LiMS-Net: A Lightweight Multi-Scale CNN for COVID-19 Detection from Chest CT Scans

AMOGH MANOJ JOSHI, Vivekanand Education Society's Institute of Technology, Mumbai, India
DEEPAK RANJAN NAYAK, Malaviya National Institute of Technology, Jaipur, India
DIBYASUNDAR DAS, National Institute of Technology, Rourkela, India
YUDONG ZHANG, University of Leicester, UK

Recent years have witnessed a rise in employing deep learning methods, especially **convolutional neural networks (CNNs)** for detection of COVID-19 cases using chest CT scans. Most of the state-of-the-art models demand a huge amount of parameters which often suffer from overfitting in the presence of limited training samples such as chest CT data and thereby, reducing the detection performance. To handle these issues, in this paper, a **lightweight multi-scale CNN** called **LiMS-Net** is proposed. The LiMS-Net contains two feature learning blocks where, in each block, filters of different sizes are applied in parallel to derive multi-scale features from the suspicious regions and an additional filter is subsequently employed to capture discriminant features. The model has only 2.53M parameters and therefore, requires low computational cost and memory space when compared to pretrained CNN architectures. Comprehensive experiments are carried out using a publicly available COVID-19 CT dataset and the results demonstrate that the proposed model achieves higher performance than many pretrained CNN models and state-of-the-art methods even in the presence of limited CT data. Our model achieves an accuracy of 92.11% and an F1-score of 92.59% for detection of COVID-19 from CT scans. Further, the results on a relatively larger CT dataset indicate the effectiveness of the proposed model.

CCS Concepts: • **Computing methodologies** → **Neural networks**; **Object recognition;**

Additional Key Words and Phrases: Deep learning, lightweight CNN, COVID-19, chest CT scan, LiMS-Net

ACM Reference format:
Amogh Manoj Joshi, Deepak Ranjan Nayak, Dibyasundar Das, and Yudong Zhang. 2023. LiMS-Net: A Lightweight Multi-Scale CNN for COVID-19 Detection from Chest CT Scans. *ACM Trans. Manage. Inf. Syst.* 14, 1, Article 5 (January 2023), 17 pages.
https://doi.org/10.1145/3551647

All authors contributed equally to this research.
This work is partially supported by the Science and Engineering Research Board (SERB), Department of Science and Technology, Govt. of India under project No. SRG/2020/001460.
Authors' addresses: A. M. Joshi, Vivekanand Education Society's Institute of Technology, Chembur, Mumbai, Maharashtra, 400 074, India; email: joshiamogh9@gmail.com; D. R. Nayak, Malaviya National Institute of Technology, Jaipur, JLN Marg, Rajasthan, 302 017, India; email: drnayak@ieee.org; D. Das, National Institute of Technology, Rourkela, Odisha, 769 008, India; email: dibyasundar@ieee.org; Y. Zhang, University of Leicester, Leicester, LE1 7RH, UK; email: yudongzhang@ieee.org.

2158-656X/2023/01-ART5 $15.00
https://doi.org/10.1145/3551647

1 INTRODUCTION

The Coronavirus disease 2019 (COVID-19) was declared a pandemic by the World Health Organization on 11 March 2020 following its rapid spread since its outbreak began in December 2019 [24]. This virus has infected around 152 million people and caused 3.19 million deaths all across the globe (including US 590k, India 211k, Brazil 404k, France 104k, etc.). Some recent outbreaks of the variants of COVID-19 virus indicate that it has a highly mutating nature. The major symptoms of this virus are fever, cough, loss of taste and smell, fatigue and muscle aches. But recently, many cases were reported where the patients, without even experiencing any of these symptoms, were tested positive due to this novel virus. A rise in such asymptomatic cases has hampered the accurate and timely diagnosis of this virus. It is still an ongoing pandemic and a leading cause of death in many countries. Hence, a quick and accurate diagnosis is imperative to control the rapid spread of this disease. The current standard testing procedure is based on **reverse transcription polymerase chain reaction (RT-PCR)** which takes around 4–6 hours to provide results and hence, is quite slow and inefficient as compared to the quick spreading of the virus [5, 6]. Besides, the scarcity of RT-PCR test kits also proves to be a major concern for timely detection of the virus and curbing its spread. Some key findings in recent studies [5, 15] reveal that **computer tomography (CT)** scans can be strongly considered as an efficient and alternative testing method because of their capability in showing clear radiological findings of COVID-19 patients at a fast speed and easy accessibility [3, 12, 33]. Also, CT scans distinctly manifest a blend of multifocal peripheral lung changes of **ground-glass opacity (GGO)** and consolidation which effectively highlight COVID-19 infections in lungs [30, 33]. However, the medical experts and radiologists require time in examining and analyzing the CT scans manually and may suffer from fatigue due to the burden of examining patients at a large scale. Thus, an automated system which can effectively analyze CT scans and classify them quickly as COVID-19 positive or not is greatly needed to tackle the current situation.

In recent years, there has been an increased effort on employing **deep learning (DL)** methods, mainly **convolutional neural networks (CNNs),** for predicting COVID-19 infection from chest radiograph images such as X-rays and CT images [24, 30]. CNNs have proved to be effective in extracting salient features from chest radiography images. Despite the promising performance of CNNs in some studies, a major issue to be noted is the large quantity of training data that they require to effectively extract the prominent features from the CT images. But the lack of publicly available datasets containing thousands of CT scans makes it harder for the CNN model to learn the accurate features that are required to correctly classify positive COVID-19 samples from negative samples. Also, most of the existing CNN models require a large number of parameters and memory space which often cause overfitting on such a small training data and thereby, reducing the prediction performance and increasing the inference time while classifying a CT image. Therefore, such heavier models may not be suitable for real-time diagnosis especially in this scenario where obtaining rapid results is highly essential.

To address the above issues, in this paper, we present a lightweight multi-scale CNN which tackles the problem of extracting prominent features even in the presence of limited CT data. Despite having smaller parameters, low memory space and shallow architecture, the proposed CNN model outperforms most of the heavy pretrained CNN models. The important contributions of this paper can be summarized as:

- We propose a lightweight CNN with a multi-scale architecture for COVID-19 detection which effectively extracts discriminable features from chest CT images even when training data is limited.

