This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

He, Lang; Tiwari, Prayag; Su, Rui; Shi, Xiuying; Marttinen, Pekka; Kumar, Neeraj

COVIDNet: An Automatic Architecture for COVID-19 Detection with Deep Learning from Chest X-ray Images

Published in:
IEEE Internet of Things Journal

DOI:
[10.1109/JIOT.2021.3126471](https://doi.org/10.1109/JIOT.2021.3126471)

E-pub ahead of print: 01/01/2021

Document Version
Peer reviewed version

Please cite the original version:

He, L., Tiwari, P., Su, R., Shi, X., Marttinen, P., & Kumar, N. (2021). COVIDNet: An Automatic Architecture for COVID-19 Detection with Deep Learning from Chest X-ray Images. *IEEE Internet of Things Journal*, PP(99), [9608952]. <https://doi.org/10.1109/JIOT.2021.3126471>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

COVIDNet: An Automatic Architecture for COVID-19 Detection with Deep Learning from Chest X-ray Images

Lang He¹, Prayag Tiwari², Rui Su, Xiuying Shi, Pekka Marttinen³, Neeraj Kumar⁴

Abstract—Up to now, the COVID-19 has been sweeping across all over the world, which has affected individual's lives in an overwhelming way. To fight efficiently against the COVID-19, radiography and radiology images are used by clinicians in hospitals. This paper presents an integrated framework, named COVIDNet, for classifying COVID-19 patients and healthy controls. Specifically, ResNet (i.e., ResNet-18 and ResNet-50) is adopted as a backbone network to extract the discriminative features first. Second, the spatial pyramid pooling (SPP) layer is adopted to capture the middle-level features from the features of ResNet. To learn the high-level features, the NetVLAD layer is used to aggregate the features representation from middle-level features. Context gating (CG) mechanism is adopted to further learn the high-level features for predicting the COVID-19 patients or not. Finally, extensive experiments are conducted on the collected database, showing the excellent performance of the proposed integrated architecture, with the sensitivity up to 97%, and specificity of 99.5% of the ResNet-18, and with the sensitivity up to 99%, and specificity of 99.4% of the ResNet-50.

Index Terms—COVID-19, Spatial Pyramid Pooling (SPP), Deep Learning, NetVLAD

I. INTRODUCTION

IN December 2019, an unknown outbreak of pneumonia was reported in a province in mainland China and soon spread to other cities of China. It was soon found to be related to

This work is supported by the Shaanxi Provincial Social Science Foundation (grant 2021K015), the Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No. 19JS028), the Academy of Finland (grants 336033, 315896), Business Finland (grant 884/31/2018), and EU H2020 (grant 101016775). (Corresponding authors: Neeraj Kumar and Prayag Tiwari)

Lang He and Prayag Tiwari contributed equally to this work.

Lang He is with the School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, Shaanxi, China, with the Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an 710121, Shaanxi, China, and with the Xi'an Key Laboratory of Big Data and Intelligent Computing, Xi'an 710121, Shaanxi, China. (E-mail: langhe@xupt.edu.cn)

Prayag Tiwari and Pekka Marttinen are with the Department of Computer Science, Aalto University, Espoo, Finland. (E-mail: prayag.tiwari@aalto.fi, pekka.marttinen@aalto.fi)

Rui Su is with the School of Foreign Languages, Northwest University, Xi'an Shaanxi, China.

Xiuying Shi is with the Affiliated Hospital of Yan'an University, Yan'an Shaanxi, China.

Neeraj Kumar is with the Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology (Deemed University, Patiala, Punjab, India). He is also associated with School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand and Department of Computer Science and Information Engineering, Asia University, Taiwan. (E-mail: neeraj.kumar@thapar.edu)

Manuscript received XXXXXX; revised XXXXXX.

a novel coronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). After that, infected cases with the coronavirus disease 2019 (COVID-19) were also found in other countries and was declared a pandemic, infecting about 170 million individuals as well as resulting in more than 3.5 million deaths as of 31 May 2021¹.

To hold back the transmission of COVID-19 disease, the efficient approach is to isolate the COVID-19 patients and healthy people. Also, early detection is significant for blocking the spread between individuals. Recently, real-time reverse transcription-polymerase chain reaction (RT-PCR) detection of SARS-CoV-2-specific ribonucleic acid (RNA) in respiratory specimens from infected patients has been adopted for COVID-19 detection. The development of RT-PCR assays for SARS-CoV-2 has gained increasing attention, and many commercial assay kits have been designed. However, only 30-60% throat swab samples can be detected for a positive infected person, which leads to infecting a great number of healthy individuals [1]–[3]. Hence, due to the lack of availability for diagnosing COVID-19, researchers from different fields try to explore complementary approaches to assist clinicians in detecting or isolate infected individuals as early as possible. Several studies investigated CT chest images for detecting COVID-19. For instance, [4] proposes to use volumetric CT chest images to estimate whether a person is infected with the COVID-19 or not. The usage of Chest CT shows high accuracy for detecting the COVID-19 [5]. While the usage of X-ray images represents certain relationships of COVID-19 [6]. Ground glass patterns around the pulmonary vessels are complicated to distinguish visually [7]. In addition, the features of symmetric patchy and diffuse airspace opacities are also considered as important cues for detecting COVID-19 [8]. Therefore, the abnormalities mentioned above can only be explained by professional radiologists. There is a contradiction between the massive suspected individuals and the shortage of trained radiologists for detecting COVID-19. To cut off the spread of COVID-19, it is necessary to design an efficient tool to help clinicians and radiologists to detect the COVID-19.

