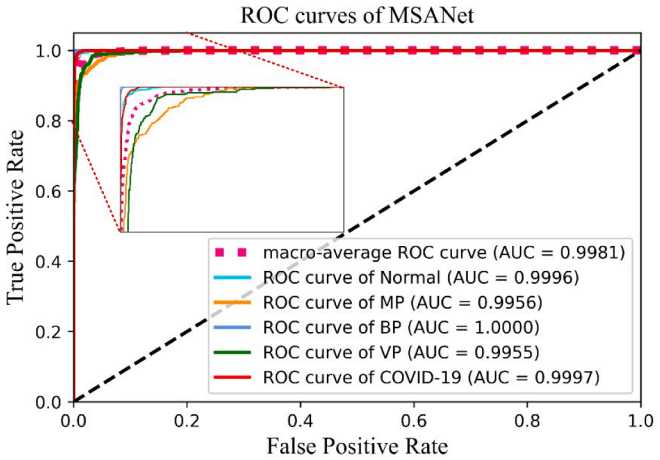

Fig. 6. ROC curves and AUCs obtained using MSANet.

5. Discussion

The outbreak of COVID-19 pneumonia poses a real threat to all countries and leads to innumerable casualties. Early diagnosis and timely treatment can alleviate the spread of the epidemic and decrease mortality [3]. Thus, automatic screening of multiple types of pneumonia and differentiation of COVID-19 pneumonia from non-COVID-19 VP, BP, MP, and healthy lung on chest CT could significantly reduce the effort of the radiologist and accelerate the diagnosis process. However, the manual identification of these types of pneumonia from chest CT is time-consuming and often reduces interobserver variability.

Several artificial intelligence systems, especially deep learning algorithms with CNNs, have been developed to save the effort of the radiologists and accelerate the diagnosis process in this pandemic. While existing studies mostly focus on binary or ternary classification, a system for multiclass pneumonia detection has not yet been developed, which is more important for clinical diagnosis. In addition, to achieve high diagnostic accuracy, existing deep learning methods usually use massive CT data to train deep networks [14–19]. However, the acquisition of a large amount of well-annotated CT data is laborious and tedious for radiologists. Driven by the desire to develop a high-quality diagnostic system for multiple types of pneumonia, an MSANet was developed to reduce the demand for CT images by effectively exploiting the multi-scale features and location characteristics of lesions in CT images. The MSFE module is the key component of MSANet, and it contains three CNN learners equipped with three spatial attention blocks, which focus on the lesion areas to acquire more detailed fine-grained information. Moreover, the MSFE module can accept images of different scales simultaneously to capture multiscale features, which is also a key radiologic marker to distinguish among the different types of pneumonia. Owning to the imbalanced property of the chest CT dataset for multiple types of pneumonia, the multiclass focal loss is used to solve this problem. The experimental results show that the proposed approach is feasible. MSANet can achieve better performance than any single-scale CNN. The MSANe with EfficientNetB0 as the backbone network achieved the best precision (97.31%), recall (96.18%), F1-score (96.71%), accuracy (97.46%), and macro-average AUC (0.9981) in distinguishing multiclass pneumonia.

This study also has some limitations. (1) Although our MSANet works well on the test dataset, it still needs to be tested on more datasets from other hospitals to prove its generalization. (2) For real clinical scenarios, multiple types of pneumonia can be found in one patient; for example, non-COVID-19 VP and BP may concurrently occur in one patient. A solution for joint detection (i.e., multilabel classification) should be considered in the future. (3) The MSANet is an image-level solution, which makes 2D image-level predictions and ignores the inherent spatial coherence of each CT image. A volume-level solution takes 3D CT scans (such as videos) as input and uses a 3D CNN to exploit volume-level information of CT scans, which can perform better than 2D CNN, although this requires numerous CT scans to train the 3D CNN. The collection of a large amount of CT data and the design of an efficient 3D CNN model are also our future study objectives.

