Automatic distinction between COVID-19 and common pneumonia using multi-scale convolutional neural network on chest CT scans

Tao Yan [a,b], Pak Kin Wong [b,*], Hao Ren [c,*], Huaqiao Wang [d], Jiangtao Wang [c], Yang Li [d,*]

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

ARTICLE INFO

ABSTRACT

The COVID-19 pneumonia is a global threat since it emerged in early December 2019. Driven by the desire to develop a computer-aided system for the rapid diagnosis of COVID-19 to assist radiologists and clinicians to combat with this pandemic, we retrospectively collected 206 patients with positive reverse-transcription polymerase chain reaction (RT-PCR) for COVID-19 and their 416 chest computed tomography (CT) scans with abnormal findings from two hospitals, 412 non-COVID-19 pneumonia and their 412 chest CT scans with clear sign of pneumonia are also retrospectively selected from participating hospitals. Based on these CT scans, we design an artificial intelligence (AI) system that uses a multi-scale convolutional neural network (MSCNN) and evaluate its performance at both slice level and scan level. Experimental results show that the proposed AI has promising diagnostic performance in the detection of COVID-19 and differentiating it from other common pneumonia under limited number of training data, which has great potential to assist radiologists and physicians in performing a quick diagnosis and mitigate the heavy workload of them especially when the health system is overloaded. The data is publicly available for further research at https://data.mendeley.com/datasets/3y55vgckg6/1https://data.mendeley.com/datasets/3y55vgckg6/1.

1. Introduction

The 2019 novel coronavirus (SARS-CoV-2) is a global threat since it emerged in early December 2019 [1]. People infected with SARS-CoV-2 will experience fever, cough, myalgia, headache, and other flu-like symptoms [2]. These virus-induced diseases are collectively named COVID-19 by the World Health Organization [3]. Early detection and treatment of presumptive patients could significantly mitigate the spread of COVID-19 and reduce mortality [1,2].

COVID-19 is typically confirmed by reverse-transcription polymerase chain reaction (RT-PCR) [4]. However, the RT-PCR test has problems such as the insufficient supply of RT-PCR kit, time-consuming and high false negatives, which may cause patients to fail to be diagnosed in time and enter standard treatment procedures [5,6]. At present, many experts have proposed to use the chest computed tomography (CT) to diagnose suspected cases because initial chest CT may present abnormal findings indicating

COVID-19 [7]. CT also has advantages of fast turnaround time, high positive rate and can provide more detailed information related to the pathology [8,9]. Although chest CT has shown great potential for diagnosing COVID-19 pneumonia, manual identification of radiographic features including peripheral ground-glass opacities often has low specificity in distinguishing COVID-19 from other types of pneumonia such as viral pneumonia and bacterial pneumonia [10]. Besides, the rapid growth of COVID-19 patients and multiple CT scans (average 300 slices per scan) of each patient has produced a large number of CT images, which is a huge challenge for radiologists, especially in the epidemic area. A potential solution to identify COVID-19 from the massive number of CT slices quickly and accurately is to develop a computer-aided detection system using convolutional neural network (CNN), which is an emerging artificial intelligence (AI) technique.

CNN is one of deep learning algorithms and can automatically learn the most predictive representations in a manner of layer-by-layer feature combinations while it is also notorious for consuming large amounts of data [11]. The current structure of CNN has achieved much success in the detection of COVID-19 from X-Rays [12], treatment, medication, screening and prediction for the

COVID -19 [13,14], etc. To data, several work groups have implemented the use of CNN for automatic detection of COVID-19 on chest CT [15–18]. Although these CNN-based AI systems have impressive specificity and sensitivity, challenges remain in the use of CNN to screen COVID-19. Firstly, most existing CNN-based methods are employed relatively large datasets. Although CNN can automatically learn the most predictive representations in a manner of layer-by-layer feature combinations while it is also notorious for consuming large amounts of well-annotated data [11]. Sufficient CT samples together with accurately annotated labels it is costly and hard to obtain, especially for some developing countries and small-scale hospitals. Secondly, few studies consider multi-scale features to cope with variations of the size and location of COVID-19 lesions. The infections of COVID-19 in CT images frequently distribute bilaterally, peripherally in lower zone predominant, and the infectious features can vary significantly in scale depending on the condition of the patients [5–10]. For instance, in mild cases, the anomalies look small, while in severe cases they appear scattered and spread over a wide range.

From previous studies in the CT image analysis field, we have found that multi-scale inputs with different levels of contextual information can improve the performance in prediction and classification tasks particularly for complicated problems involving a limited number of images. For example, Wang et al. [19] adopted a new multiscale rotation-invariant CNN model for classifying various lung tissue types on CT images. The model employed a Gabor-local binary pattern that introduces a good property in image analysis-invariance to image scales and rotations. Liu et al. [20] used a multi-scale CNN for lung nodule classification and achieved an error rate of 5.41% and 13.91% for binary and ternary classifications, respectively. Yan et al. [21] proposed a lesion annotation network to extract multi-scale features, experiments show promising qualitative and quantitative results on lesion retrieval, clustering, and classification in CT images.

Thus, the purpose of this study is to develop an AI system based on a multi-scale convolutional neural network (MSCNN) for automatic differentiation of COVID-19 from other common pneumonia. Our major contributions are summarized as follows:

1) We publish a chest CT data set, which includes 416 COVID-19 positive CT scans and 412 common pneumonia (CP) CT scans. Compared with the existing open data set, our data has been double confirmed and more refined. COVID-19 patients were confirmed positive by RT-PCR. CP patients were laboratory-confirmed bacterial pneumonia, mycoplasma pneumonia, fungal pneumonia, and viral pneumonia. All the images contain lesions were re-confirmed by two experienced radiologists.

2) We proposed a novel MSCNN architecture that learns feature representations of multi-scale inputs that can achieve better performance without large-scale training data.

3) Through multi-scale spatial pyramid decomposition, data augmentation, transfer learning, and other strategies, our AI system achieves comparable diagnostic performance than experienced radiologists.

2. Related works

This section presents related works in terms of automated COVID-19 screening on chest CT images and involved methods of our work.