- Unlike other sequential state-of-the-art CNN models, we introduce two feature learning blocks in the proposed model which aims to extract multi-scale features from suspicious regions of CT scans by applying filters of different sizes in parallel and thereafter, fuses them to achieve higher performance. An additional filter is then applied in each block to further enhance the feature learning capability.
- To test the efficacy of the LiMS-Net, an extensive set of experiments is carried out using a publicly available COVID-19 CT dataset[1] which carries 349 COVID-19 and 397 non-COVID samples. Several ablation studies are performed to find the best hyperparameter values for effective COVID-19 diagnosis. Further, we verify the generalization ability of the model using a comparatively larger CT dataset.
- We compare the effectiveness of the proposed LiMS-Net with ImageNet pretrained CNN models in terms of classification performance, computational cost and memory. Also, we validate the state-of-the-art COVID-19 detection methods using our datasets.

The paper is organized as follows: Section 2 reviews the recent works on DL based COVID-19 diagnosis and Section 3 presents the dataset used. Section 4 describes the proposed LiMS-Net in detail, while Section 5 presents the experimental results, comparisons and the ablation studies. Finally, we conclude the paper in Section 6.

2 RELATED WORK

Soon after the outbreak of COVID-19, an ample amount of studies were performed to automatically interpret chest radiograph images such as X-rays and CT scans to correctly identify COVID-19 infection [5, 24]. Many studies [16, 18] explained the importance and role of DL methods (mostly CNNs) in screening and diagnosing COVID-19 patients. Xu et al. [31] developed a DL based early diagnostic system for distinguishing COVID-19 from pneumonia and normal cases using CT images which obtained an accuracy of 86.7%. Wang et al. [29] designed a DL algorithm by modifying Inception V3 model to achieve an accuracy of 79.3%. Zhang et al. [32] proposed a 7-layer CNN model integrated with stochastic pooling strategy for COVID-19 diagnosis from chest CT scans. He et al. [9] proposed a self supervised transfer learning DL approach for COVID-19 detection from CT scans and achieved an accuracy of 86%. Hasan et al. [7] proposed a multi encoder ensemble network named CVR-Net using ResNet50 and Xception as encoders for COVID-19 detection from CT scans. Saqib et al. [22] implemented various state-of-the-art CNN models like ResNet, EfficientNet, MNasNet to determine the efficacy of CNN models in detecting COVID-19 from CT scans. Wang et al. [30] developed a CT scan based deep learning model called CCSHNet to screen COVID-19 which used two best pretrained CNN models to learn salient features and then fused these features by a discriminant correlation analysis method. Recently, Kaur et al. [13] proposed an automatic method using deep features derived from a MobileNetv2 and a **parameter free BAT optimized fuzzy K-nearest neighbor classifier (PF-FKNN)** to achieve higher accuracy for COVID-19 diagnosis.

All these studies either used heavy CNN models (large model size) or pretrained models with transfer learning which are computationally expensive and need large memory space. Further, many models were not focused on learning multi-scale features which are essential for effective COVID-19 diagnosis. Another issue that many studies faced is of validating models using limited training data which greatly influences the model's performance. To handle the problems faced by heavy models, Polsinelli et al. [20] built a lightweight CNN model based on SqueezeNet using CT images which achieved a classification accuracy of 85.03% for COVID-19 diagnosis. However,

[1] https://github.com/UCSD-AI4H/COVID-CT.

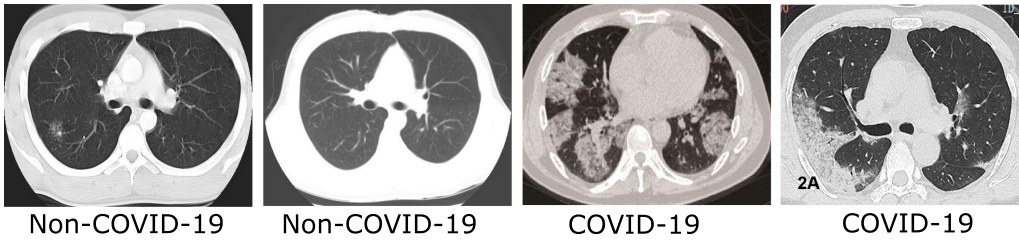

Fig. 1. Sample CT scans of COVID-19 and Non-COVID-19 from COVID-19 CT dataset.

Table 1. Description of COVID-19 CT Dataset

Set	COVID-19	Non-COVID-19
Train	191	234
Validation	60	58
Test	98	105
Total Images	349	397

the accuracy is not reliable for real-time COVID-19 diagnosis and the model doesn't effectively help the medical diagnostic process. Apart from these, several studies were also conducted for automated COVID-19 diagnosis using chest X-ray images [17, 19]. Thus, a lightweight CNN which could effectively learn prominent features from the limited training data is highly in demand. In this study, we aim to design a CNN model which learns multi-scale features from the limited chest CT images while retaining the small model size.

3 DATASET

To validate the proposed LiMS-Net model, we consider a large COVID-19 CT dataset1 available to date which is reported in He et al. [9]. The dataset contains 349 CT images of COVID-19 positive cases and 397 CT scans of COVID-19 negative cases. The positive COVID-19 samples were collected from 143 patients, whereas the negative COVID samples were collectively obtained from various online resources such as **PubMed Central (PMC)** and MedPix. The images were varied in height and width from 153 to 1853 and 124 to 1485, respectively and were hence resized to 224×224. The sample CT images of patients with positive COVID-19 and negative COVID-19 are shown in Figure 1. This dataset was already available in train, validation, and test sets. Table 1 below summarizes the number of CT images in each set.

3.1 Data Augmentation

Medical imaging datasets face a common problem of small-size dataset due to a shortage of sufficient data samples. A lesser number of training samples often causes ineffective feature extraction using CNN models, and thus results in poor performance [25]. Although the dataset considered is comparatively larger in this domain, it is still not adequate to effectively train a CNN model. Augmenting the data is an effective way to tackle this problem. Hence, each image in the training set was augmented with four types of transformations: (a) Gaussian noise (with mean and standard deviation as 0 and 0.01, respectively), (b) horizontal flipping, (c) anticlockwise rotation (angle 5°), and (d) clockwise rotation (angle 5°).