6. Conclusions

In this study, a novel MSANet was successfully developed for the automatic differentiation of multiple types of lung pneumonia (i.e., COVID-19, non-COVID-19 VP, BP, MP, and normal lung) from CT images. The main contributions of this study are summarized as follows: (1) Multicentre and multiclass pneumonia CT dataset was constructed and published. The published multiclass dataset is beneficial for clinical applications and promotes research on AI diagnosis of COVID-19 and pneumonia. (2) For the screening problem of multiple types of pneumonia, this is the first study to propose an MSANet based on advanced deep learning techniques and endow it with the ability to capture multiscale features and fine-grade characteristics of different pneumonia lesions. (3) A spatial attention block is proposed to effectively focus on the salient parts and fine-grained characteristics of the CT image. (4) A multiclass focal loss is designed to better handle the data imbalance problem. Comprehensive experiments show that the proposed MSANet can obtain excellent diagnostic accuracy for a few CT images. It is believed that the proposed MSANet can significantly assist in clinical applications.

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.

Acknowledgements

This work was funded by The Science and Technology Development Fund, Macau SAR (File no. 0021/2019/A); and Project of Xiangyang Science and Technology on Medical and Health Field (2020YL12, 2020ZD02).

References

[1] S.N. Grief, J.K. Loza, Guidelines for the evaluation and treatment of pneumonia, Primary Care: Clin. Off. Pract. 45 (3) (2018) 485–503, https://doi.org/10.1016/j.pop.2018.04.001.
[2] Pneumonia. https://www.hopkinsmedicine.org/health/conditions-and-diseases/pneumonia, Accessed 13 June 2021.
[3] N.a. Zhu, D. Zhang, W. Wang, X. Li, B.o. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A novel coronavirus from patients with pneumonia in China, 2019, New Engl. J. Med. 382 (8) (2020) 727–733, https://doi.org/10.1056/NEJMoa2001017.
[4] I. Rudan, C. Boschi-Pinto, Z. Biloglav, K. Mulholland, H. Campbell, Epidemiology and etiology of childhood pneumonia, Bull. World Health Organ 86 (2008) 408–416B, https://doi.org/10.2471/BLT.07.048769.
[5] D.S. Kermany, M. Goldbaum, W. Cai, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, Cell 172 (5) (2018) 1122–1131. e9, https://doi.org/10.1016/j.cell.2018.02.010.
[6] Y.-E. Claessens, M.-P. Debray, F. Tubach, et al., Early chest computed tomography scan to assist diagnosis and guide treatment decision for suspected community-acquired pneumonia, Am. J. Resp. Crit. Care 192 (8) (2015) 974–982, https://doi.org/10.1164/rccm.201501-0017OC.
[7] B. Xu, Y. Xing, J. Peng, et al., Chest CT for detecting COVID-19: a systematic review and meta-analysis of diagnostic accuracy, Eur. Radiol. 30 (10) (2020) 5720–5727, https://doi.org/10.1007/s00330-020-06934-2.
[8] Y. Fang, H. Zhang, J. Xie, et al., Sensitivity of chest CT for COVID-19: comparison to RT-PCR, Radiology 296 (2) (2020) E115–E117, https://doi.org/10.1148/radiol.2020200432.
[9] S. Tahan, B.A. Parikh, L. Droit, M.A. Wallace, C.A. Burnham, D. Wang, SARS-CoV-2 E gene variant alters analytical sensitivity characteristics of viral detection using a commercial RT-PCR assay, J. Clin. Microbiol. 26:JCM-00075 (2021), https://doi.org/10.1128/JCM.00075-21.
[10] J.G. Bartlett, L.M. Mundy, Community-acquired pneumonia community-acquired pneumonia, New Engl. J. Med. 333 (24) (1995) 1618–1624, https://doi.org/10.1056/NEJM199512143332408.
[11] O. Ruuskanen, E. Lahti, L.C. Jennings, D.R. Murdoch, Viral pneumonia, Lancet 377 (9773) (2011) 1264–1275, https://doi.org/10.1016/S0140-6736(10)61459-6.
[12] J.S. Suri, S. Agarwal, S.K. Gupta, et al., A narrative review on characterization of acute respiratory distress syndrome in COVID-19-infected lungs using artificial intelligence, Comput. Biol. Med. 130 (2021) 104210, https://doi.org/10.1016/j.compbiomed.2021.104210.
[13] N.Y. Khanday, S.A. Sofi, Deep insight: Convolutional neural network and its applications for COVID-19 prognosis, Biomed Signal Proces 69 (2021) 102814, https://doi.org/10.1016/j.bspc.2021.102814.
[14] L. Li, L. Qin, Z. Xu, et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, Radiology 296 (2020) E65–E71, https://doi.org/10.1148/radiol.2020200905.
[15] K. Zhang, X. Liu, J. Shen, et al., Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography, Cell 181 (6) (2020) 1423–1433.e11, https://doi.org/10.1016/j.cell.2020.04.045.
[16] H.X. Bai, R. Wang, Z. Xiong, et al., Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT, Radiology 296 (3) (2020) E156–E165, https://doi.org/10.1148/radiol.2020201491.