2.1. Automatic screening of COVID-19 on chest CT

There are already some published studies about the automatic screening of COVID-19 on chest CT. Here we briefly review some representative studies. Zhang et al. [15] used a total of 617,775 CT images from 4154 patients to develop a CNN model that can accurately diagnose COVID-19 and differentiate it from other pneumonia, their AI system can identify and quantify some typical features indicating COVID-19 from CT images, including ground-glass opacities, multifocal patchy consolidation, and/or interstitial changes with a peripheral distribution. Li et al. [16] used 4356 chest CT scans from 3322 patients to develop a 3D CNN model to achieve a per-exam sensitivity of 90% and a per-exam specificity of 96% in detecting COVID-19 from community-acquired pneumonia. Ouyang et al. [17] collected 4982 CT scans from 3645 patients provided by 8 collaborative hospitals to develop a dual-sampling attention network to diagnosis COVID-19 from community-acquired pneumonia, the algorithm can identify the COVID-19 images with the area under the receiver operating characteristic curve (AUC) value of 0.944, the accuracy of 87.5%, the sensitivity of 86.9% and specificity of 90.1%. Bai et al. [18] based on the cohort of 1186 patients (132,583 CT images) and EfficientNetB4 to develop a 2D COVID-19 diagnosis system which achieved a test accuracy of 96%, sensitivity 95% and specificity of 96% with AUC of 0.95 and Precision-Recall (PR) AUC of 0.90.

Although the above studies have demonstrated promising results by using chest CT for the diagnosis of COVID-19, most existing methods are based on large well-annotated datasets. The selection and labeling of data require considerable manpower, however, at such an outbreak situation radiologists have limited time to perform the tedious manual drawing, researchers have to wait a long time before they gather enough data to train a high-quality model. In this study, we will use a multi-scale convolutional neural network to overcome the lack of data.

2.2. Multi-scale convolutional neural network for CT image analysis

One challenge in medical image domain is that regions of interest are often scale-invariant, i.e., visually similar patterns often appear in varying scales, multi-scale CNN has been successfully utilized to learn scale-invariant patterns in a variety of medical image analysis tasks, such as breast MRI malignancy classification [22], cancer subtype classification from histopathological images [23], macular optical coherence tomography image classification [24], etc. Here we focus on review multi-scale CNN used in CT images[19–21,25–27]. Kim et al. [25] developed a multi-scale gradual integration CNN for the false-positive reduction in pulmonary nodule detection, in their experiments on the LUNA16 challenge datasets, the model achieved the highest performance. Liu et al. [26] proposed a novel multi-view multi-scale CNN for lung nodule type classification from CT images, the experimental results shown the promising classification performance even with complex ground-glass opacities and non-nodule types. Shen et al. [27] applied a hierarchical multi-scale CNN to capture nodule heterogeneity by extracting discriminative features from alternatively stacked layers. The Experimental results demonstrate the effectiveness of their method on classifying malignant and benign nodules without nodule segmentation.

Inspired by the above-mentioned researches, in this study we propose a novel multi-scale CNN for the diagnosis of COVID-19. To fully explore multiple features describing COVID-19 from multi-scales, we will be based on multi-scale spatial pyramid decomposition and the latest CNN algorithm to develop the MSCNN model.

3. Materials and methods

3.1. Dataset preparation

This study was approved by the institutional review board of Xiangyang Central Hospital and Xiangyang No.1 People's Hospital in the Hubei province of China. Written informed consent

Table 1
Demographics of selected patients with COVID-19 and other common pneumonia.

Overall	COVID-19 $n = 206$	CP $n = 412$	p-value
Age (year)			<0.001
Mean age	44±17	57±18	
<60	148 (72)	222 (54)	
≥60	58 (28)	190 (46)	
Sex			0.306
Male	97 (47)	213 (52)	
Female	109 (53)	199 (48)	
Xiangyang Central Hospital			
Age (year)			<0.001
Mean age	46±15	60±17	
<60	74 (71)	136 (55)	
≥60	30 (29)	112 (45)	
Sex	0.641		
Male	50 (48)	127 (51)	
Female	54 (52)	121 (49)	
Xiangyang No.1 People's Hospital			
Age (year)			<0.001
Mean age	43±20	52±19	
<60	74 (73)	72 (44)	
≥60	28 (27)	78 (56)	
Sex			0.377
Male	47 (46)	86 (52)	
Female	55 (54)	78 (48)	

Note: Value in the parenthesis represents a percentage. Ages are reported as mean± standard deviation. p-values about the sex and age distributions are calculated by the chi-square test and student *t*-test, respectively. COVID-19: Coronavirus disease 2019. CP: common pneumonia. n: number of patients.

was waived by the institutional review board for the retrospective study. We retrospectively acquired 416 three-dimensional (3D) chest CT scans with abnormal findings from 206 patients who were confirmed positive COVID-19 by RT-PCR between January 1 and May 1, 2020. The same patient underwent one or multiple CT scans at various timelines during the course of the disease (The average CT scan per patient is 2, with a range from 1 to 5). A total of 412 patients and their chest CT scans (one scan per patient) with laboratory-confirmed were retrospectively identified from the participating hospitals between January 1, 2018, and December 1, 2019. These CP patients were randomly selected and consisted of non-COVID-19 viral pneumonia (such as influenza virus), bacterial pneumonia, and fungal pneumonia. The selected patients and their CT scans are randomly divided into 80% for training, 10% for validation, and 10% for testing. Furthermore, the test cases of CT scans are selected from patients who have not been included in the training stage. The demographics of the selected patients is summarized in Table 1. No significant difference exists between COVID-19 and CP groups in terms of sex distribution (*p*-value > 0.05). However, the average age of patients with COVID-19 is lower than that of patients with CP (*p*-value < 0.001).

Lung window (window width = 1500 Hu, window level = −600 Hu) is performed over all slices of the 828 CT scans to increase the internal contrast of the lung. The entire image slices are transferred to JPG and normalized by mapping pixel values in the range of 0–255. Since the AI system is trained by 2D images, so a slice level selection is performed, an image slice contains pulmonary parenchyma and with lesions (COVID-19 or CP) is selected out. Two experienced radiologists review the final selected CT slices and reach a decision in consensus. Sample exclusion, inclusion, and distribution of the sample data are described in Fig. 1. Fig. 2 shows some typical images for testing our AI system. It can be observed from Fig. 2 that COVID-19 and CP are not easy to be distinguished.