Currently, deep learning (DL) has been adopted in various fields, e.g., computer vision, medical image processing, natural language processing (NLP), and proven to obtain excellent performance [9]. Advanced computer-aided diagnosis (CAD) methods for medical image analysis are mainly on the state-of-the-art approaches, e.g., Transformer [10], and TransUNet

¹<https://www.worldometers.info/coronavirus/>

[11]. Especially in the work of [11], TransUNet leverages the advantage of Transformers and U-Net for medical image segmentation. TransUNet obtains promising performances to existing methods on several medical applications, e.g., cardiac segmentation and multi-organ segmentation, and so on.

Since 2019, COVID-19 has been widespread worldwide and still lingered to affect people's daily lives. Therefore, many researchers dedicate their efforts to use computer vision and DL methods to address severe situations. [12] adopts DarkNet to distinguish the COVID-19 patients and healthy persons and obtains the accuracy of 98.08%. In addition, the method of DarkNet obtains 87.02% from the COVID-19, healthy persons, and pneumonia individuals. The authors of [13] also introduce a new architecture for distinguishing COVID-19 based on chest X-ray (CXR) images that obtained an overall performance of 89.6%, and a recall accuracy for COVID-19 cases with 93% in classifying COVID-19, healthy conditions, bacterial pneumonia, and viral pneumonia. In addition, three deep architectures (i.e., Inception-V3, Inception-ResNet-V2, and ResNet) are adopted to differentiate COVID-19 and healthy persons, and the binary accuracy of 98% is achieved [14]. [15] proposes a framework that can accurately recognize COVID-19, which includes two procedures: the Discrimination-DL and the Localization-DL. The Discrimination-DL can be used for modeling the lung features from CXR radiographs. And the Localization-DL was trained with 406-pixel patches and utilized to recognize X-ray radiographs for localizing and arranging them into the right lung, left lung, or bipulmonary. The proposed architecture obtains the accuracy of 98.71% for Discrimination-DL and 93.03% for Localization-DL. The contribution of this work is that the DL approach can help the radiologists in detecting the localization for COVID-19 efficiently and accurately. [16] proposes a determined model named CCSHNet for detecting COVID-19 and other lung infectious diseases. The CCSHNet obtains the F1 score of 97.04%. It was found in [17] that ResNet-18 is adopted to check the disease from chest CT images, which obtained an accuracy of 92% for diagnosing COVID-19. Moreover, in [18], a new method employs DL to recognize clinical abnormalities from CXR at a certain level exceeding professional radiologists.

Computer-aided detection (CAD) has been used successfully in clinical applications. However, the studies mentioned earlier have been proposed to adopt various DL methods to detect COVID-19. However, there still exist the following issues. Up to the present, some available databases are collected for research purposes [19]. First, the training data of COVID-19 is scarce to train the deep models for detecting COVID-19 and healthy persons. Second, the traditional medical image processing method contains two steps, i.e., hand-crafted feature extraction, classification, or regression by machine learning methods. In this paper, to overcome this issues, we first re-check and clean the available database. Then, we leverage the dataset provided by the work of [19], and double-check the resolution, labels of images to ensure the robustness of the training models. And then introduce an end-to-end framework named COVIDNet, which can automatically predict the COVID-19 from CXR images. Due to the

advantage of ResNet [20], we adopt it as a backbone network to extract the discriminative features in the first place. Second, the spatial pyramid pooling (SPP) layer is adopted to capture the middle-level features from the features of ResNet. After that, to learn the high-level features, the NetVLAD layer is used to aggregate the features representation from middle-level features. After performing feature aggregation operation, the context gating (CG) mechanism is adopted to further learn the high-level features for predicting the COVID-19 patients. Finally, extensive experiments are conducted on the collected database, showing the excellent performance of the proposed integrated architecture.

To sum up, the contribution of this work is three-fold:

- 1) An end-to-end architecture COVIDNet is presented that automatically models the deep learned features for predicting the COVID-19 based on X-ray images.
- 2) SPP is adopted to model the middle-level features from the obtained features of ResNet. And then, NetVLAD and CG mechanisms are adopted for feature aggregation to learn high-level features for COVID-19 detection.
- 3) Extensive experiments are carried out on the collected database and obtain the proposed architecture's promising performance.

The remained of the present paper is structured as follows. Section II details previous study on COVID-19 detection. Our method is introduced in Section III. Section IV discusses the experimental performance. Conclusions and future studies are planned in Section V.