[17] A.A. Ardakani, A.R. Kanafi, U.R. Acharya, N. Khadem, A. Mohammadi, Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks, Comput. Biol. Med. 121 (2020) 103795, https://doi.org/10.1016/j.compbiomed.2020.103795.
[18] X.i. Ouyang, J. Huo, L. Xia, et al., Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia, IEEE T Med. Imag. 39 (8) (2020) 2595–2605, https://doi.org/10.1109/TMI.2020.2995508.
[19] M. Rahimzadeh, A. Attar, S.M. Sakhaei, A fully automated deep learning-based network for detecting covid-19 from a new and large lung CT scan dataset, Biomed. Signal Proces. 68 (2021) 102588, https://doi.org/10.1016/j.bspc.2021.102588.
[20] G. Gilanie, U.I. Bajwa, M.M. Waraich, et al., Coronavirus (COVID-19) detection from chest radiology images using convolutional neural networks, Biomed. Signal. Proces. 66 (2021) 102490, https://doi.org/10.1016/j.bspc.2021.102490.
[21] T. Yan, P.K. Wong, H. Ren, H. Wang, J. Wang, Y. Li, Automatic distinction between COVID-19 and common pneumonia using multi-scale convolutional neural network on chest CT scans, Chaos Soliton Fract. 140 (2020) 110153, https://doi.org/10.1016/j.chaos.2020.110153.
[22] E. Adeli, X. Li, D. Kwon, Y. Zhang, K.M. Pohl, Logistic regression confined by cardinality-constrained sample and feature selection, IEEE T Pattern Anal. 42 (7) (2020) 1713–1728, https://doi.org/10.1109/TPAMI.2019.2901688.
[23] X.W. Gao, R. Hui, Z. Tian, Classification of CT brain images based on deep learning networks, Comput. Meth. Prog. Bio. 138 (2017) 49–56, https://doi.org/10.1016/j.cmpb.2016.10.007.
[24] X. Qian, H. Fu, W. Shi, et al., M3Lung-Sys: A deep learning system for multi-class lung pneumonia screening from CT imaging, IEEE J. Biomed. Health 24 (12) (2020) 3539–3550, https://doi.org/10.1109/JBHI.2020.3030853.
[25] He X, Wang S, Shi S et al (2020) Benchmarking deep learning models and automated model design for COVID-19 detection with chest CT scans. medRxiv. https://doi.org/10.1101/2020.06.08.20125963.
[26] P. Angelov, E. Almeida Soares, SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification, MedRxiv (2020), https://doi.org/10.1101/2020.04.24.20078584.
[27] M. Maftouni, A.C.C. Law, B. Shen, et al., A robust ensemble-deep learning model for COVID-19 diagnosis based on an integrated CT scan images database, in: Proceedings of the 2021 Institute of Industrial and Systems Engineers Annual Conference (IISE), 2021, pp. 632–637.
[28] W. Ning, S. Lei, J. Yang, et al., Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning, Nat. Biomed. Eng. 4 (12) (2020) 1197–1207, https://doi.org/10.1038/s41551-020-00633-5.
[29] Chen LC, Papandreou G, Schroff F, Adam H (2017) Rethinking atrous convolution for semantic image segmentation. ArXiv preprint arXiv:1706.05587.
[30] R. Rasti, H. Rabbani, A. Mehridehnavi, F. Hajizadeh, Macular OCT classification using a multi-scale convolutional neural network ensemble, IEEE T Med. Imag. 37 (4) (2018) 1024–1034, https://doi.org/10.1109/TMI.2017.2780115.
[31] V. Das, S. Dandapat, P.K. Bora, Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images, Biomed Signal Proces 54 (2019) 101605, https://doi.org/10.1016/j.bspc.2019.101605.
[32] E.H. Adelson, C.H. Anderson, J.R. Bergen, P.J. Burt, J.M. Ogden, Pyramid methods in image processing, RCA Eng. 29 (6) (1984) 33–41.
[33] Lin M, Chen Q, Yan S (2013) Network in network. ArXiv preprint arXiv:1312.4400.
[34] X. Liang, P. Hu, L. Zhang, J. Sun, G. Yin, MCFNet: Multi-layer concatenation fusion network for medical images fusion, IEEE Sens. J. 19 (16) (2019) 7107–7119, https://doi.org/10.1109/JSEN.2019.2913281.
[35] W. Zhang, J. Zhong, S. Yang, et al., Automated identification and grading system of diabetic retinopathy using deep neural networks, Knowl.-Based Syst. 175 (2019) 12–25, https://doi.org/10.1016/j.knosys.2019.03.016.
[36] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE T Pattern Anal 20 (11) (1998) 1254–1259, https://doi.org/10.1109/34.730558.
[37] Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K (2015) Spatial transformer networks. ArXiv preprint arXiv:1506.02025.
[38] C. Xu, M. Lou, Y. Qi, Y. Wang, J. Pi, Y. Ma, Multi-Scale Attention-Guided Network for mammograms classification, Biomed. Signal. Proces. 68 (2021) 102730, https://doi.org/10.1016/j.bspc.2021.102730.
[39] A. He, T. Li, N. Li, K. Wang, H. Fu, CABNet: Category attention block for imbalanced diabetic retinopathy grading, IEEE T Med. Imag. 40 (1) (2021) 143–153, https://doi.org/10.1109/TMI.2020.3023463.
[40] H. Wang, S. Wang, Z. Qin, Y. Zhang, R. Li, Y. Xia, Triple attention learning for classification of 14 thoracic diseases using chest radiography, Med. Image Anal. 67 (2021) 101846, https://doi.org/10.1016/j.media.2020.101846.

