Non-invasive screening of COVID-19 cases using deep learning based multilevel fusion model with attention mechanism

Abstract

Automatic and fast screening for COVID-19 infections has become an urgent need of the ongoing pandemic. Various AI functionalities can effectively be used to screen (diagnose) and predict coronavirus disease 2019 (COVID-19) infections and propose timely response (remedial action) to minimize the spread and impact of the virus. Motivated by this, a deep learning-based CNN fusion model has been proposed in this research to accelerate analysis of chest X-ray (CXR) images in automatic screening of COVID-19 cases. We use different fine-tuned CNN models coupled with an attention network to extract salient features that are robust to overfitting and leverage a locally connected layer for a weighted contribution of these networks for final classification of COVID-19 cases. We validate the effectiveness of our proposed model using two publicly available datasets comprising images from normal, COVID-19 and other pneumonia infected categories. Experimental results demonstrate the effectiveness of the proposed fusion model in screening COVID-19 from CXR images even with a limited number of training samples.

Keywords: COVID-19, Chest X-Ray (CXR) imaging, Model fusion, Diagnosis, Deep learning, Attention Mechanism

1. Introduction

Due to the coronavirus disease 2019 (COVID-19), the world is going through unprecedented situations in recent human history with massive economic losses and global health crisis. The virus initially identified in December 2019 in the city of Wuhan in China has rapidly spread throughout the world within a very short period of time resulting in an ongoing pandemic. Since the outbreak it has affected over two hundred countries and territories across the globe with more than 20 million cases reported [15]. The outbreak was declared as Public Health Emergency of International Concern (PHEIC) by the World Health Organization (WHO) [16] on January 30, 2020. The virus is extremely contagious and primarily transmitted between people through close contact. Various common symptoms are found in the infected patients such as cough, fever, shortness of breath, fatigue, loss of smell, and pneumonia. The complications include pneumonia, acute respiratory distress syndrome, and other infections. Precise and on time diagnosis are being hampered due to undiscovered treatment, scarcity of resources, and harsh conditions of laboratory environment. This has increased the challenges to curb the spread of the virus.

Precise and timely diagnosis are being hampered due to undiscovered treatment, scarcity of resources, and harsh conditions of laboratory environment. This has increased the challenges to curb the spread of the virus. Accurate and speedy identification of suspected patients at the early phase may possibly play a critical role in timely quarantine and progressive cure. Thus, swift identification of potential infection by coronavirus is incredibly crucial for timely control of epidemic and public health welfare.

Identification of coronavirus infection is primarily done by nucleic acid test also called a PCR (polymerase chain reaction) test which examines for the existence of antibodies for an infection. However, results from recent studies show that this type of pathogenic laboratory testing though being a diagnostic gold standard suffers from limitations since it is time-consuming and produces high false negative cases [5]. Furthermore, deploying COVID-19 tests at a large scale is very expensive and is not affordable by many developing and underdeveloped countries. Hence, development of Artificial Intelligence (AI) based diagnosis and testing methods will be very beneficial. In favor of this, researchers are taking global initiatives to use AI as a potentially powerful tool to come up with cost-effective and fast diagnostic procedures to control the ongoing epidemic. The key research goals include COVID-19 transmission, its early diagnosis, development of effective treatment, and understanding its socio-economic impact.

Computer aided diagnosis (CAD) systems capable of processing chest X-ray (CXR) images and computed tomography (CT) scans along with state-of-the-art deep learning techniques could be very beneficial for the health professionals in diagnosing COVID-19 cases. Some studies [11, 21] in the literature have already demonstrated the effectiveness of using various deep learning techniques to identify positive COVID-19 cases from chest X-ray (CXR) images and computed tomography (CT) scans and to monitor the disease progress over time. Since deep learning algorithms generally require huge amount of training data to produce effective prediction results, the existing methods trained on limited training samples (due to the lack of large COVID-19 public dataset availability) are likely to suffer from model generalizability to new data. To alleviate this problem of data scarcity, researchers adopted various techniques such as data augmentation and generative adversarial network (GAN) [6, 7]. Nevertheless, these techniques are highly dependent on the appropriate selection of parameters. Various hand-tuned data augmentation technique suffers from over-fitting problem [6] whereas techniques related to generating images through GAN face challenges in emulating real patient data which leads to unanticipated bias during model testing [7].