II. RELATED WORKS

After the outbreak of COVID-19, various AI models are designed to detect it. The earlier work of [21] introduces a novel scheme for diagnosing COVID-19 based on CT scans. The framework adopts UNet++ [22] to detect coronavirus with CT scans. The authors collected a database consisting of 40,000 images from 106 subjects to generate the deep model. The performances showed that the proposed model could reduce the efforts of radiologists to detect the COVID-19.

Xu et al. [23] introduces a deep framework with DL to diagnose COVID-19 from CT scans. Inception network [24] is fine-tuned with the collected training data.

In addition to CT scans, CXR images are also investigated for COVID-19 detection. For instance, in [25], Wang et al. presents a deep architecture COVID-Net from CXR image to detect COVID-19, healthy persons, and individuals with pneumonia. In addition, the works of [26], [27] propose a deep architecture to diagnose COVID-19 from CXR images. Interestingly, [28] designed healthcare on 5G framework from CXR and CT images for the COVID-19 detection.

The aforementioned works of this section are mainly focused on investigating the DL methods for COVID-19 detection. Though the works have obtained promising performance using DL to detect COVID-19. However, the limited number of data can affect the performance of COVID-19 detection. Hence, the works of [29] show that radiologists pay attention to the limited data to train the deep DL models for COVID-19 detection. Further, to overcome the issue of the scarce training

Fig. 1. The pipeline of the proposed framework for COVID-19 detection. (a) Deep architecture (i.e., ResNet-18) is fine-tuned to extract deep-learned features. (b) SPP is used to aggregate the middle features from the features obtained of (a). (c) To learn the high-level features, NetVLAD is adopted. Finally, two fully connected layers are utilized to project the high-level features into a new feature space to obtain the classification of COVID-19 and Non-COVID.

data, the authors of [30] introduce an architecture based on DL that is capable of obtaining important improvement for detecting COVID-19 from scarce data samples with CT scans. Transfer learning and self-supervised contrastive learning are adopted for modeling unbiased features to address the over-fitted issue by designing a synergistic approach. Comparable performance obtained by conducting extensive experiments on the introduced database. After that, Chen et al. [31] presents a meta-learning model based on the integrated network for diagnosing COVID-19 based on chest CT images by a small size of data.

Existing works using CXR images to detect COVID-19 mainly utilize the traditional CNN framework or pre-trained deep models, which need a great number of training data to generate feature representation. By contrast, we not only leverage the CNN to extract the deep learned features, but adopt the custom feature aggregation to learn the high-level features to distinguish COVID-19 cases by limited CXR images.

III. OUR PROPOSED METHOD

In this sub-section, an overview is made to introduce architecture in Section III-A, and describe the feature extraction in Section III-B. SPP is introduced in Section III-C. In Section III-D, VLAD is described to aggregate the obtained features. CG mechanism is introduced in Section III-E.

A. System Overview

To illustrate the proposed method, the ensemble framework is shown in Fig. 1. We first collect the database (the detailed description will be given in Section IV-A) from the Internet based on the available data (CovidChestxray-Dataset:², ChexPert-Dataset). We first adopt deep architectures (i.e., ResNet-18, ResNet-50) to extract the discriminative feature

representations. To mine the informative features and filter useless features, SPP is adopted to learn the middle-level features. After that, to learn the high-level features, NetVLAD is used. Finally, two fully connected layers and softmax function are utilized to classify the COVID-19 or Non-COVID. In the following, we expatiate each part of the proposed framework.

B. Representation of Deep Learned Features

In general, as a result of the limited data samples of medical images, the studies in this field often fine-tune the pre-trained deep models to make special tasks. For instance, a large deep model trained on ImageNet (which consists of a great number of images) is to adapt for COVID-19 detection with a limited database. Furthermore, to further make the COVID-19 detection, transfer learning is adopted. Therefore, the parameters of the deep model can be learned with good initial values, and new tasks can be better managed with only a few small modifications.

In addition, the pre-trained model can be divided into two ways for detecting COVID-19. In a first way, the model is adopted to extract the deep features, and a machine learning model is adopted for classification. The other way is that the entire network or its sub-networks are fine-tuned to adapt to the special study. Hence, the weights of the deep model are considered as the parameter value of the new task and updated them in the training stage.

As for this situation, the training images of COVID-19 is complicated to collect from hospitals. Hence, to extract the informative features, ResNet is first used to extract the informative deep features, followed by using SPP to extract the middle-level features, and NetVLAD is adopted to aggregate the high-level features.

In our work, ResNet-18 (the red dotted rectangle in Fig. 1) is used for feature extraction. As shown in Fig. 1, sixteen convolution layers, two downsampling layers, and two fully

²<https://github.com/ieee8023/covid-chestxray-dataset>

connected layers are included in ResNet-18. For the ResNet-18, the convolution layer is 224×224 with the kernel size of 7×7 . As is known, the kernel size of other layers is 3×3 . Followed by the last convolution layer, an eigenvector is obtained by full connection. In addition, for the ResNet-50 with the homologous architect as the ResNet-18, the only difference is that it contains more layers.

C. Non-Variant Feature Representation–SPP

The main goal of our work is to detect people infected with COVID-19 from healthy persons. To capture the features for COVID-19 detection, the deep representation has to capture discriminative features in different scales; hence a significant number of instances in all its natural cases are needed in training.