To explain better, assume $|S|$ to be the cardinality of the a set S of images, i.e., set S contains $|S|$ images. Since we perform augmentation operations on the original train set T, it contains $|T|$

Table 2. Description of Augmented Dataset

Set	COVID-19	Non-COVID-19
Train (augmentation)	955	1170
Validation	60	58
Test	98	105
Total Images	1073	1333

images such that

$$I_t \in T,\ t = 1, 2, \ldots, |T| \tag{1}$$

Now, we can define the above augmentation operations mathematically as follows:

3.1.1 Gaussian Noise. Gaussian noise was injected over the training images to generate new noisy images. Let GN be the Gaussian noise function such that

$$I_T^{gn} = GN(I_t, \mu^{GN}, \sigma^{GN}),\ t = 1, 2, \ldots, |T| \tag{2}$$

where, I_T^{gn} represents the set of augmented noisy images, μ^{GN} denotes the mean, and σ^{GN} denotes the standard deviation. The value of μ^{GN} and σ^{GN} were set as 0 and 0.01, respectively.

3.1.2 Horizontal Flip. Each training image I_t was flipped horizontally to obtain the new image. Let $FLIP$ denote the flip function which requires a parameter flipcode fc and is defined as follows.

$$I_T^{fl} = FLIP(I_t, fc),\ t = 1, 2, \ldots, |T| \tag{3}$$

where, I_T^{fl} represents the set of horizontally flipped images. For horizontal flip, the value of $fc = 1$ and for vertical flip, $fc = 0$.

3.1.3 Counterclockwise Rotation. Image rotation is one of the most widely used augmentation techniques. Each training image I_t was rotated counterclockwise to obtain the new image. Let $ROTA$ be the function which requires the angle of rotation θ^{POS} and can be defined as

$$I_T^{ra} = ROTA(I_t, \theta^{POS}),\ t = 1, 2, \ldots, |T| \tag{4}$$

where, I_T^{ra} represents the set of counterclockwise rotated images. We set the value of θ^{POS} as 5°.

3.1.4 Clockwise Rotation. Similar to counterclockwise rotation, we also rotated each training image I_t clockwise with an angle $\theta^{NEG} = 5°$ to obtain the new image which can be expressed as

$$I_T^{rc} = ROTC(I_t, \theta^{NEG}),\ t = 1, 2, \ldots, |T| \tag{5}$$

where, I_T^{rc} represents the set of clockwise rotated images.

After performing the above augmentation operations, every image I_t produced four new augmented images such that

$$I_t \Rightarrow \left\{I_t^{gn}, I_t^{fl}, I_t^{ra}, I_t^{rc}\right\},\ t = 1, 2, \ldots, |T| \tag{6}$$

Thus, the train set size is increased by 5 times (including the original image). The final dataset arrangement after augmentation is summarized in Table 2.

4 PROPOSED METHODOLOGY

The objective behind proposing a lightweight multi-scale CNN model was to enable efficient feature extraction from CT images which results in high performance while preserving the small model size. In this section, we brief upon the importance of having multi-scale feature learning and describe the proposed LiMS-Net model.

4.1 Importance of Multi-Scale Feature Learning

ImageNet pretrained models like VGGNet [26], ResNet [8], and the like, and a majority of the CNN models proposed in the above mentioned studies have single branched convolution layers connected in a linear fashion, stacked one below the other. Such networks mostly focus on extracting homogeneously scaled features and fall short in extracting multi-scale features which have been proved to be effective for complex image classification tasks [14]. Generally, the most vital features are missed out and these models could not extract exact features which leads to a poor learning capability and generalization performance. Medical images like CT scans don't contain significant features with respect to some specific locations; rather, they can be found in any location over them. Further, these single branched models when trained on such medical datasets have a high chance of learning unimportant features which affects their diagnosis performance. Some recent works have proven the effectiveness of having multi-scaled convolutional filters in the CNN architecture [14, 27]. Multi-scaled filters enable the model to extract and learn discriminant features which improve the performance of the model to a large extent. Hence, we incorporate multi-scaled filters in the proposed LiMS-Net model which facilitates efficient feature learning and improved performance.

4.2 Importance of Lightweight CNN

Heavy CNN models are generally known for their good classification performance; however, they are computationally expensive and demand large memory. They also require more computational resources and take larger inference time. In general, heavy models usually have a larger number of parameters which often face an issue of overfitting when trained on small training data like medical data. Further, such heavy models may not be feasible as far as real-time deployment is concerned. Therefore, to handle the above issues, we aim at developing a lightweight CNN model which learns discriminant features from the limited chest CT dataset while retaining the model size.

4.3 Proposed CNN Model

The proposed model is a shallow network which mainly incorporates two multi-scale feature extraction blocks, a **convolution (CONV)** layer, **batch normalization (BN)** layers, **max-pooling (MP)** layers, and **fully connected (FC)** layers as shown in Figure 2. The input image is first passed through a CONV layer followed by a BN and MP layer to initiate the feature learning process and reduce the feature map size. The model then includes two feature learning blocks which facilitate multi-scale feature extraction using different scaled convolution filters. These blocks incorporate a two level architecture. In the first level, two different CONV filters of size 3×3 and 5×5 are applied simultaneously to capture distinctive features at different scales from various regions of CT images. The 3×3 filter helps in extracting the detailed features like patchy shadows, small patches, and GGO which are found in the CT scans of COVID-19 positive patients, whereas the 5×5 filter helps to learn the coarse features like shape of the lung region. These features are then fused using concatenation and then a supplementary CONV filter of size 3×3 is employed in the second level to further capture the enriched features and thereby, enhancing the feature learning capability.

In the parallel CONV layers, padding is kept the same to ensure the same output dimension from the parallel stream which is required for concatenation. It is worth mentioning here that each CONV layer in the proposed model is followed by a BN and MP layer. The BN layer is included in order to improve the training convergence and reduce overfitting [20]. The same block is repeated once again. The functioning of both the blocks is same and can be expressed mathematically as follows.