3.2. Architecture of proposed AI system

Our idea of the multi-scale strategy is motivated by the clinical fact that radiographic features of COVID-19 have varied shapes, locations, and sizes. The peripheral ground-glass opacities in the early stage are very small in size and need to be analyzed at a finer scale, the pulmonary consolidation in the late stage can be visualized in a coarse-scale [5–10]. Besides, from an information quantity perspective, it may be reasonable to use morphological and structural features in different scales and thus effectively integrating multi-scale contextual information [25,27].

According to the above motivations, we introduce the multi-scale spatial pyramid (MSSP) decomposition [28] to create multi-scale views of the CT image to capture key multi-scale information for better classification even though the training data is small. The MSSP creates an image pyramid by the reduced and Gaussian low pass filtered versions of the image of the previous level [28]. Although CNN algorithms benefit from spatial pooling for providing some inherent invariance to distorted, scaled, and translated inputs [29], but in the ensembled MSCNN module, the MSSP decomposition can be used to reduce the time complexity and effective parameters of the overall model to reduce the chance of overfitting and to obtain a promising performance in practice [24]. In this study, three levels of the multi-scale version of Gaussian low-pass image pyramids are considered to simulate the multi-scale perception of the proposed MSCNN module, we chose three levels because the image resolution becomes too small to capture essential diagnostic information. To do this, the image pyramids are calculated for each slice using a symmetric pyramidal decomposition method. Suppose that a CT image g is at the zero level ($l = 0$) of the Gaussian pyramid, the multi-scale image at the level l given the image at level (l-1) is calculated as Eq. (1), where g_l is the image obtained at scale l and w is the kernel function.

$$g_l(i, j) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} w(m, n) g_{l-1}(2i + m, 2j + n) \tag{1}$$

The AI system contains an MSSP decomposed module and an MSCNN module. The MSSP decomposed images are then fed to the proposed MSCNN module. The MSCNN module is ensembled by three different CNNs which have the same structure but different input scales. The images at different scales (g_l, $l \in \{0, 1, 2\}$) are fed to different CNNs (CNN$_l$, $l \in \{0, 1, 2\}$) for extracting scale specific information. The architecture of the proposed AI system is shown in Fig. 3. In this study, EfficientNetB0 is the backbone of the three CNNs. The latest EfficientNetB0 algorithm is pre-trained to achieve 93.5% (top-five) accuracy on 1000 categories of ImageNet and its params size is just 5.3 Megabytes [30]. The main building block of EfficientNetB0 is the mobile inverted bottleneck which is tuned more efficiently by carefully balancing network depth, width, and resolution, resulting in better performance. To construct the MSCNN module, the last fully connected layer of each EfficientNetB0 is firstly dropped and a global average-pooling layer, which can capture more informative features by enforcing the correspondence between features and classes, is added. Then, a dense layer with a dimension of 64 is added, followed by the addition of a dropout layer with a probability of 0.5 which is used to reduce overfitting. After this, the outputs of three CNNs are concatenated for better classification. Finally, a new fully-connected layer with one output node is added with a sigmoid activation function to generate continuous numbers between 0 and 1 which indicates the probability for COVID-19. To secure high sensitivity for COVID-19 diagnosis, the cut-off number is set to be 0.5. This means that if the final score is greater than 0.5, the slice is diagnosed as COVID-19.

To improve the interpretability of the AI system, the heat map calculated by output feature maps of the last convolutional layer of each EfficientNetB0 is overlaid upon the original images to obtain an attention map. An attention map can present to the doctor about the important areas that have significant contribution to the
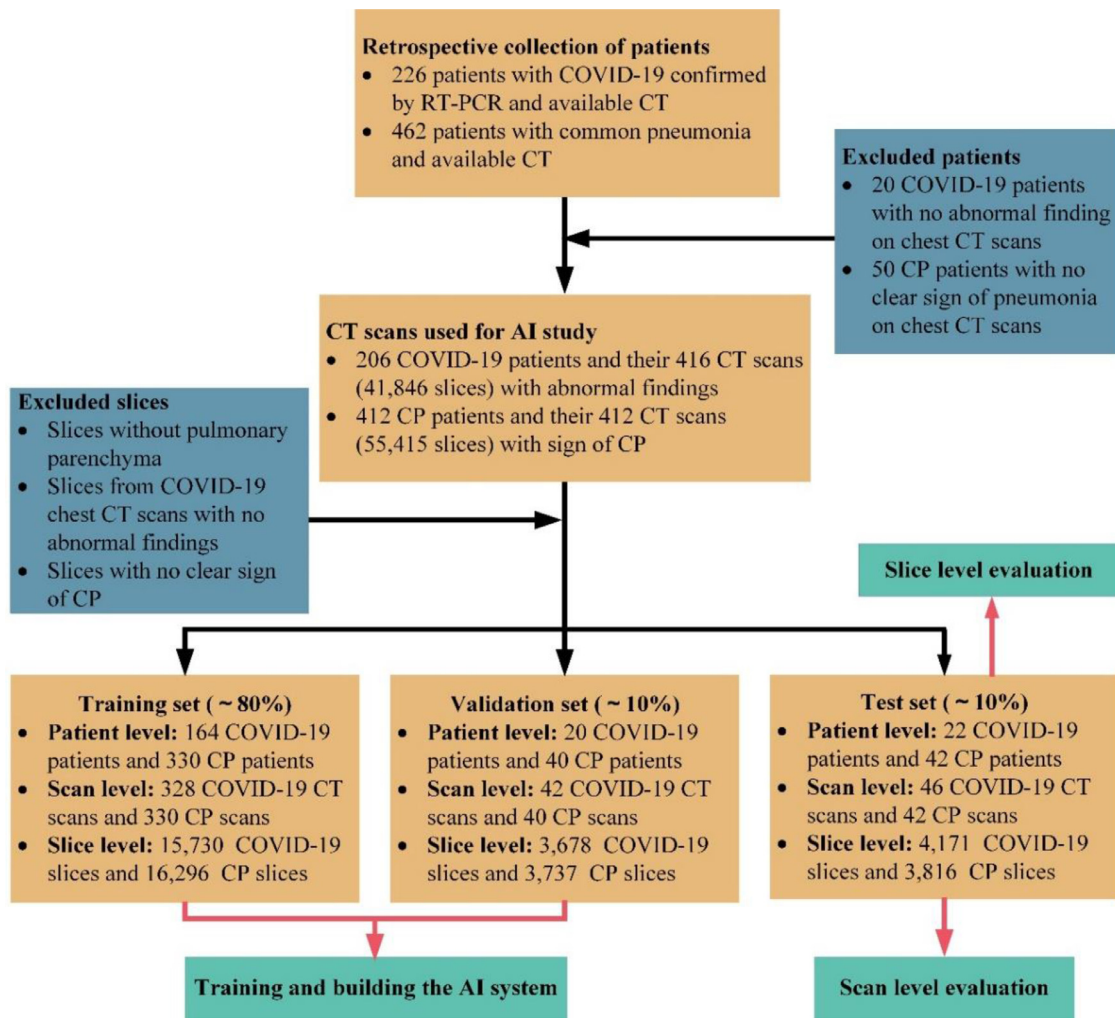
Fig. 1. Sample exclusion, inclusion, and distribution for construction of proposed AI system. COVID-19: Coronavirus disease 2019; CP: Common pneumonia; RT-PCR: Reverse-transcription polymerase chain reaction; AI: Artificial intelligence.