Lately, there has been growing interest in developing fusion model from heterogenous

technologies [1] and also explicitly training deep fusion networks where multiple deep learning models (wide and deep) are fused together as a weighted combination of these networks to make superior prediction. Thus, fusion representation of deep learning models allows us to assimilate the strength of various individual networks to achieve promising performance. At the end, there will have a single fused multi-headed network that is meant to work reliably for the classification of unseen medical images. To this end, we have proposed a deep learning based fusion model with attention mechanism that exploits the benefit of a weighted combination of multiple deep CNN models in extracting salient features from the input CXR images that are fused to obtain robust classification of these images. We use three fine-tuned CNN models called ResNet50V2, VGG16, and InceptionV3 as base network models each one coupled with an attention network to extract feature embeddings from the input images. Furthermore, we use a locally connected layer for a weighted contribution of these networks in extracting salient features to be used in final classification. More specifically, we have used CXR images from two public datasets to train our fusion model.

The rest of the paper is organized as follows. Related work is presented in Section 2. Section 3 and 4 present method, dataset, and experiments with performance results and discussion. Lastly, Section 5 concludes the paper with future work.

2. Related Studies

Relevant to the proposed research, our literature study will largely contain existing research effort in the area of COVID-19 diagnosis using AI techniques. Deep learning which is a specialized form of machine learning in the domain of AI has shown great potential in medical image analysis during the last decade [9]. Substantial research has been conducted using deep learning in various medical fields such as disease prediction, diagnosis of pulmonary nodules, and classification of benign and malignant tumors and so on. According a recent study from the researchers at UN Global Pulse [10], it is shown that AI applications can exhibit human-like accuracy in detecting COVID-19 and offer faster and cheaper solutions in diagnosing the virus than standard test kits thus saving radiologists' valuable time. As part of this, researchers are primarily concentrating on techniques based on statistical learning for the detection of potential coronavirus infection from CXR images and computed tomography (CT scans). Some research initiatives in progress are provided below.

A relatively earlier (in the beginning of the outbreak) effort done by a group of researchers in Renmin University of Wuhan, China [11] proposed an AI model for the diagnosis of COVID-19 cases using CT scans. The model uses UNet++ [12] architecture for coronavirus detection using CT scan features and makes use of more than 40,000 images from 106 patients for model training.

Experimental results demonstrate that the radiologists' efforts in terms of time can substantially be decreased by using this model.

Xu et al. [13] presented a deep learning model to screen coronavirus disease from viral pneumonia (of type Influenza-A) and normal cases with pulmonary CT scans. They have first identified candidate infection regions using segmentation and the separated images are then classified using a classification model based on location attention. The model was trained and evaluated using 618 CT samples consisting of 219 COVID-19, 224 Influenza-A, and 175 healthy cases. The samples were collected from three hospitals in China that are designated for COVID-19 treatment. The model showed only a moderate level (86.7%) of accuracy score using the curated dataset. In a subsequent effort, Wang et al. [5] proposed a robust diagnostic model based on deep neural network which works based on graphical features generated from CT scans and saves crucial disease control time. They used a fine-tuned Inception [14] pre-trained model in their architecture and utilized a dataset containing CT scans of COVID-19 patients and patients with other non-COVID viral pneumonia.

Besides CT scans, several studies have used CXR images for the detection of COVID-19. Since it is relatively easier to find CXR images than CT scans especially in rural areas, they can be a viable alternative to CT images. Wang and Wong [18] proposed an AI system called COVID-Net to diagnose COVID-19 from chest X-ray images containing samples from healthy, COVID-19 and other pneumonia infected patients. The limitation of this study is that the authors trained and tested their model using an imbalanced dataset which contains very few (less than 100) COVID-19 images as opposed to about 16,000 images from healthy and other non-COVID pneumonia patients. Chakraborty [20] also developed a CXR based model using deep neural network that can achieve significant performance improvement even when the size of the dataset is limited. Nevertheless, the model lacks generalizability and needs fine-tuning to produce more stable results.

Another laudable effort made by researchers from Delft Imaging project [21] which developed an AI model for diagnosing COVID-19 from CXR images. Their model is called CAD4COVID which is built upon an existing AI model previously developed for diagnosing tuberculosis. It triages COVID-19 suspected patients. Hossain et al. [38] presented a healthcare framework based on 5G network that makes use of CXR and CT images for COVID-19. Some other work [22-24] have also shown potential in bringing deep learning techniques for the diagnosis of COVID-19 from CXR images using CNN and other pre-trained models such as ResNet [3]. Besides, a few research efforts [26, 27] have made contribution in interpreting their predictions by extracting critical features related to COVID-19 to gain deeper understanding.