In our task, to learn the scale-invariant characteristics, an SPP layer [32] is used to learn multi-scales on the top of the output of convolution features of ResNet-18 or ResNet-50. The idea of SPP is to divide the feature map into different divisions from finer to coarser scales, finalized with an aggregation of local features. Thus, the SPP layer can improve scale invariance and avoid the overfitting issue. In our study, we adopt the original weighting definition on the spatial pyramid kernel in [33]. Thus, the features with finer resolutions are associated with a heavier weight, and coarser resolutions with a lower weight.

The SPP layer can be adopted for COVID-19 detection in that it can project features from multi-level spatial bins at multiple scales to calculate the features over the whole slice. In addition, the SPP layer can promote the capacity of the proposed model to project the input size with the fixed-length features for COVID-19 detection.

D. NetVLAD

As a common feature aggregation method, Vector of Locally Aggregated Descriptors (VLAD) has been widely adopted in computer vision field. It aggregated the informative patterns of local descriptors over the image.

Formally, let $\mathbf{x}_i \in \mathbb{R}^{N \times D}$ be the local image descriptor, where N and D represent the number and the dimension of the local descriptors. After performing the VLAD operation, the output V is a $K \times D$ matrix, where K represents the “visual words” ($\mathbf{c}_k$). V is denoted with $K \times D$ matrix, and then convert into a vector to image classification. The (j, k) element of V can be written as:

$$V(j, k) = \sum_{i=1}^N a_k(\mathbf{x}_i)(x_i(j) - c_k(j)) \quad (1)$$

where $x_i(j)$ and $c_k(j)$ represent the j -th dimensions of the i -th descriptor and k -th cluster center. $a_k(\mathbf{x}_i)$ represents the membership of the descriptor $\mathbf{x}_i$ to k -th visual word, i.e., if $a_k(\mathbf{x}_i)=1$, which means the cluster $\mathbf{c}_k$ is close to the descriptor $\mathbf{x}_i$, and vice versa. Specifically, for each column k with D -dimension of V matrix, which means the sum of residuals $\mathbf{x}_i - \mathbf{c}_k$ for the cluster $\mathbf{c}_k$. After that, L2-normalized operation is performed on the matrix V to keep the data at the same scale.

To leverage the advantage of the CNN and VLAD, an integrated trainable layer named NetVLAD is introduced in [34]. For the original attribute of VLAD, hard assignment $a_k(\mathbf{x}_i)$ of descriptors $\mathbf{x}_i$ is applied to the cluster center $\mathbf{c}_i$. Therefore, to perform the operation discernable, soft assignment operation is replaced to multiple clusters:

$$\bar{a}_k(\mathbf{x}_i) = \frac{e^{-\alpha \|\mathbf{x}_i - \mathbf{c}_k\|^2}}{\sum_{k'} e^{-\alpha \|\mathbf{x}_i - \mathbf{c}_{k'}\|^2}} \quad (2)$$

which arranges the parameters of descriptor $\mathbf{x}_i$ to cluster $\mathbf{c}_k$ according to their proximity, but relative to proximities to other cluster centers. The range of $\bar{a}_k(\mathbf{x}_i)$ is from 0 to 1, the threshold near to 1 means that the closest to cluster center. α is a positive constant used to weight the decay of the response by the magnitude of the distance. Special emphasis is put on that $\alpha \rightarrow +\infty$ means the original VLAD as $\bar{a}_k(\mathbf{x}_i)$ would be equal to 1 if $\mathbf{x}_i$ closest cluster is k and 0 otherwise.

Fig. 2. Illustration of the supervised VLAD. Red and green circles represent different local descriptors from two images, which have the same cluster. After the VLAD encoding operation, a residual vector is computed as the difference between the cluster’s anchor point and the descriptor. The anchor point $\mathbf{c}_k$ is the origin of a new coordinate system local to the specific cluster k . For the traditional VLAD, the anchor is cluster centre ($\bullet$) to assign the residuals across the database. While for the NetVLAD, which can learn an anchor $\star$ that is the small distance between the new residuals.

To perform the expanding operation on the Equ. 2, which can be re-written as follows:

$$\bar{a}_k(\mathbf{x}_i) = \frac{e^{\mathbf{w}_k^T \mathbf{x}_i + b_k}}{\sum_{k'} e^{\mathbf{w}_{k'}^T \mathbf{x}_i + b_{k'}}} \quad (3)$$