[41] N. Srivastava, G. Hinton, A. Krizhevsky, et al., Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn Res. 15 (56) (2014) 1929–1958.
[42] M.M. Rahman, D.N. Davis, Addressing the class imbalance problem in medical datasets, Internat. J. Mach. Learn. Comput. 3 (2) (2013) 224, https://doi.org/10.7763/IJMLC.2013.V3.307.
[43] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal loss for dense object detection, IEEE T Pattern Anal. 42 (2) (2020) 318–327, https://doi.org/10.1109/TPAMI.2018.2858826.
[44] Zhang Z, Sabuncu MR (2018) Generalized cross entropy loss for training deep neural networks with noisy labels. ArXiv preprint arXiv:1805.07836.
[45] S.V. Stehman, Selecting and interpreting measures of thematic classification accuracy, Remote Sens. Environ. 62 (1) (1997) 77–89.
[46] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1) (1977) 159, https://doi.org/10.2307/2529310.
[47] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition (CVPR), 2016, pp. 770–778, 10.1109/CVPR.2016.90.
[48] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1251–1258, 10.1109/CVPR.2017.195.
[49] Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. ArXiv preprint arXiv:1409.1556.
[50] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818–2826, 10.1109/CVPR.2016.308.
[51] M. Sandler, A. Howard, M.L. Zhu, A. Zhmoginov, L.C. Chen, MobileNetV2: inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510–4520, 10.1109/CVPR.2018.00474.
[52] G. Huang, Z. Liu, K. Weinberger, L. van der Maaten, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700–4708, 10.1109/CVPR.2017.243.S.
[53] Tan M, Le QV (2019) EfficientNet: rethinking model scaling for convolutional neural networks. ArXiv preprint arXiv: 1905.11946.
[54] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE T Evolut. Comput. 1 (1) (1997) 67–82, https://doi.org/10.1109/4235.585893.