Let $B^j(I)$ be the output of block B^j, where j is the block number and I is the input to the block. Since the block contains two levels, let ${B_{int}}^j(I)$ represents the intermediate output of the first level

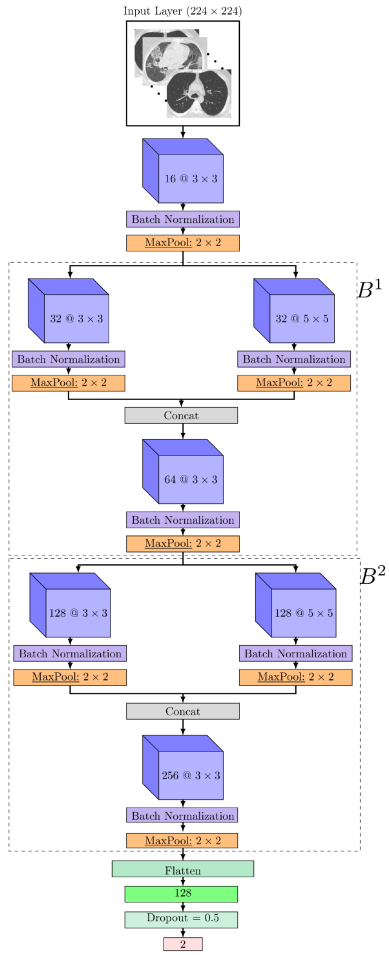

Fig. 2. Overall structure of proposed LiMS-Net model.

which is then passed on to the second level. Let $\psi_p^{f,f,N}(I)$ represents convolution operation with N number of kernels of size $f \times f$ and padding p, δ represents the BN function, φ represents max-pooling operation and $\oplus$ indicates the concatenation operation. Since our model contains only two blocks, $j \in [1, 2]$. The response of each block can be defined as follows.

$$B_{int}{}^{j}(I) = \left\{\varphi\left(\delta\left(\psi_{p=1}^{3,3,2^{2j+3}}(I)\right)\right) \oplus \varphi\left(\delta\left(\psi_{p=2}^{5,5,2^{2j+3}}(I)\right)\right)\right\} \tag{7}$$

$$B^{j}(I) = \varphi\left(\delta\left(\psi_{p=1}^{3,3,2^{2j+4}}(B_{int}{}^{j}(I))\right)\right) \tag{8}$$

The BN layer used in each block normalizes the inputs of each layer to solve the problem of internal covariate shift. It calculates the mean μ_M and variance σ_M^2 across the mini batch size and normalizes the layer inputs using the previously calculated batch statistics. Let M be the mini-batch having size m such that $M = \{I_1, I_2, \ldots, I_m\}$, then the BN of M is calculated as follows.

$$\mu_M = \frac{1}{m}\sum_{i=1}^{m} I_i \tag{9}$$

$$\sigma_M^2 = \frac{1}{m}\sum_{i=1}^{m}(I_i - \mu_M)^2 \tag{10}$$

$$\bar{I}_i = \frac{I_i - \mu_M}{\sqrt{\sigma_M^2 + \epsilon}} \tag{11}$$

where, $\bar{I}_i$ is the normalized input with zero mean and unit variance and ϵ is a small arbitrary constant added in the denominator for numerical stability. Finally, this normalized input is scaled and shifted using two learnable parameters γ and β to obtain the output of the BN layer. Thus, the batch normalization process can be summarized as follows

$$O_i^{BN} = \gamma I_i + \beta \tag{12}$$

where, O_i^{BN} is the BN output.

The final block is followed by an FC layer of 128 neurons and a dropout layer with a value of 0.5. Following that, a final classification layer of two neurons is introduced with softmax activation to classify the CT images as COVID-19 or Non-COVID-19.

Unlike other state-of-the-art CNN models, the proposed LiMS-Net model consists of novel multi-scale feature learning blocks that allows the model to learn the salient features including the minute and coarse ones, and thereby facilitating effective classification of COVID-19 positive samples from negative samples. The model is lightweight and hence demands less computational cost and memory. Further, it can capture prominent discriminable features even when the training data is limited.

5 EXPERIMENTS AND RESULTS

In this section, we present the implementation details, performance metrics, and experimental results of the LiMS-Net model. To verify effectiveness of LiMS-Net, the results were compared with ImageNet based pretrained models as well as the existing methods. Further, an extensive set of experiments were performed as a part of ablation studies to analyze the effect of each important factor such as filter size, activation function, weight initialization paradigm, pooling strategy, and number of blocks.

5.1 Implementation Details and Performance Metrics

The proposed LiMS-Net was validated on a COVID-19 CT image dataset with 349 COVID-19 positive and 397 COVID-19 negative cases which takes an input resolution of 224×224. To reduce overfitting and provide good generalization, we applied transformations such Gaussian noise, horizontal flip, and both clockwise and anticlockwise rotations as a part of data augmentation. There are two neurons in the classification and and the loss function considered was categorical cross-entropy. The model was trained for 50 epochs with Adam optimizer. The mini-batch size and the learning rate was set to 16 and 0.00001, respectively. The random initialization method was chosen to initialize the weights. All models were implemented using Keras with Tensorflow as backend. To further verify the effectiveness of the model, a comparatively larger CT dataset was taken into consideration.

To evaluate LiMS-Net as well as other existing models, we used several different evaluation metrics: accuracy (Acc), F1-score (F1), specificity (Spec), sensitivity (Sens), and AUC.

5.2 Evaluation of Proposed Model on COVID-19 CT Dataset

The training curves obtained by our LiMS-Net model are shown in Figure 3. From the graph, it can be seen that the model is converged well within 50 epochs. The model achieved a training accuracy of 100%, validation accuracy of 94.07%, and test accuracy of 92.11%. Apart from accuracy, we also

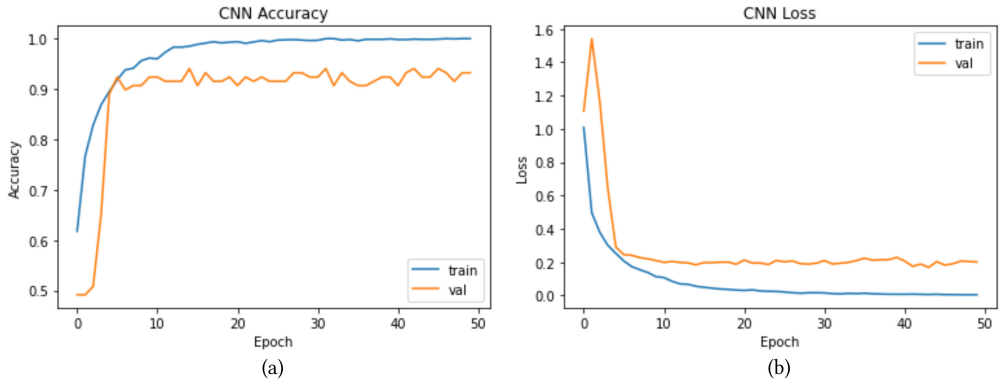

Fig. 3. Training curves on COVID-19 CT dataset: (a) accuracy vs. epoch, and (b) loss vs. epoch.