Where $\mathbf{w}_k$ represents a vector with the value of $\mathbf{w}_k = 2\alpha \mathbf{c}_k$, and b_k denotes a scalar with the value of $b_k = -\alpha \|\mathbf{c}_k\|^2$. Then the final NetVLAD layer can be formed by inserting the soft-assignment (Equ. 3) into the VLAD descriptor (Equ. 1) as follows:

$$V(j, k) = \sum_{i=1}^N \frac{e^{\mathbf{w}_k^T \mathbf{x}_i + b_k}}{\sum_{k'} e^{\mathbf{w}_{k'}^T \mathbf{x}_i + b_{k'}}} (x_i(j) - c_k(j)) \quad (4)$$

where $\mathbf{w}_k$, b_k and $\mathbf{c}_k$ represent the parameters of cluster k . The same as the original VLAD, the NetVLAD layer can aggregate the first order statistics of residuals $\mathbf{x}_i - \mathbf{c}_k$ on the different descriptor space via the soft-assignment $\bar{a}_k(\mathbf{x}_i)$ of descriptor $\mathbf{x}_i$ to cluster k . Furthermore, the primary difference is that the NetVLAD layer contains three parameters, i.e., $\mathbf{w}_k$, b_k and $\mathbf{c}_k$, while the original VLAD only has one parameter $\mathbf{c}_k$. As shown in Fig. 2, the NetVLAD has greater flexibility when compared the VLAD. For the parameters of NetVLAD, which are learnt via an end-to-end manner for different tasks.

TABLE I
THE NUMBER OF IMAGES IN THE TRAINING AND TEST PARTITION.

Partition	COVID-19	Non-COVID
Training Set	420	2000
Test Set	100	3000

Meanwhile, as shown in NetVLAD component (the black dashed rectangle) of Fig. 1, the NetVLAD layer has the following attributes. Firstly, the first term of Equ. 4 can be considered as a softmax function $\sigma_k(\mathbf{z}) = \frac{\exp(z_k)}{\sum_{k'} \exp(z_{k'})}$. Hence, the procedure of the soft-assignment for the input descriptor $\mathbf{x}_i$ with K clusters can be explained as follows: 1) a convolution 1×1 with the parameters (K filters $\mathbf{w}_k$ and biases b_k), generating the output $s_k(\mathbf{x}_i) = \mathbf{w}_k^T \mathbf{x}_i + b_k$; 2) the output $s_k(\mathbf{x}_i)$ is performed with the softmax function σ_k to generate the final output $\bar{a}_k(\mathbf{x}_i)$. Finally, normalization operation is adopted for the final output to obtain a representation $F_{netvlad}$ with the size of $(K \times D) \times 1$.

E. Context Gating (CG)

The CG mechanism is adopted to project the output feature $F_{netvlad}$ into an another representation F_{cg} as

$$F_{cg} = \psi(\bar{W}F_{netvlad} + b) \circ F_{netvlad} \quad (5)$$

where $F_{netvlad} \in \mathbb{R}^n$ represents the input feature vector, ψ denotes the element-wise sigmoid activation and $\circ$ represents the element-wise multiplication. $\bar{W} \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$ are the trainable parameters. The weights vector $\psi(\bar{W}F_{netvlad} + b) \in [0, 1]$ denotes a series of learned gates utilized to the respective dimensions of the feature $F_{netvlad}$.

In our task, the reasons for adopting CG are as follows. Firstly, non-linear interactions among activations of the $F_{netvlad}$ feature are learned by CG. Secondly, the strengths of different activations of the $F_{netvlad}$ through a self-gating mechanism are recalibrated by CG.

After performing the aforementioned methods, two fully connected layers and Cross-entropy function are adopted for the final classification.

IV. EXPERIMENTS

This section describes the experimental evaluation for the proposed architecture. In Section IV-A, we introduce the collected database adopted in this paper. Evaluation metrics are introduced in Section IV-B2. In Section IV-C, we discuss the experimental performance for COVID-19 detection.

A. Datasets

To evaluate the introduced architecture, the database of the work [19] is adopted in our study. The database COVID-Xray-5k are collected based on the Covid-Chestxray-Dataset and ChexPert databases. As shown in Table I, the COVID-19 data samples consist of 100 images of the test partition, and 84 images of the training partition. The non-COVID data samples include 3000 images on the test set and 2000 images on the training set. In addition, the radiologist from the affiliated hospital of Yan'an university, Shaanxi, China, were asked to

Fig. 3. The random selected images of COVID-Xray-5k database. The four images of the first row represent the COVID-19. The four images of the second row denote the Non-COVID in the ChexPert database [35]. For the rest rows, which selected eight images from the ChexPert.

re-check and re-label the data samples to ensure the accuracy of the COVID-19 detection.

As shown in Fig. 3, sixteen images are selected in for COVID-Xray-5k database, containing four COVID-19 images in the first row, four Non-COVID images in the second row from the ChexPert database, and eight images from ChexPert in the third and fourth row with one of the 13 diseases.

B. Experimental Setup and Evaluation Measures

This section provides a demonstration of the experimental evaluation of the introduced architecture for COVID-19 detection. Extensive experiments are performed on the collected database to estimate the proposed architecture. The aim of the experiments is twofold.

- Evaluate the capability for the architecture COVIDNet for COVID-19 detection.
- Apply the designed architecture COVIDNet to the clinical usage.

1) *Evaluation Setup*: In our work, to estimate the capability of the architecture, ResNet-18 and ResNet-50 networks are fine-tuned with 50 epochs. The batch size is 48. SGD optimizer is adopted to optimize the Cross-entropy loss function. The learning rate is set to 0.0001, and the momentum is set to 0.9. All the images are set to 224×224 to input the ResNet-18 and ResNet-50. Our deep models are trained with the PyTorch platform (<https://pytorch.org/>). The experiments are performed on two Titan-X GPU (each with 12G memory).