classification result in the CT slices. Different from the previous researches [31], in this study, each CNN in the MSCNN module generates a heat map. Then these heat maps are resized to the same size and the corresponding pixel values are averaged to generate a new heat map. Finally, this new heat map will overlay the original image to get more accurate attention regions which is a pioneering work. To alleviate the scarcity of training data and improve diagnostic performance, data augmentation is performed during the training process to increase the size of the training samples, each training image is randomly rotated, cropped, flipped, shifted and zoomed [32]. The system is built by Keras (https://keras.io/) and trained on an Intel i7 CPU with a GeForce RTX 2080Ti GPU personal computer.

3.3. Experiment design

Two rounds of binary classification experiments are carried out to evaluate the AI system. We start with an evaluation of the ability to detect slice level COVID-19 since the training is on 2D slice level and tuning of hyper-parameters is easier than 3D scan level and. In fact, we have also tried a 3D classification network, but the performance is not satisfactory due to the limited number of training CT scans and limited memory of GPU. At the slice level analysis, we will compare the diagnostic performance of different CNNs in the MSCNN module. Moreover, we also proceed with an evalu-

ation of the detection ability at the scan level for COVID-19 vs CP patients because the radiologist diagnosis is on scan level, which is consistent with clinical practice. As one scan is COVID-19 positive when any one of its slices is also COVID-19 positive, the top 3 highest scores of all slices of a scan are averaged as the scan-level score [33]. As a result, though training and validation are done on the slice level, the AI system can take the whole 3D CT scan into account and generate a single prediction on the scan-level. Two radiologists with more than 15 years of chest CT diagnosis experience individually evaluate the CT scans in the test set and compare their average performance with the AI system.