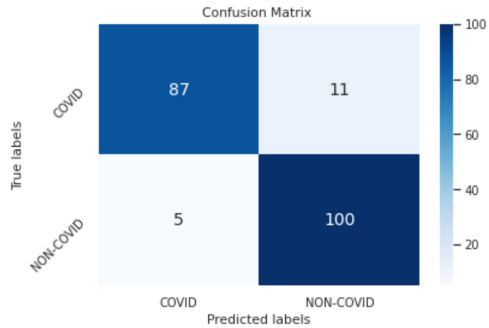

Fig. 4. Confusion matrix obtained by LiMS-Net on test set.

Table 3. COVID-19 Classification Results (in %) of our LiMS-Net on COVID-19 CT Dataset

Metric	Value
Accuracy	92.11
F1-score	92.59
AUC Score	92.00
Specificity	95.23
Sensitivity	88.77
Precision	94.56

evaluated its performance using various metrics as shown in Table 3. Our model achieved an F1 score of 92.59%, AUC score of 92.00%, specificity of 95.23%, sensitivity of 88.77%, and precision of 94.56%. The test set contained 203 images of which 98 were COVID-19 positive samples and the rest 105 were COVID-19 negative samples. Figure 4 shows the confusion matrix obtained by our LiMS-Net on the test set. It can be observed that out of 105 samples only 5 were wrongly classified as COVID-19 positive and only 11 out of 98 images were misclassified as COVID-19 negative.

To verify the stability of the proposed model, it has been repeatedly evaluated for 20 runs, and the results are plotted in Figure 5. It can be observed that the accuracy does not vary to a large extent and instead is similar in most runs, indicating the stability of the proposed model.

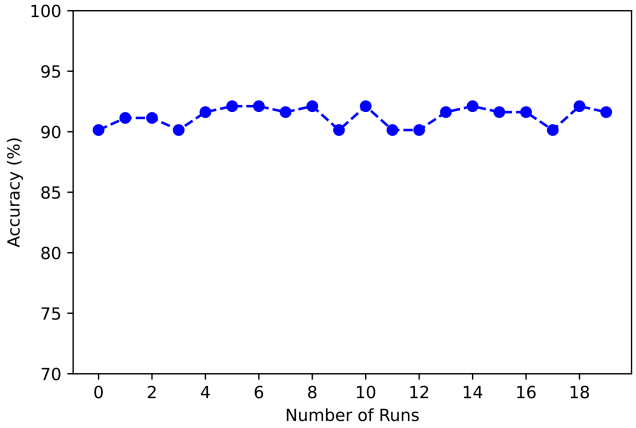

Fig. 5. Accuracy obtained by LiMS-Net in different runs.

5.2.1 Comparison with ImageNet Pretrained CNN Models on COVID-19 CT Dataset. To demonstrate the robustness and efficacy of our LiMS-Net, we investigated the performance of ImageNet pretrained CNN architectures by finetuning them on the considered COVID-19 CT dataset and compared their results with that of our model. The pretrained models included VGG-16 [26], ResNet-50 [8], ResNet-101 [8], DenseNet-121 [11], DenseNet-169 [11], Inception-V3 [28], Xception [2], and MobileNet [10]. First the comparison was made on model size in the context of model parameters and memory space required as listed in Table 4. It can be seen that the ImageNet pretrained models are extremely heavy like VGG-16 with 134.26M, ResNet-101 with 42.66M, and ResNet-50 with 23.5M parameters as compared to our model which requires just 2.53M parameters (about 1.88% of VGG-16, 5.93% of ResNet-101, and 10.77% of ResNet-50). Also, it demands less parameter compared to the most widely used lightweight model called MobileNet. Further, it needs a small amount of memory space and less training time (in sec) as compared to others and thus, is more suitable for real-time COVID-19 diagnosis using CT images.

Table 5 shows the classification results of the compared ImageNet pretrained models on COVID-19 CT dataset. It can be also noticed that our model achieved comparable or better performance than many models. Although DenseNet-169 and ResNet-101 achieved a slightly better classification performance, they required comparatively a very large number of model parameters to optimize than our LiMS-Net, i.e., DenseNet-169 required nearly about 5 times more and ResNet-101 required 17 times more parameters. Despite being extremely lightweight, our LiMS-Net outperforms many pretrained heavy CNN architectures like ResNet-50, DenseNet-121, Inception-V3, Xception, and VGG-16 in terms of classification performance which further proved the efficacy of LiMS-Net. Specifically, compared to DenseNet-121 and Inception-V3 which learns multi-scale features, our LiMS-Net gets higher performance with less model parameters and memory space. It is worth mentioning here that the ImageNet models were implemented under similar experimental set up.

5.2.2 Comparison with Existing Methods on COVID-19 CT Dataset. We further compared the performance of LiMS-Net with the recent existing approaches for automated COVID-19 diagnosis and the results are tabulated in Table 6. It is evident that our LiMS-Net yielded higher classification accuracy than state-of-the-art COVID-19 diagnosis methods [7, 9, 13, 17, 20, 22, 29, 32]. It is worth mentioning here that all the existing methods were validated using the same dataset (COVID-19 CT dataset) and were implemented under similar experimental setting to derive a fair comparison. It can also be seen that LiMS-Net outperformed other multi-scale based CNN methods proposed by

Table 4. Comparison with ImageNet CNN Models in Terms of Model Parameters, Size, and Training Cost

Model	Depth	Parameter (M)	Size (MB)	Train time (Sec)
VGG-16	23	134.26	512.27	7.59×10^2
ResNet-101	101	42.66	163.68	7.52×10^2
ResNet-50	50	23.5	90.48	7.3×10^2
Inception-V3	159	21.8	84	7.46×10^2
Xception	126	20.86	79.97	7.41×10^2
DenseNet-169	169	12.64	49.76	7.24×10^2
DenseNet-121	121	7.03	27.94	7.19×10^2
MobileNet	55	3.23	13	6.31×10^2
Our LiMS-Net	**7**	**2.53**	**9.7**	5.13×10^2