2) *Evaluation Measures*: The performance of COVID-19 detection is measured according to the sensitivity, specificity, and F1-score between the predicted and reported labels. The sensitivity, specificity, F1-score can be written as follows:

TABLE II

EVALUATIONS FOR THE VARIOUS COMBINATIONS OF THE ARCHITECTURE USING RESNET-18 WITHOUT PRE-TRAINING ON THE TEST SET OF THE COLLECTED DATABASE.

Model Setting	ROC	Sensitivity	Specificity	Accuracy	F1-score
A1: ResNet-18_FC	0.749	73%	71%	70%	70%
B1: ResNet-18_FC_SPP	0.725	75%	74%	71%	72%
C1: ResNet-18_FC_SPP_VLAD	0.714	76%	75%	75%	71%
D1: ResNet-18_FC_SPP_VLAD_CG	0.733	77%	75%	77%	72%

TABLE III

EVALUATIONS FOR THE VARIOUS COMBINATIONS OF THE ARCHITECTURE USING RESNET-50 WITHOUT PRE-TRAINING ON THE TEST SET OF THE COLLECTED DATABASE.

Model Setting	ROC	Sensitivity	Specificity	Accuracy	F1-score
A2: ResNet-50_FC	0.721	74%	74%	72%	70%
B2: ResNet-50_FC_SPP	0.754	76%	76%	71%	72%
C2: ResNet-50_FC_SPP_VLAD	0.712	75%	71%	70%	73%
D2: ResNet-50_FC_SPP_VLAD_CG	0.796	77%	71%	71%	73%

TABLE IV

EVALUATIONS FOR THE VARIOUS COMBINATIONS OF THE ARCHITECTURE USING RESNET-18 ON THE TEST SET OF THE COLLECTED DATABASE.

Model Setting	ROC	Sensitivity	Specificity	Accuracy	F1-score
A3: ResNet-18_FC	0.988	97%	94.2%	97%	96.7%
B3: ResNet-18_FC_SPP	0.996	98%	86.6%	99.5%	92.8%
C3: ResNet-18_FC_SPP_VLAD	0.961	85%	92.4%	99.5%	95.8%
D3: ResNet-18_FC_SPP_VLAD_CG	0.977	90%	96.17%	99.5%	97.9%

$$Specificity = D / (D + B) \times 100 \quad (6)$$

Where D represents the correctly predicted as COVID-19, B is the wrongly predicted as Non-COVID.

$$Sensitivity = A / (A + C) \times 100 \quad (7)$$

Where A represents the correctly predicted as Non-COVID, C is the wrongly predicted as COVID-19.

$$F1 - score = 2A / (2A + C + B) \times 100 \quad (8)$$

C. Results

In this part, the experiments are conducted by ablation study for evaluating the overall performance of each sub-framework of the proposed framework. Then a comparison is made with other methods to show the performance of the framework.

1) *Ablation Study*: The performance of each sub-framework of the designed scheme is evaluated on the collected database. To make a clear explanation, as shown in Table II, III, IV and V, the model setup “ResNet-18_FC” and “ResNet-50_FC” denotes that ResNet-18 with two fully connected layers, the rest are like this observation. In addition, other model setting in Tables II, III, IV and V equips two or three modules to learn the discriminative patterns among them.

To compare the accuracy for COVID-19 detection, ResNet-50 without pre-training is also evaluated. As shown in Table II, for the metric of accuracy, all the combinations obtain the comparable performance for COVID-19 detection. As shown in Table III, to further validate the performance of our method, ResNet-50 without pre-training is also adopted. From both tables, we can observe that ResNet-18 and ResNet-50 without pre-training can obtain poor performance for COVID-19 detection.

As shown in Table IV, for the metric of accuracy, all the combinations obtain the comparable performance of 99.5% except A3. In addition, for the sensitivity metric, A1 obtains the best accuracy, and D3 obtains the weak accuracy for COVID-19 detection. The reason is that an unbalanced test set (3000 Non-COVID data samples, 100 COVID-19 data samples) is adopted in our task for COVID-19 detection. In addition, one can note that A3 obtains the best sensitivity of 97%, and D3 achieves 90% for COVID-19 detection. The reason is twofold: 1) The structure of D1 is more complicated for training the deep models, while for A3, only two append two fully connected layers are appended; 2) The sacrifice of data samples also causes the discrepancy of the performance.

However, as shown in Table IV, the specificity demonstrates another representation for Non-COVID samples. One can note that D3 obtains the best performance of 96.17% for classifying the Non-COVID images. Also, for A3, it obtains better performances with 94.2%. From these results, conclusion can be drawn that the data samples can affect the performance of COVID-19 detection.

As shown in Table V, to further test the performance for our method, ResNet-50 is adopted. The same observation found as the ResNet-18, A4 also obtains the best performance for the sensitivity of 99%, and D4 obtains the best result of 99.4%. These results indicate that the observation concluded for the Table IV. These results also show the availability of the architecture, which can learn the patterns for COVID-19 detection.