Although a recent report [28] has shown the success of different Chinese hospitals in deploying machine learning based radiology technologies in combating COVID-19, radiologists have shown their concern [29] that the shortage of available data to train the AI diagnostic models is a major challenge. This is substantiated by the fact that large body of the AI models in the literature have used datasets containing limited COVID-19 images.

Narin et al. [31] presented a transfer learning-based approach to the classification of X-ray images into COVID-19 and normal categories. They have used three pre-trained models such as InceptionV3, ResNet50, and InceptionResNetV2 in their system achieved the highest 98% accuracy with ResNet50 for binary classification. However, the number of COVID-19 images in the curated dataset is only 50. In another effort, Oh et al. [30] have introduced a patch-based technique to train and fine-tune ResNet18 CNN model. They have used patches extracted from input CXR images to train the model. A majority voting strategy was used to obtain the final classification decision and the model attained a moderate accuracy of 88.9% in a multi-label classification scenario. Ozturk et al. [27] took a different approach than majority of the AI based detection model for COVID-19 where they proposed an objected detection-based technique. They have trained a DarkNet model for the classification of CXR images for COVID-19 detection. The experimental results showed a high level of accuracy (98.08 %) for binary classification. However, the model exhibited relatively poor performance for multi-class classification and attained an accuracy of 87.02% only.

3. Methodology

From the discussion in the literature study, we observe that the potential of fusion technology has not been fully leveraged for robust diagnosis of COVID-19 encounters. To this end, we have proposed a deep learning based fusion model that exploits the benefit of a weighted combination of multiple deep CNN models in extracting salient features from the input CXR images that are fused to obtain robust classification of these images into COVID-19, normal, and pneumonia categories. In this section, we start with the problem description from the perspective of a classification network based deep CNN fusion model. Then, we proceed to describe various components of our proposed system and the underlying technology to realize COVID-19 screening from the supplied CXR data.

3.1 Problem Formulation

There has been growing interest in explicitly training deep fusion networks where multiple deep learning models (wide and deep) are fused together as a weighted combination of these networks to make superior prediction. Thus, fusion representation of deep learning models allows us to assimilate the strength of various individual networks to achieve promising performance. At the end, we will have a single fused multi-headed network that is meant to work reliably for the classification of unseen CXR images. An illustration of our fusion problem is given in Fig. 1.

Given a CXR image dataset, $\boldsymbol{D} = \{(\boldsymbol{x_n}, \boldsymbol{y_n}),\ \boldsymbol{n} = \boldsymbol{1}, \ldots, \boldsymbol{N}\}$, where for the nth instance of the input image, $\boldsymbol{y_n}$ denotes its class label and its attribute values are represented by $\boldsymbol{x_n}$. In addition, we consider M total base deep learning models where *i*-th model is denoted by Θi, for i = 1, …, M. For each image, $\boldsymbol{x_n}$, in $\boldsymbol{D}$, let $\boldsymbol{z_n}^{(i)}$is the feature vector generated by model Θi and $\boldsymbol{z_n}^{(i)}$: $\mathbf{R}^{f_i} \rightarrow \mathbf{R}^{f}$ $(\boldsymbol{i} = \boldsymbol{1}, \ldots, \boldsymbol{M})$ over the complete fusion domain, *f*. Now, we obtain a weighted combination of the feature vectors generated from all M models:

$$\boldsymbol{z} = [\, \boldsymbol{w_1 z_n}^{(1)}, \boldsymbol{w_2 z_n}^{(2)}, \ldots, \boldsymbol{w_i z_n}^{(i)}] \qquad (1)$$

Now, given the combined or fused feature vector output from all the based models, $\boldsymbol{z} \in \mathbf{R}^{f}$or $\boldsymbol{z} \in \mathbf{R}^{\sum f_i}$, final classification results are obtained by another deep learning model, *F*:

$$\boldsymbol{p} = \boldsymbol{F}(\boldsymbol{z}) \quad \boldsymbol{where\ F}: \mathbf{R}^{f} \rightarrow \mathbf{R}^{C}\ \boldsymbol{with\ C\ class\ labels} \qquad (2)$$