Table 5. Comparison of Classification Results (in %) with ImageNet Pretrained CNN Models on COVID-19 CT Dataset

Model	Acc	F1	Sens	Spec	AUC
DenseNet-169	93.10	93.39	91.84	94.29	93.06
ResNet-101	92.61	93.02	89.80	95.24	92.61
ResNet-50	92.11	92.59	88.77	95.23	92.00
DenseNet-121	91.62	92.01	89.80	93.33	91.54
Inception-V3	91.13	91.50	89.80	92.38	91.08
Xception	91.13	91.66	87.76	94.29	91.02
MobileNet	91.13	91.66	87.76	94.29	91.02
VGG-16	90.14	90.82	85.71	94.29	90.00
Our model	**92.11**	**92.59**	**88.77**	**95.23**	**92.00**

Wang et al. [29], He et al. [9], and Hasan et al. [7]. The CRNet [9] obtained the least performance and this is due to the fact that CRNet followed VGG like architecture and did not extract multi-scale features from the CT images.

5.3 Ablation Studies

To gain deeper insights about the effectiveness of different vital elements on our LiMS-Net, we performed various ablation studies. The vital elements include the choice of filter size, activation function, weight initialization scheme, pooling operation, number of blocks, and effect of augmentation. It is worth noting here that all the ablation studies were preformed on COVID-19 CT dataset.

5.3.1 Choice of Activation Function. In this experiment, we analyzed the impact of activation function on our LiMS-Net. The performance of LiMS-Net with different activation functions such as ReLU, leaky ReLU, and Swish [21] are reported in Table 7. Generally, leaky ReLU is an improvised version of ReLU which solves the dying gradient problem of ReLU and gives better performance. Swish is also known to achieve better performance than ReLU in many tasks. But surprisingly, for our LiMS-Net, ReLU activation provided higher classification results as demonstrated in Table 7.

5.3.2 Choice of Weight Initialization. The results obtained so far were using random weight initialization. After selecting a proper filter size and activation function, we verified the effect of different weigh initialization schemes such as random, Kaiming and Xavier initialization to

Table 6. Performance Comparison with Existing Works on COVID-19 CT Dataset

Reference	Method	Acc (%)	F1 (%)
Wang et al. [29]	M-Inception	83.74	83.74
He et al. [9]	DenseNet-169 with Self-Trans	86.00	85.00
He et al. [9]	CRNet	72.00	76.00
Zhang et al. [32]	7L-CNN	86.69	87.20
Saqib et al. [22]	EfficientNet-B4	80.30	81.13
Saqib et al. [22]	DenseNet169	87.10	38.46
Nayak et al. [17]	ResNet-34	91.13	91.58
Hasan et al. [7]	CVR-Net	78.00	78.00
Polsinelli et al. [20]	SqueezeNet	86.20	86.91
Kaur et al. [13]	MobileNetv2 and PF-FKNN	84.24	83.16
Our model	LiMS-Net	**92.11**	**92.59**

Table 7. Accuracy of LiMS-Net with Different Activation Functions

Epochs	Activation function		
	ReLU	Leaky ReLU	Swish
25	90.14	87.19	88.17
50	92.11	88.66	89.16

Table 8. Accuracy of LiMS-Net with Different Weight Initializations

Epochs	Weight initialization		
	Random	Kaiming	Xavier
25	90.14	89.16	90.14
50	92.11	90.14	86.69

initialize the weights of LiMS-Net and the results are reported in Table 8. From the table, it can be seen that the performance of our LiMS-Net was improved with random weight initialization, i.e., improved by 1.97% and 5.42% when compared with Kaiming and Xavier initialization, respectively.

5.3.3 Choice of Pooling Strategies. In this experiment, the effect of different pooling operations was investigated on our model. The two major types of pooling are max pooling and average pooling. Instead of using a pooling layer, another way of reducing the dimension of feature maps is by using CONV layer with a stride of 2 in which the filter map moves in steps of 2, thus exactly reducing the dimension of feature map by half. The performance of each of these operations was validated with our model and the results are summarized in Table 9. It can be seen that there is no improvement in the performance after 25 epochs in case of average pooling and CONV with stride 2. However, max pooling effectively preserved the dominant features and resulted in an improved classification performance for COVID-19 diagnosis.

5.3.4 Choice of Filter Size. In any CNN architecture, selection of proper CONV filter sizes plays an important role in extracting relevant features.The 3×3 filters were used in two blocks of our LiMS-Net with an aim to extract local contextual information and hence, were kept fixed

Table 9. Accuracy of LiMS-Net with Different Pooling Strategies

Epochs	Downsampling operation		
	Max Pool	Avg Pool	CONV with stride 2
25	90.14	89.65	86.69
50	92.11	89.65	86.69

Table 10. Accuracy of LiMS-Net with Different Filter Sizes

	Combination of Filters					
	(3,5)	(3,7)	(3,9)	(5,7)	(5,9)	(7,9)
Accuracy	92.11	91.13	88.17	90.14	89.65	90.14
Parameters (M)	2.53	2.74	3.02	2.88	3.16	3.36

(x, y) denotes the combination of filters with size $x \times x$ and $y \times y$.

Table 11. Performance of LiMS-Net (in %) with Different Number of Blocks

Number of Blocks	Acc	F1	AUC
2	92.11	92.59	92.00
3	85.71	85.99	85.74
4	86.20	87.03	86.08

Table 12. Performance of LiMS-Net (in %) With and Without Augmented Trained Data

Mode	Acc	F1	AUC
Without augmentation	71.42	76.03	70.85
With augmentation	92.11	92.59	92.00

throughout the layers. To extract multi-scale features, we employed a 5×5 CONV filter along with a 3×3 filter simultaneously. We performed a comprehensive analysis on the combination of filter sizes to be adopted for achieving the best features, thus resulting in better performance. The results are shown in Table 10. From the table, it can be seen that the best results were obtained using a combination of 3×3 and 5×5 filters which effectively capture minute and coarse features. Moreover, increasing the filter size also increases the parameters which might affect the model's performance.