To further validate the proposed method for COVID-19 detection, SqueezeNet and Densenet-121 are adopted. As a result, as shown in Table VI, D4 obtains the best performance for COVID-19 detection. In addition, of course, the four methods have a high sensitivity rate exceeding 92% and obtain a comparable specificity rate.

TABLE V
EVALUATIONS FOR THE VARIOUS COMBINATIONS OF THE ARCHITECTURE USING RESNET-50 ON THE TEST SET OF THE COLLECTED DATABASE.

Model Setting	ROC	Sensitivity	Specificity	Accuracy	F1-score
A4: ResNet-50_FC	0.991	99%	85.2%	99.8%	91.2%
B4: ResNet-50_FC_SPP	0.999	99%	95.8%	99.8%	97.0%
C4: ResNet-50_FC_SPP_VLAD	0.995	97%	95.3%	99.8%	97.5%
D4: ResNet-50_FC_SPP_VLAD_CG	0.999	92%	99.4%	99.8%	99.6%

TABLE VI
DIFFERENT DEEP NETWORKS ARE USED FOR COVID-19 DETECTION ON THE TEST SET OF THE COLLECTED DATABASE.

Model Setting	ROC	Sensitivity	Specificity	Accuracy	F1-score
D3	97.7%	90%	96.17%	99.5%	97.9%
D4	99.9%	92%	99.4%	99.8%	99.6%
SqueezeNet_FC_SPP_VLAD_CG	0.969	90%	94%	98.7%	92.1%
Densenet-121_FC_SPP_VLAD_CG	0.92	89%	92%	96.5%	91.8%

TABLE VII
PERFORMANCE COMPARISON WITH DIFFERENT DEEP ARCHITECTURES FOR COVID-19 DETECTION.

Model Setting	ROC	Sensitivity	Specificity	Accuracy	F1-score
D3 (Ours)	97.7%	90%	96.17%	99.5%	97.9%
D4 (Ours)	99.9%	92%	99.4%	99.8%	99.6%
ResNet-18 [19]	98%±2.7%	90.7%±1.1%	95.3%	—	—
ResNet-50 [19]	98%±2.7%	89.6%±1.1%	—	—	—
SqueezeNet [19]	98%±2.7%	92.9%±0.9%	—	—	—
Densenet-121 [19]	98%±2.7%	75.1%±1.5%	—	—	—
TLCov-VGG-16 [36]	—	—	—	97.67%	96.59%

Fig. 4. The confusion matrix of the proposed architect with different combinations.

Comparison is also provided between our results, using the combination of ResNet, SPP, and NetVLAD methods, with those state-of-the-art methods with other deep learning methods. The results on the database are shown in Table VII. From the reported results, it can be noted that the proposed methods obtain comparable performance for COVID-19 detection.

In addition, to further illustrate the right image samples for Non-COVID and COVID-19, confusion matrices for the ResNet-18 and ResNet-50 Fig. 4 for the 3100 images on the test set.

Furthermore, to further visually illustrate the results of COVID-19 detection, the method of [37] is used to visualize the deep architectures. As shown in Fig. 5, the first row is the randomly selected COVID-19 images, and the other row is the heatmap for the first row. One can note that the proposed method can detect the most discriminative regions for COVID-19 detection.

V. CONCLUSION

We present a learnable framework named COVIDNet for detecting COVID-19 from X-ray images in this paper. In the framework, ResNet (i.e., ResNet-18 and ResNet-50) is utilized to the backbone network to extract the discriminative representations. Second, the SPP layer is adopted to capture the middle-level features from the features of ResNet. After that, to learn the high-level features, the NetVLAD layer is used to aggregate the features representation from middle-level features. After performing feature aggregation operation, the CG mechanism is adopted to learn further the high-level representations for predicting the COVID-19 patients and NON-COVID. Lastly, we carry out extensive experiments to estimate the proposed architecture performance for COVID-19 detection, with the sensitivity of up to 97%, and specificity of 99.5% of the ResNet-18, and with the sensitivity up to 99%, and specificity of 99.4% of the ResNet-50.

In future work, we will augment the database for COVID-19 detection and also collaborate with the hospitals to track the COVID-19 patients to study the depression symptoms.

REFERENCES

- [1] Y. Yang, M. Yang, C. Shen, F. Wang, J. Yuan, J. Li, M. Zhang, Z. Wang, L. Xing, J. Wei *et al.*, "Laboratory diagnosis and monitoring the viral shedding of 2019-ncov infections," *MedRxiv*, 2020.
- [2] K. H. Abdulkareem, M. A. Mohammed, A. Salim, M. Arif, O. Geman, D. Gupta, and A. Khanna, "Realizing an effective covid-19 diagnosis system based on machine learning and iot in smart hospital environment," *IEEE Internet of Things Journal*, 2021.
- [3] K. Qian, M. Schmitt, H. Zheng, T. Koike, J. Han, J. Liu, W. Ji, J. Duan, M. Song, Z. Yang *et al.*, "Computer audition for fighting the sars-cov-2 corona crisis—introducing the multi-task speech corpus for covid-19," *IEEE Internet of Things Journal*, 2021.
- [4] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, J. Cui, W. Xu, Y. Yang, Z. A. Fayad *et al.*, "Ct imaging features of 2019 novel coronavirus (2019-ncov)," *Radiology*, vol. 295, no. 1, pp. 202–207, 2020.