3.4. Statistical analysis

The diagnostic performances of the AI system and radiologists are evaluated using sensitivity, specificity and accuracy together with their 95% confidence intervals. These three metrics are calculated as follows: sensitivity = (true positive identified)/(true positive samples + false negative samples); Specificity = (true negative identified)/(false positive samples + true negative samples); Accuracy = (true positive identified + true negative identified)/(all samples). Additionally, the AUC values are also calculated according to the area under the receiver operating characteristic (ROC) curves, if the AUC is close to 1, it indicates an excellent classifier. The sex and age distributions among COVID-19 and CP groups are
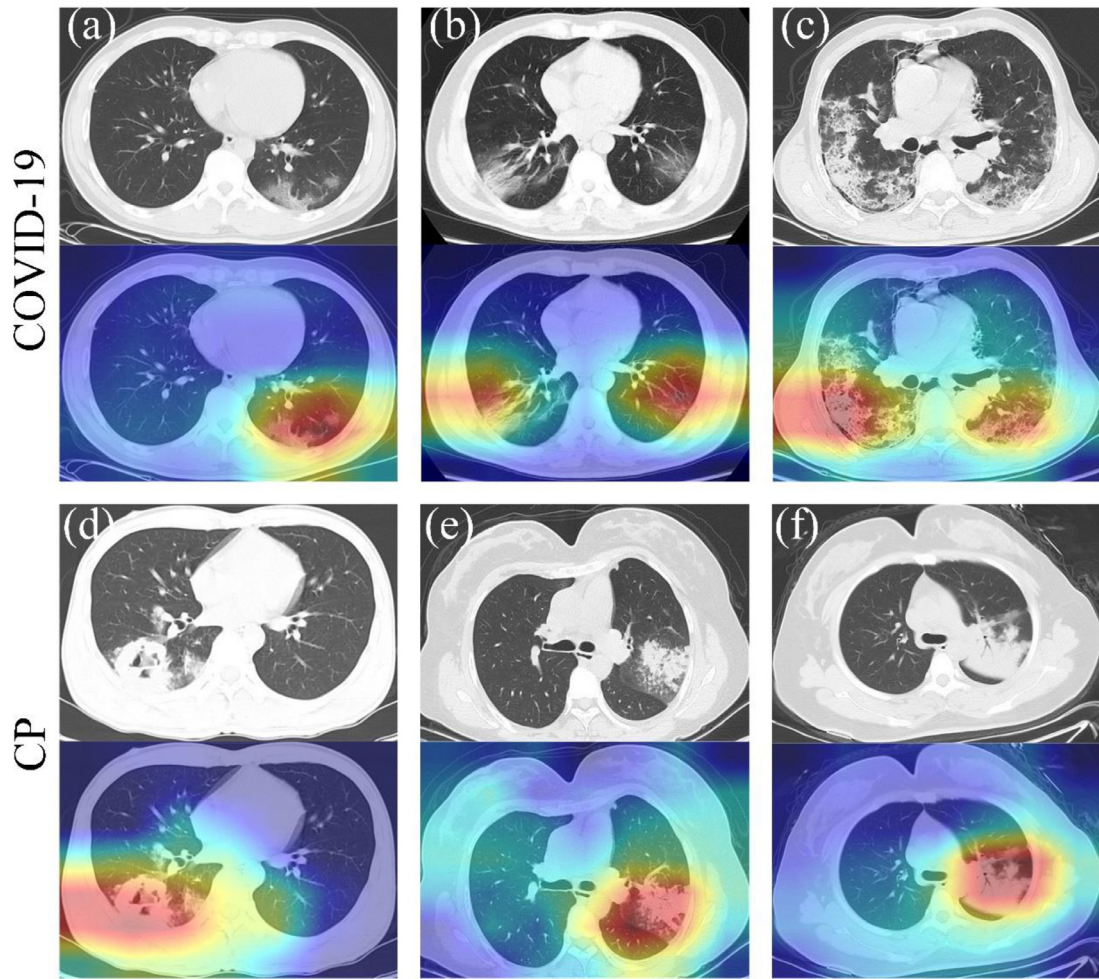
Fig. 2. Representative examples and attention maps that highlight the region of lesions for CT slice with COVID-19 and CP. (a) A 35-year-old male with confirmed COVID-19. (b) A 53-year-old male with confirmed COVID-19; (c) A 67-year-old female with confirmed COVID-19; (d) A 24-year-old male with confirmed CP; (e) A 36-year-old female patient with confirmed CP; (f) A 21-year-old female patient with confirmed CP. COVID-19: Coronavirus disease 2019; CP: Common pneumonia.
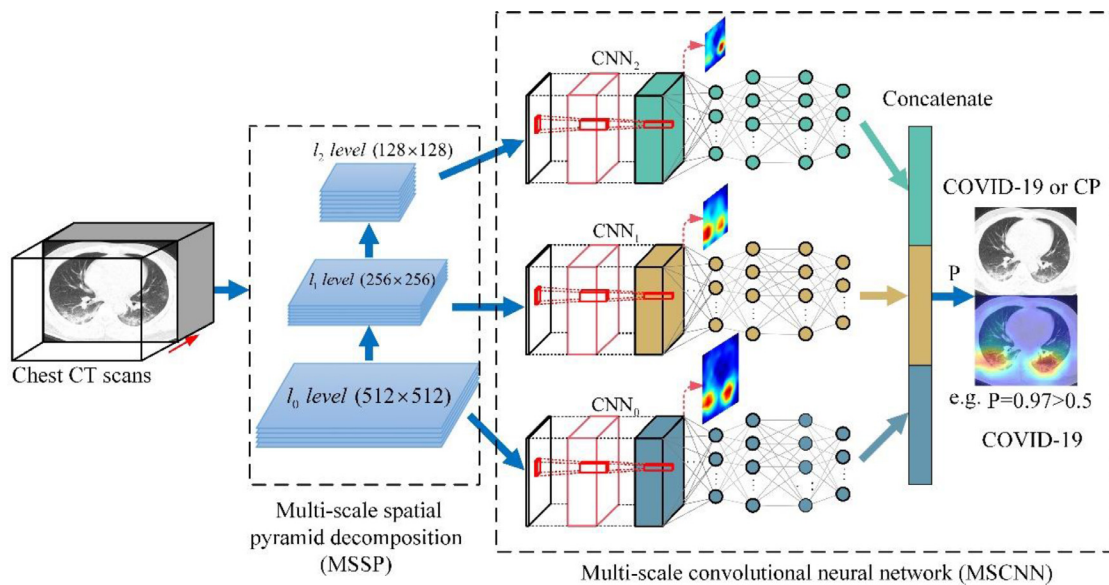
Fig. 3. Architecture of proposed AI system. P refers to the probabilities of COVID-19 predicted by the MSCNN; COVID-19: Coronavirus disease 2019; CP: Common pneumonia.

Table 2
Results of slice level analysis of three CNNs and AI system.

	CNN_0	CNN_1	CNN_2	AI system
Per slice sensitivity,%(95% CI)	96.5 (4023/4171)(95.8–97.0)	99.3 (4140/4171)(98.9–99.5)	81.6 (3405/4171)(80.4–82.8)	99.5 (4152/4171)(99.3–99.7)
Per slice specificity,%(95% CI)	96.2 (3669/3816)(95.5–96.7)	93.0 (3548/3816)(92.1–93.8)	93.6 (3573/3816)(92.8–94.4)	95.6 (3669/3816)(94.9–96.2)
Per slice accuracy,%(95% CI)	96.3 (7692/7987)(95.9–96.7)	96.3 (7688/7987)(95.8–96.7)	87.4 (6978/7987)(86.6–88.1)	97.7 (7800/7987)(97.3–98.0)
AUC	0.952	0.951	0.944	0.962

Note: Per slice sensitivity is the ratio of true positive identified to all CT slices with COVID-19 lesions; Per slice specificity is the ratio of true negative identified to all CT slices without COVID-19 lesions; Per slice accuracy is the ratio of all true values identified to all CT slices. AI: artificial intelligence; CI: confidence interval; AUC: area under the receiver operating characteristic curve.

evaluated by the chi-square test and student *t*-test, respectively. A 2-sided McNemar test is used to compare the performance difference between the AI system and human experts. A p-value less than 0.05 is considered statistically significant. Confusion matrixes are also plotted to make it easy for checking any confusion between two classes (i.e. mislabeling one as another). All statistical analyses are performed by using Python 3.7.6 and Sklearn 0.22.1.