5.3.5 Choice of Number of Blocks. To further study and verify the architecture of the proposed LiMS-Net, we performed a study of the effect of the number of blocks on the model's performance as shown in Table 11. It can be seen that the performance was decreased as the number of blocks increased which can be attributed to learning of irrelevant features in the later blocks of the model. Thus, two blocks sufficiently extract the prominent features which results in a superior performance.

5.3.6 Effect of Augmentation. To study and verify the effect of data augmentation, we analyzed the model's performance on both original and augmented trained COVID-CT dataset. Table 12 shows the results of our LiMS-Net with data augmentation and without data augmentation. It can be clearly seen that data augmentation significantly increases the accuracy by 21% due to a considerable increase (5 times) in the number of training samples.

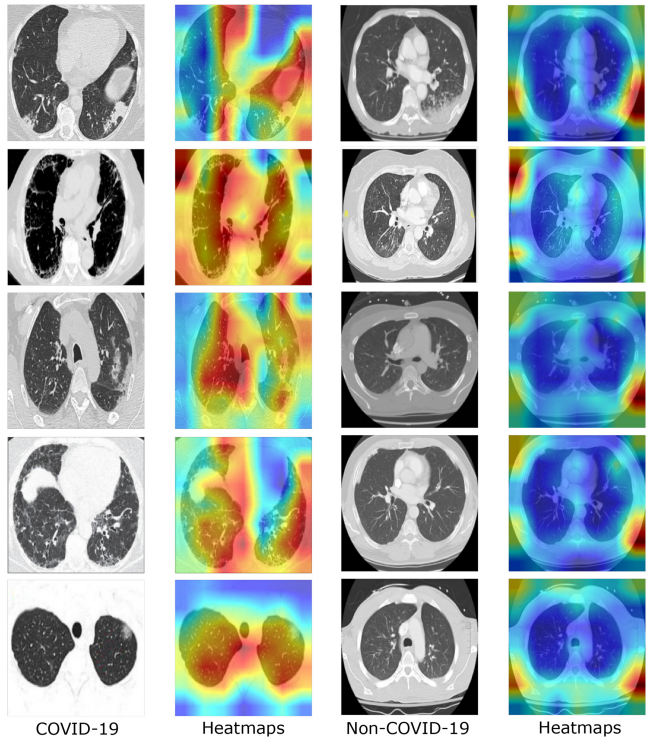

Fig. 6. Grad-CAM visualization results of LiMS-Net on both COVID-19 and non-COVID-19 CT images. The first and third columns provide the original CT images. The second and fourth columns provide their corresponding heatmaps.

5.4 Grad-CAM Visualization of LiMS-Net

The comparison with ImageNet pretrained models and existing DL-based COVID-19 diagnosis methods demonstrated the efficacy of our LiMS-Net model. But, to get a deeper understanding of the behaviour of the proposed model, we provide the visual explanation of the model predictions using Grad-CAM [23]. Grad-CAM helps to crosscheck whether the model is learning relevant features from the image instead of learning background or other irrelevant features and predicting the output based on them. Figure 6 depicts the heatmap results for both COVID-19 and non-COVID-19 samples using Grad-CAM visualizations. The figure shows that the heatmaps obtained by our model can cover infected lung regions in most COVID-19 CT scans. While for non-COVID-19 CT scans, there is no indication of infected regions in the heatmaps. These results indicate better interpretability of the classification results by our model and hence, can be helpful for clinical COVID-19 diagnosis.

5.5 Evaluation of Proposed Model on SARS-CoV-2 CT-Scan Dataset

For further verifying the effectiveness of our proposed model, we used a dataset "SARS-CoV-2 CT-Scan" [1] which is comparatively larger than the COVID-CT dataset. The SARS-CoV-2 CT-Scan dataset consists of 1,252 CT scans from 60 COVID-19 positive patients and 1,230 CT scans obtained from 60 patients non-infected by COVID-19. To ensure a fair comparison, we adopted the same data split as adopted in [1]. We divided the train set into train and validation sets using 80:20 ratio, respectively. The number of CT images in each set is summarized in Table 13.

Table 13. Description of SARS-CoV-2 CT-Scan Dataset

Set	COVID-19	Non-COVID-19
Train	802	788
Validation	200	196
Test	250	246
Total Images	1252	1230

Table 14. Classification Performance Comparison (in %) with ImageNet Pretrained CNN Models on SARS-CoV-2-CT-Scan Dataset

Model	Acc	F1	AUC	Spec	Sens
DenseNet-169	98.99	98.99	99.00	100	98.00
ResNet-101	97.78	97.76	97.78	97.97	97.60
ResNet-50	98.79	98.79	98.80	100	97.60
DenseNet-121	98.79	98.78	98.79	99.19	98.40
Inception-V3	97.98	98.00	98.00	100	96.00
Xception	97.78	97.76	97.78	97.97	97.60
MobileNet	98.79	98.78	98.79	97.63	100
VGG-16	98.79	98.78	98.79	99.19	98.40
Our LiMSNet	**99.19**	**99.20**	**99.08**	**99.59**	**98.80**

The performance comparison of the proposed model with ImageNet pretrained CNN models on this is shown in Table 14. It can be observed that the model achieves higher performance than the ImageNet models in spite of being significantly lighter.

6 CONCLUSION

In this paper, we proposed a lightweight multi-scale CNN model termed as LiMS-Net to effectively detect COVID-19 infection from chest CT scans. The LiMS-Net introduced two feature learning blocks that contains two layers of convolutions using different filter sizes such as 3×3 and 5×5 to extract multi-scale features. This two layer lightweight structure helped in learning discriminable features even in the presence of limited CT images. The LiMS-Net was compared with several ImageNet pretrained models along with existing DL based COVID-19 detection methods using a publicly available COVID-19 CT dataset. Experimental results indicated that LiMS-Net achieved comparable or better classification performance with far fewer model parameters and memory space which makes it suitable for real-time COVID-19 diagnosis. Our future work is to validate LiMS-Net using an even more large and diverse dataset. Also, we intend to incorporate attention mechanisms into our model to enhance the performance by enabling it to focus merely on essential features of the CT image. Further, in the future, the validation performance of the model could be estimated in detail using methods such as bootstrap [4].