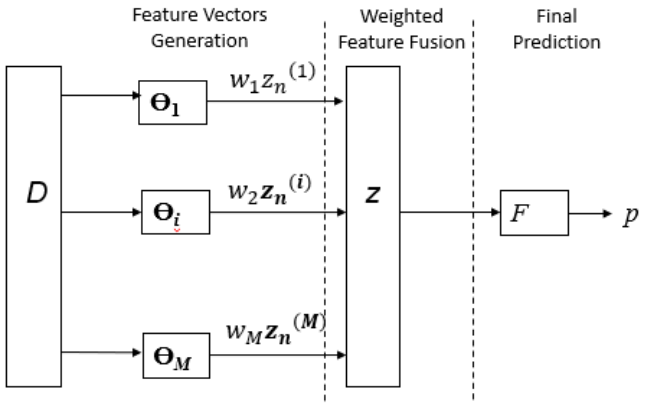


Fig. 1. Illustration of fusion problem

3.2 Proposed System

We present a schematic diagram of our proposed system (as shown in Fig. 2) for non-invasive screening of COVID-19 cases using weighted fusion of deep CNN models. First, feature extraction from input CXR images are performed using three widely used deep

convolutional networks, namely VGG-16 [8], InceptionV3 [2], and ResNet50V2 [3]. Then, we perform feature fusion from the weighted combination of the features extracted from these networks. Multi-label classification on the fused features are conducted by another convolution and a fully connected layer to realize coronavirus infected cases. Finally, model interpretation and feature representation are demonstrated through two different visualization techniques. A detailed description of the system is given in the following subsections.

3.2.1 Feature Extraction by Base CNN Models

In our system, we have used the above mentioned pre-trained (on ImageNet dataset [33]) CNN models along with their weights from convolutional layers for feature extraction from input CXR images. The VGG-16 network architecture introduced by Simonyan and Zisserman [8] represents a simple network using only 3 x 3 convolutional layers stacked together in increasing depth. In addition, volume size is reduced using max pooling layers. As the name suggests it has a total of 16 weight layers with last two fully connected layers, each with 4,096 nodes which are then followed by a softmax classifier layer. It has won the ILSVRC-2012 classification challenge on the ImageNet dataset containing over 14 million images across 20,000 categories by outperforming its pioneer AlexNet [4] with the help of smaller sized

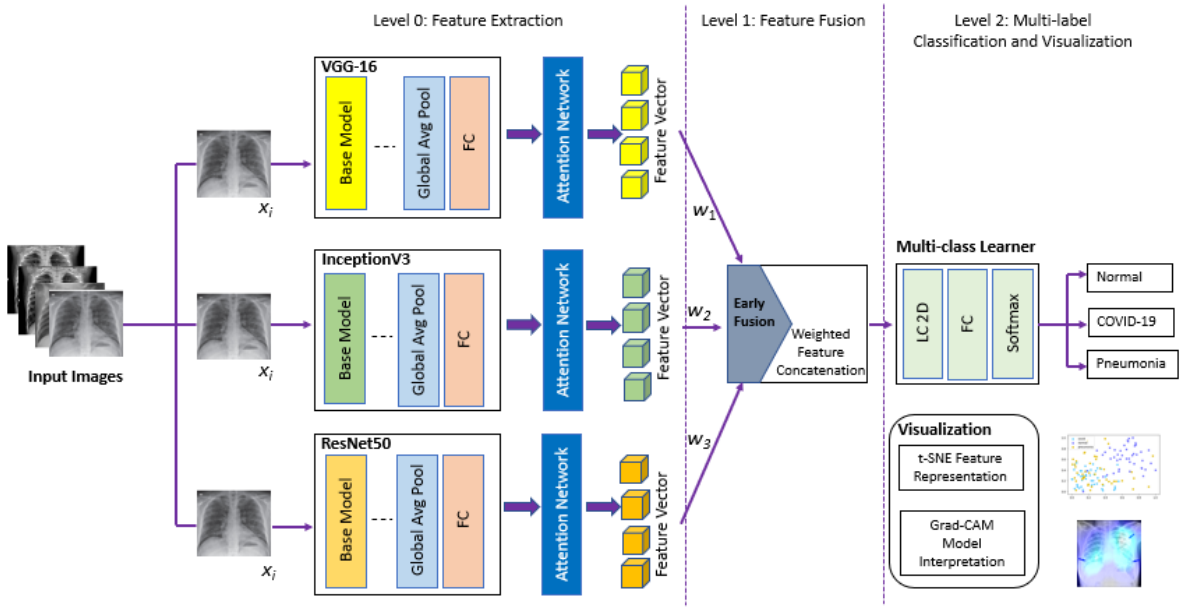


Fig. 2. Schematic diagram for COVID-19 diagnosis from CXR images using multi-level fusion network (Zooming may be required for enhanced view)

filters. Traditional sequential deeper networks such as VGG-16 suffer from vanishing gradient