4. Experimental results

4.1. Binary classification at slice level

A total of 7987 test CT slices, including 4171 slices with abnormal findings from 46 chest CT scans of 22 COVID-19 patients and 3816 slices with clear sign of CP from 42 chest CT scans of 42 non-COVID-19 patients, are used for a slice level analysis. The performance of the three classifiers and the AI system at the slice level is shown in Table 2. For CNN_0, the per slice sensitivity, specificity and accuracy are 96.5% (95% confidence interval (CI), 95.8–97.0), 96.2% (95% CI, 95.5–96.7) and 96.3% (95% CI, 95.9–96.7), respectively. The per slice sensitivity, specificity and accuracy of CNN_1 are 99.3% (95% CI, 98.9–99.5), 93.0% (95% CI, 92.1–93.8) and 96.3% (95% CI, 95.8–96.7), respectively. CNN_2 achieves a per slice sensitivity, specificity and accuracy of 81.6% (95% CI, 80.4–82.8), 93.6%

(95% CI, 92.8–94.4) and 87.4% (95% CI, 86.6–88.1), respectively. The AUC values of the three CNNS are 0.952, 0.951, and 0.944, respectively. Among the three classifiers, CNN_1 has the highest sensitivity and CNN_0 achieves the highest specificity. When we test the AI system, the AI system could diagnose COVID-19 with a per slice sensitivity, specificity, and accuracy of 99.5% (95% CI, 99.3–99.7), 95.6% (95% CI, 94.9–96.2) and 97.7% (95% CI, 97.3–98.0), respectively. The three diagnostic performance indexes of the AI system are better than those of the independent CNN. Fig. 4(a)–(e) shows the confusion matrices of the three CNNs and AI system at slice level analysis. The aforesaid confusion matrices depict that there are still some cases for the AI system to misclassify CP as COVID-19. Nevertheless, experienced radiologists also have similar rates of errors.

4.2. Binary classification at scan level

To further validate the effectiveness of the proposed AI system and align with clinical practice. scan level analysis is performed. 46 chest CT scans from 22 COVID-19 patients and 42 chest CT scans from 42 CP patients are used for scan level analysis. Two experienced radiologists are evaluated with the same test dataset. All the radiologists are blinded to the results of the AI diagnosis. In the human-machine contest, 41 out of 46 COVID-19 positive CT scans are correctly detected by the AI system and 36 out of 42
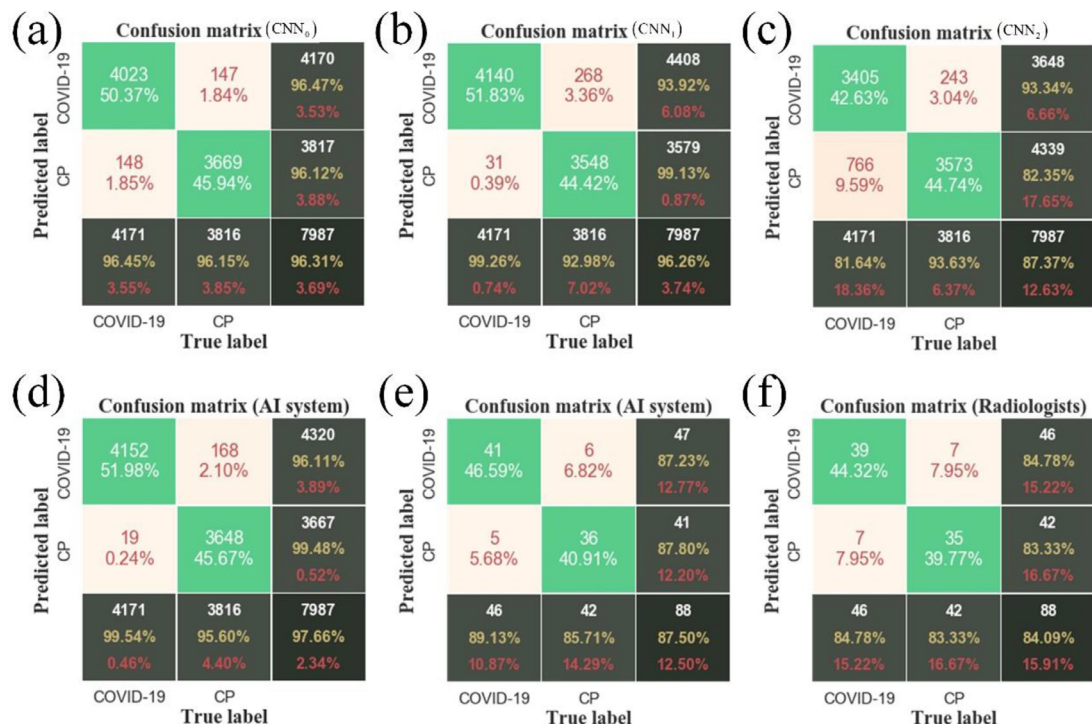

Fig. 4. Confusion matrixes. (a) Slice level analysis by CNN_0; (b) Slice level analysis by CNN_1; (c) Slice level analysis by CNN_2; (d) Slice level analysis by AI system; (e) Scan level analysis by AI system; (f) Scan level analysis by radiologists.

Table 3
Results of scan level analysis of radiologists and AI system.

	AI system	Radiologists	*p*-value
Per scan sensitivity*,%(95% CI)	89.1 (41/46)(75.6–95.9)	84.8 (39/46)(70.5–93.2)	0.727
Per scan specificity**,%(95% CI)	85.7 (36/42)(70.8–94.1)	83.3 (35/42)(68.0–92.5)	1.000
Per scan accuracy***,%(95% CI)	87.5 (77/88)(78.8–93.0)	84.1 (74/88)(74.9–90.4)	0.804
AUC	0.934	n/a	n/a