REFERENCES

[1] Plamen Angelov and Eduardo Almeida Soares. 2020. SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. *MedRxiv* (2020).

[2] François Chollet. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1251–1258.

[3] Michael Chung, Adam Bernheim, Xueyan Mei, Ning Zhang, Mingqian Huang, et al. 2020. CT imaging features of 2019 novel Coronavirus (2019-nCoV). *Radiology* 295, 1 (2020), 202–207.

[4] Jesse Dodge, Suchin Gururangan, Dallas Card, Roy Schwartz, and Noah A. Smith. 2019. Show your work: Improved reporting of experimental results. *arXiv preprint arXiv:1909.03004* (2019).
[5] Di Dong, Zhenchao Tang, Shuo Wang, Hui Hui, Lixin Gong, Yao Lu, Zhong Xue, Hongen Liao, Fang Chen, Fan Yang, et al. 2020. The role of imaging in the detection and management of COVID-19: A review. *IEEE Reviews in Biomedical Engineering* (2020).
[6] Yicheng Fang, Huangqi Zhang, Jicheng Xie, Minjie Lin, Lingjun Ying, Peipei Pang, and Wenbin Ji. 2020. Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. *Radiology* (2020), 200432.
[7] Md. Hasan, Md. Alam, Md. Elahi, E. Toufick, Shidhartho Roy, Sifat Redwan Wahid, et al. 2020. CVR-Net: A deep convolutional neural network for Coronavirus recognition from chest radiography images. *arXiv preprint arXiv:2007.11993* (2020).
[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.
[9] Xuehai He, Xingyi Yang, Shanghang Zhang, Jinyu Zhao, Yichen Zhang, Eric Xing, and Pengtao Xie. 2020. Sample-efficient deep learning for COVID-19 diagnosis based on CT scans. *MedRxiv* (2020).
[10] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017).
[11] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4700–4708.
[12] Yifan Jiang, Han Chen, M. H. Loew, and Hanseok Ko. 2020. COVID-19 CT image synthesis with a conditional generative adversarial network. *IEEE Journal of Biomedical and Health Informatics* (2020).
[13] Taranjit Kaur, Tapan K. Gandhi, and Bijaya K. Panigrahi. 2021. Automated diagnosis of COVID-19 using deep features and parameter free BAT optimization. *IEEE Journal of Translational Engineering in Health and Medicine* (2021).
[14] Shaohua Li, Yong Liu, Xiuchao Sui, Cheng Chen, Gabriel Tjio, Daniel Shu Wei Ting, and Rick Siow Mong Goh. 2019. Multi-instance multi-scale CNN for medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 531–539.
[15] Xiaoming Li, Wenbing Zeng, Xiang Li, Haonan Chen, et al. 2020. CT imaging changes of Corona virus disease 2019 (COVID-19): A multi-center study in southwest China. *Journal of Translational Medicine* 18 (2020), 1–8.
[16] Xueyan Mei, Hao-Chih Lee, Kai-yue Diao, Mingqian Huang, Bin Lin, et al. 2020. Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. *Nature Medicine* 26, 8 (2020), 1224–1228.
[17] Soumya Ranjan Nayak, Deepak Ranjan Nayak, Utkarsh Sinha, Vaibhav Arora, and Ram Bilas Pachori. 2021. Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study. *Biomedical Signal Processing and Control* 64 (2021), 102365.
[18] Thanh Thi Nguyen, Quoc Viet Hung Nguyen, Dung Tien Nguyen, Edbert B. Hsu, Samuel Yang, and Peter Eklund. 2020. Artificial intelligence in the battle against Coronavirus (COVID-19): A survey and future research directions. *arXiv preprint arXiv:2008.07343* (2020).
[19] Yujin Oh, Sangjoon Park, and Jong Chul Ye. 2020. Deep learning COVID-19 features on CXR using limited training data sets. *IEEE Transactions on Medical Imaging* 39, 8 (2020), 2688–2700.
[20] Matteo Polsinelli, Luigi Cinque, and Giuseppe Placidi. 2020. A light CNN for detecting COVID-19 from CT scans of the chest. *Pattern Recognition Letters* 140 (2020), 95–100.
[21] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. 2017. Swish: A self-gated activation function. *arXiv preprint arXiv:1710.05941* (2017).
[22] Muhammad Saqib, Saeed Anwar, Abbas Anwar, Michael Blumenstein, et al. 2020. COVID-19 detection from radiographs: Is Deep Learning able to handle the crisis?
[23] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*. 618–626.
[24] Feng Shi, Jun Wang, Jun Shi, Ziyan Wu, Qian Wang, Zhenyu Tang, Kelei He, Yinghuan Shi, and Dinggang Shen. 2020. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. *IEEE Reviews in Biomedical Engineering* (2020).
[25] Hoo-Chang Shin, Holger R. Roth, Mingchen Gao, Le Lu, et al. 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging* 35, 5 (2016), 1285–1298.
[26] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
[27] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1–9.

[28] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2818–2826.

[29] Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, et al. 2021. A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). *European Radiology* (2021), 1–9.

[30] Shui-Hua Wang, Deepak Ranjan Nayak, David S. Guttery, Xin Zhang, and Yu-Dong Zhang. 2021. COVID-19 classification by CCSHNet with deep fusion using transfer learning and discriminant correlation analysis. *Information Fusion* 68 (2021), 131–148.

[31] Xiaowei Xu, Xiangao Jiang, Chunlian Ma, Peng Du, Xukun Li, Shuangzhi Lv, Liang Yu, Qin Ni, Yanfei Chen, Junwei Su, et al. 2020. A deep learning system to screen novel Coronavirus disease 2019 pneumonia. *Engineering* 6, 10 (2020), 1122–1129.

[32] Yu-Dong Zhang, Suresh Chandra Satapathy, Li-Yao Zhu, Juan Manuel Górriz, and Shui-Hua Wang. 2020. A seven-layer convolutional neural network for chest CT based COVID-19 diagnosis using stochastic pooling. *IEEE Sensors Journal* (2020).

[33] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie. 2020. COVID-CT-dataset: A CT scan dataset about COVID-19. 490 (2020). *arXiv preprint arXiv:2003.13865*.

Received 4 June 2021; revised 14 March 2022; accepted 13 July 2022