Fig. 5. The heatmaps of COVID-19 detected by the ResNet-18 deep model. The six X-ray images are randomly selected from the test set of the collected database. Thus, the first row is the original images, and the second row is the heatmap for the first row.

- [5] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: a report of 1014 cases," *Radiology*, vol. 296, no. 2, pp. E32–E40, 2020.
- [6] J. P. Kanne, B. P. Little, J. H. Chung, B. M. Elicker, and L. H. Ketai, "Essentials for radiologists on covid-19: an update—radiology scientific expert panel," 2020.
- [7] D. M. Hansell, A. A. Bankier, H. MacMahon, T. C. McLoud, N. L. Muller, and J. Remy, "Fleischner society: glossary of terms for thoracic imaging," *Radiology*, vol. 246, no. 3, pp. 697–722, 2008.
- [8] J. Rodrigues, S. Hare, A. Edey, A. Devaraj, J. Jacob, A. Johnstone, R. McStay, A. Nair, and G. Robinson, "An update on covid-19 for the radiologist—a british society of thoracic imaging statement," *Clinical radiology*, vol. 75, no. 5, pp. 323–325, 2020.
- [9] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.
- [11] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.
- [12] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of covid-19 cases using deep neural networks with x-ray images," *Computers in biology and medicine*, vol. 121, p. 103792, 2020.
- [13] A. I. Khan, J. L. Shah, and M. M. Bhat, "Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105581, 2020.
- [14] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *Pattern Analysis and Applications*, pp. 1–14, 2021.
- [15] Z. Wang, Y. Xiao, Y. Li, J. Zhang, F. Lu, M. Hou, and X. Liu, "Automatically discriminating and localizing covid-19 from community-acquired pneumonia on chest x-rays," *Pattern recognition*, vol. 110, p. 107613, 2021.
- [16] S.-H. Wang, D. R. Nayak, D. S. Guttery, X. Zhang, and Y.-D. Zhang, "Covid-19 classification by ccshnet with deep fusion using transfer learning and discriminant correlation analysis," *Information Fusion*, vol. 68, pp. 131–148, 2021.
- [17] X. Mei, H.-C. Lee, K.-y. Diao, M. Huang, B. Lin, C. Liu, Z. Xie, Y. Ma, P. M. Robson, M. Chung *et al.*, "Artificial intelligence-enabled rapid diagnosis of patients with covid-19," *Nature medicine*, vol. 26, no. 8, pp. 1224–1228, 2020.
- [18] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz *et al.*, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the cheXnext algorithm to practicing radiologists," *PLoS medicine*, vol. 15, no. 11, p. e1002686, 2018.
- [19] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning," *Medical image analysis*, vol. 65, p. 101794, 2020.
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [21] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, Q. Chen, S. Huang, M. Yang, X. Yang *et al.*, "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography," *Scientific reports*, vol. 10, no. 1, pp. 1–11, 2020.
- [22] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 3–11.
- [23] B. Xu and X. Meng, "A deep learning algorithm using ct images to screen for corona virus disease (covid-19)," *Preprint*.
- [24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [25] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.
- [26] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 635–640, 2020.
- [27] S. Chakraborty, "An attempt-detection of covid-19 presence from chest x-ray scans using cnn & class activation maps," 2020.
- [28] M. S. Hossain, G. Muhammad, and N. Guizani, "Explainable ai and mass surveillance system-based healthcare framework to combat covid-19 like pandemics," *IEEE Network*, vol. 34, no. 4, pp. 126–132, 2020.
- [29] C. Ross and R. Robbins, "Debate flares over using ai to detect covid-19 in lung scans," *Stat*, vol. 30, p. 160, 2020.
- [30] X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, and P. Xie, "Sample-efficient deep learning for covid-19 diagnosis based on ct scans," *MedRxiv*, 2020.
- [31] X. Chen, L. Yao, T. Zhou, J. Dong, and Y. Zhang, "Momentum contrastive learning for few-shot covid-19 diagnosis from chest ct images," *Pattern Recognition*, vol. 113, p. 107826, 2021.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [33] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 2169–2178.
- [34] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "Netvlad: Cnn architecture for weakly supervised place recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5297–5307.
- [35] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghighi, R. Ball, K. Shpankaya *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 590–597.
- [36] A. K. Das, S. Kalam, C. Kumar, and D. Sinha, "Tlcov-an automated covid-19 screening model using transfer learning from chest x-ray images," *Chaos, Solitons & Fractals*, vol. 144, p. 110713, 2021.
- [37] B. Zhou, A. Khosla, A. Lapiedra, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE*

conference on computer vision and pattern recognition, 2016, pp. 2921–2929.