Note: The *p*-values are calculated by comparing the AI system with radiologists using a 2-sided McNemar test. *Per scan sensitivity is the ratio of true positive identified to all CT scans with COVID-19 lesions; **Per scan specificity is the ratio of true negative identified to all CT scans without COVID-19 lesions; ***Per patient accuracy is the ratio of all true values identified to all CT scans. AI: artificial intelligence; CI: confidence interval; AUC: area under the receiver operating characteristic curve. n/a: not applicable.
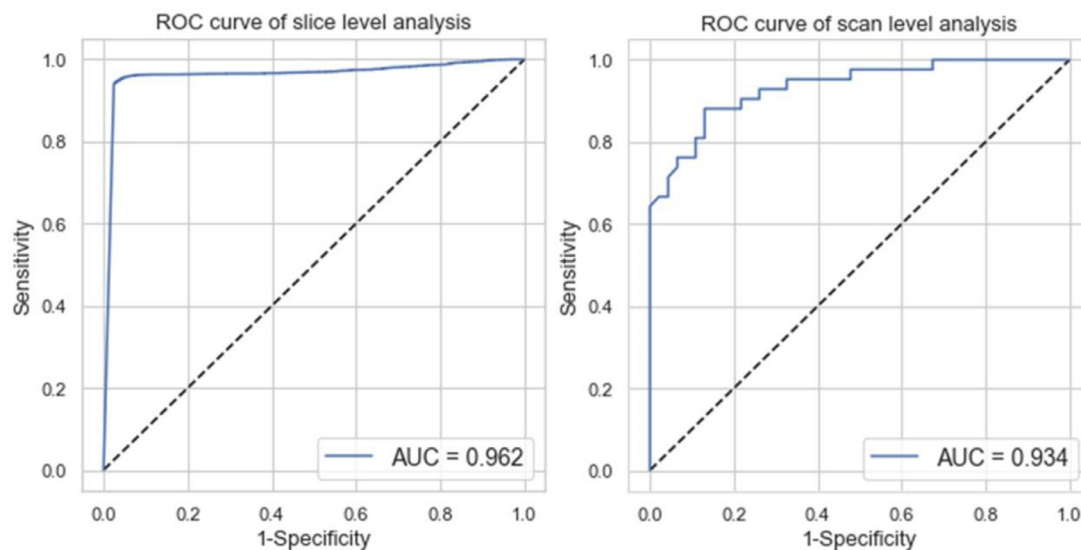
Fig. 5. Receiver operating characteristic curve (ROC) and area under ROC (AUC) obtained by the AI system at slice level analysis and scan level analysis.

non-COVID-19 CT scans are correctly diagnosed, whereas 39 of 46 COVID-19 positive CT scans and 35 of 42 non-COVID-19 CT scans are correctly diagnosed by the experienced radiologists, on average. Table 3 presents that the per scan sensitivities, specificities and accuracies of the scan level analysis are 89.1% (95% CI, 75.6–95.9), 85.7% (95% CI, 70.8–94.1) and 87.5% (95% CI, 78.8–93.0) for the AI system and 84.8% (95% CI, 70.5–93.2), 83.3% (95% CI, 68.0–92.5) and 84.1% (95% CI, 74.9–90.4) for the experienced radiologists, respectively. Fig. 4(e) and (f) shows the confusion matrices of the AI system and experienced radiologists at scan level analysis. The ROC curves of both slice level and scan level analysis are plotted in Fig. 5 and the corresponding AUC values are 0.962 and 0.934 respectively, which are also close to 1, it further indicates an excellent diagnosis system.

5. Discussion

Our pilot study on automatic distinction between COVID-19 and other common pneumonia on chest CT using MSCNN AI system demonstrates a promising result. Compared with single scale input, the MSCNN can achieve better performance.

COVID-19 pneumonia has affected the world with its rapid spread [1]. Many affected patients quickly develop acute respiratory failure with a very poor prognosis and a high mortality rate [2]. Driven by the desire to develop an AI system for the rapid diagnosis of COVID-19 to assist radiologists and clinicians to combat with this pandemic, we utilize two important strategies, MSSP and MSCNN to develop an AI system which does not require a large number of labeled CT data and can achieve satisfactory results. Our AI system achieves a high per slice sensitivity of 99.5% (95% CI, 99.3–99.7) and a high specificity of 95.6% (95% CI, 94.9–

96.2) in 7987 independent test slices. The per scan sensitivities, specificities and accuracies of the AI system against the human experts and their corresponding p-values are (89.1% vs. 84.8%, p-value = 0.724), (85.7% vs. 83.3%, p-value = 1.000) and (87.5% vs. 84.1%, p-value = 0.804), respectively. Even though the three indexes of the AI system are slightly higher than the human experts, but all the *p*-values > 0.05, this means there is no significant difference in the diagnostic performance between the AI system and the experienced radiologists. Although the diagnostic accuracy of the AI system is not significantly larger than experienced radiologists, the AI system can diagnose a CT scan with a time of 0.17 min (10 s per scan), whereas experienced radiologists require an average of 10 min to read a scan. Besides, the AI system can localize the lesions or other key structures in the image when the diagnosis result is given. As shown in Fig. 2, the areas highlighted in the attention maps are the true areas that radiologists will consider the most appropriate to predict COVID-19 and CP. Therefore, the proposed AI system shows great potential to improve the diagnosis time and mitigate the heavy workload of radiologists for differentiating COVID-19 from other common pneumonia.

Nevertheless, our study still has some limitations. Firstly, although our AI system works well on the test dataset of 88 CT scans, it still needs to test on large CT dataset to prove its generalization. Another fundamental limitation arises from the black box, which is the nature of deep networks, although attention maps aid interpretation by highlighting the dominant areas, they are still not sufficient to visualize what unique features are used by the CNN algorithm to distinguish between COVID-19 and CP. Thirdly, this study just cares about the general pattern on chest CT, more detailed clinical information is excluded, further improvement and integration of multidisciplinary approaches are necessary for ex-

tending the application of the AI system. Finally, the training data only uses axial view CT slices, combining the lung regions in the axial direction with coronal and sagittal views for diagnosing more diseases together with their severities is a future work of our research.

6. Conclusion

In sum, this AI system shows a good diagnostic performance for the detection and differentiation of COVID-19 based on a small number of chest CT data. In many developing countries and small-scale hospitals, the number of chest CT scans of COVID-19 is limited, so the number of training samples for building a low-cost intelligent COVID-19 diagnosis system for their own use is always small. In this research, MSSP, MSCNN, and data augmentation are used together to alleviate the scarcity of training data to improve the diagnostic performance of the AI system. In order to defeat COVID-19 and encourage further research in this area, we have shared the dataset at https://data.mendeley.com/datasets/3y55vgckg6/1. It is believed that this system can provide valuable support for radiologists and physicians in performing a fast and accurate diagnosis in the initial screening of COVID-19 and mitigate the heavy workload of them especially when the health system is overloaded.

Author contributions

TY and PKW have contributed to study the concept and design. HR and YL have contributed to the acquisition of data. TY, HR, PKW, HQW, JTW and YL have contributed to analysis and interpretation of data. TY and PKW have contributed to drafting of the manuscript. All authors have approved the manuscript submitted.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

References

[1] Zhu N, Zhang DY, Wang WL, et al. A Novel Coronavirus from patients with Pneumonia in China, 2019. N Engl J Med 2020;382:727–33.

[2] Guan W, Ni Z, Hu Y, et al. Clinical characteristics of Coronavirus disease 2019 in China. N Engl J Med 2020;382:1708–20.

[3] World Health Organization Coronavirus disease (COVID-19) pandemic, Geneva: World Health Organization; 2000. (accessed 25 May 2020). https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[4] Ai T, Yang Z, Hou H, et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology 2020:200642 Feb 26. doi:10.1148/radiol.2020200642.

[5] Xie X, Zhong Z, Zhao W, et al. Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. Radiology 2020:200343 Feb 12. doi:10.1148/radiol.2020200343.

[6] Fang Y, Zhang H, Xie J, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology 2020:200432 Feb 19. doi:10.1148/radiol.2020200432.

[7] Pan Y, Guan H, Zhou S, et al. Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China. Eur Radiol 2020:1–4 Feb 13. doi:10.1007/s00330-020-06731-x.

[8] Bai HX, Hsieh B, Xiong Z, et al. Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. Radiology 2020:200823 Mar 10. doi:10.1148/radiol.2020200823.

[9] Xu B, Xing Y, Peng J, et al. Chest CT for detecting COVID-19: a systematic review and meta-analysis of diagnostic accuracy. Eur Radiol 2020:1 May 15. doi:10.1007/s00330-020-06934-2.

[10] Li X, Fang X, Bian Y, Lu J. Comparison of chest CT findings between COVID-19 pneumonia and other types of viral pneumonia: a two-center retrospective study. Eur Radiol 2020:1–9 May 12. doi:10.1007/s00330-020-06925-3.

[11] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.

[12] Panwar H, Gupta P K, Siddiqui M K, et al. Application of Deep Learning for Fast Detection of COVID-19 in X-Rays using nCOVnet. Chaos Solitons Fractals 2020:109944. doi:10.1016/j.chaos.2020.109944.

[13] Lalmuanawma Samuel, Hussain Jamal, Chhakchhuak Lalrinfela. Applications of Machine Learning and Artificial Intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. Chaos Solitons Fractals 2020:110059. doi:10.1016/j.chaos.2020.110059.

[14] Shi F, Wang J, Shi J, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. IEEE Rev Biomed Eng 2020. doi:10.1109/RBME.2020.2987975.

[15] Zhang K, Liu X, Shen J, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 Pneumonia using computed tomography. Cell 2020 May 4. doi:10.1016/j.cell.2020.04.045.

[16] Li L, Qin L, Xu Z, et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology 2020:200905 Mar 19. doi:10.1148/radiol.2020200905.

[17] Ouyang X, Huo J, Xia L, et al. Dual-sampling attention network for diagnosis of COVID-19 from community acquired Pneumonia. IEEE Trans Med Imaging 2020 May 18. doi:10.1109/TMI.2020.2995508.

[18] Christodoulidis S, Anthimopoulos M, Ebner L, Christe A, Mougiakakou S. Multisource transfer learning with convolutional neural networks for lung pattern analysis. IEEE J Biomed Health Inform 2017;21:76–84.

[19] Wang QC, Zheng YJ, Yang GP, Jin WD, Chen XJ, Yin YL. Multiscale rotation-invariant convolutional neural networks for lung texture classification. IEEE J Biomed Health Inform 2018;22:184–95.

[20] Liu K, Kang G. Multiview convolutional neural networks for lung nodule classification. Int J Imaging Syst Technol 2017;27:12–22.

[21] Yan K, Peng Y, Sandfort V, Bagheri M, Lu Z, Summers RM. Holistic and comprehensive annotation of clinically significant findings on diverse CT images: learning from radiology reports and label ontology. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2019. p. 8523–32.

[22] Haarburger C, Baumgartner M, Truhn D, Broeckmann M, Schneider H, Schrading S, Kuhl C, Merhof D. Multi scale curriculum CNN for context-aware breast MRI malignancy classification. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham; 2019. p. 495–503. Oct 13. doi:10.1007/978-3-030-32251-9_54.

[23] Hashimoto N, Fukushima D, Koga R, et al. Multi-scale domain-adversarial multiple-instance CNN for cancer subtype classification with unannotated histopathological images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020. p. 3852–61.

[24] Rasti R, Rabbani H, Mehridehnavi A, Hajizadeh F. Macular OCT classification using a multi-scale convolutional neural network ensemble. IEEE Trans Med Imaging 2018;37:1024–34.

[25] Kim BC, Yoon JS, Choi JS, Suk HI. Multi-scale gradual integration CNN for false positive reduction in pulmonary nodule detection. Neural Netw 2019;115:1–10.

[26] Liu XL, Hou F, Qin H, Hao AM. Multi-view multi-scale CNNs for lung nodule type classification from CT images. Pattern Recognit 2018;77:262–75.

[27] Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-scale convolutional neural networks for lung nodule classification. In: International conference on information processing in medical imaging. Springer, Cham; 2015. p. 588–99. Jun 28. doi:10.1007/978-3-319-19992-4_46.

[28] Burt PJ, Adelson EH. The Laplacian pyramid as a compact image code. IEEE Trans Commun 1983;31:532–40.

[29] Farabet C, Couprie C, Najman L, LeCun Y. Learning hierarchical features for scene labeling. IEEE Trans Pattern Anal Mach Intell 2013;35:1915–29.

[30] Tan M, Le QV. Efficientnet: rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946. 2019 May 28.

[31] Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621. 2017 Dec 13.

[32] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis 2020;128:336–59.

[33] Jin C, Chen W, Cao Y, et al. Development and evaluation of an AI system for COVID-19 diagnosis. medRxiv. 2020 Jan 1. doi:10.1101/2020.03.20.20039834.