problem where accuracy becomes saturated and drops abruptly with the increasing depth. ResNet architecture addresses this problem by skipping through less important layers with the help of residual modules while training the network with standard SGD optimizer. The version of ResNet (ResNet50V2) used in this study contains 50 weight layers demonstrating significant decrease in size of the model and the number of FLOPs. The third CNN model used in our fusion network is Inception introduced by Szegedy et al. [2]. It appears as a micro-architecture with the goal to use a multi-level feature extractor by computing convolutions of different sizes (1×1, 3×3, and 5×5) within the same module of the network. Outputs from these filters are stacked on top of each other along the channel dimension and then fed into the next layer. The original manifestation of this network was called GoogleNet, but the subsequent incarnations are simply named as Inception with appropriate versions. In this study, we have used Inception V3 [32] which is an updated version of the Inception module achieving further improve in ImageNet classification accuracy. InceptionV3 is characterized by its weights which are less than both ResNet and VGG.

3.2.2 Attention Network

We have used the base CNN models in our fusion network where each of them is coupled with a visual mechanism to aid the CNN models to focus on parts of the input image rather than concentrating on every part of it with the same importance. Inspired by the work presented by Jetley et al. [35] in which they have used a soft attention module in CNN architecture for improving multilabel classification accuracy, we have added an attention layer after the last fully connected layer of each base CNN model as shown in Fig. 2. The feature input from the base model is fed to a batch normalization (BN) layer which standardizes the input for each mini-batch with the aim to achieve a steady learning process and to substantially reduce the number of training steps. We employ a number of 1×1 convolutions to downsample a very large number of feature maps. In the attention part, we modify the global average pooling (GAP) layer which appears to be very simplistic to highlight the regions of interest since some regions are more pertinent than others. To achieve this, we develop an attention module that alters the pixel values in GAP layer ahead of pooling and then rescales the output in proportion with the pixel counts through a Lambda layer. The model could be considered as a weighted version of global average pooling. The detailed architecture of the attention layer is given in Fig. 3.

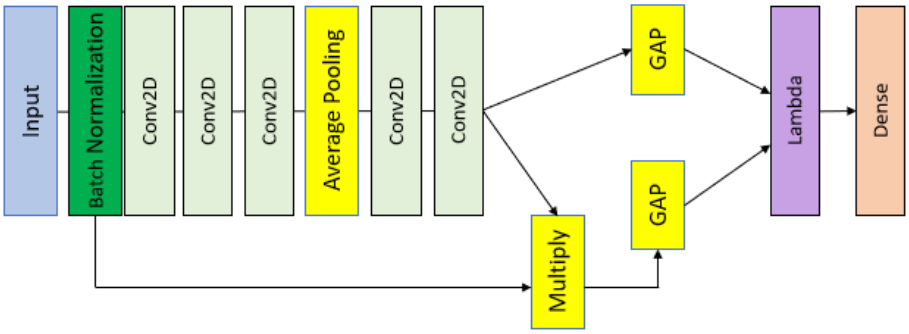


Fig. 3. Detailed architecture of the proposed attention mechanism used in the feature extraction level (input is propagated from the fully connected layer of preceding base CNN model)

3.2.3 Weighted Feature Fusion

As depicted in Fig. 2, we carry out fusion of feature vectors extracted from the base CNN models coupled with attention layers. This feature fusion can be considered as a joint representation which refers to a concatenation of features obtained from individual CNN models. This process is also known as "early fusion". The joint feature representation is then passed through a sequence of convolution and dense layers before final prediction. Thus, our fusion model can be treated as end-to-end trainable and capable of both learning feature representation and performing multi-label classification through prediction. This early fusion and late prediction strategy requires fewer parameters as compared to its early prediction and late fusion counterpart due to the use of reduced number of hidden convolutional and dense layers. More importantly, the fused features can provide a richer representation of images compared with the individual CNN features. We introduce weighted fusion of features in our system largely motivated by the feature fusion technique used in [9] for diabetic fundus image classification.

Our system exploits the concept of a locally connected (LC) layer to fuse the independent CNN models. In essence, an LC layer is very similar to convolutional layer but without using any sharing of weights. In a convolutional layer, the same filter weights are shared by all the pixel positions whereas a LC layer uses non-sharing filters over spatial arrangements to learn distinct weights in each local field [36]. For instance, researchers in Facebook built DeepFace [37] employing LC layers with non-sharing filters to obtain more discriminating facial features. In a slightly different context, we use LC layers to learn adaptive weights indicating the importance for different base models and procure an enhanced fused representation of features. To demonstrate the benefit of locally connected fusion technique used in our system, Fig. 4

shows a comparison with two other generally used fusion types namely, fully connected fusion without learning any weights and as mentioned already convolutional fusion with sharing weights.

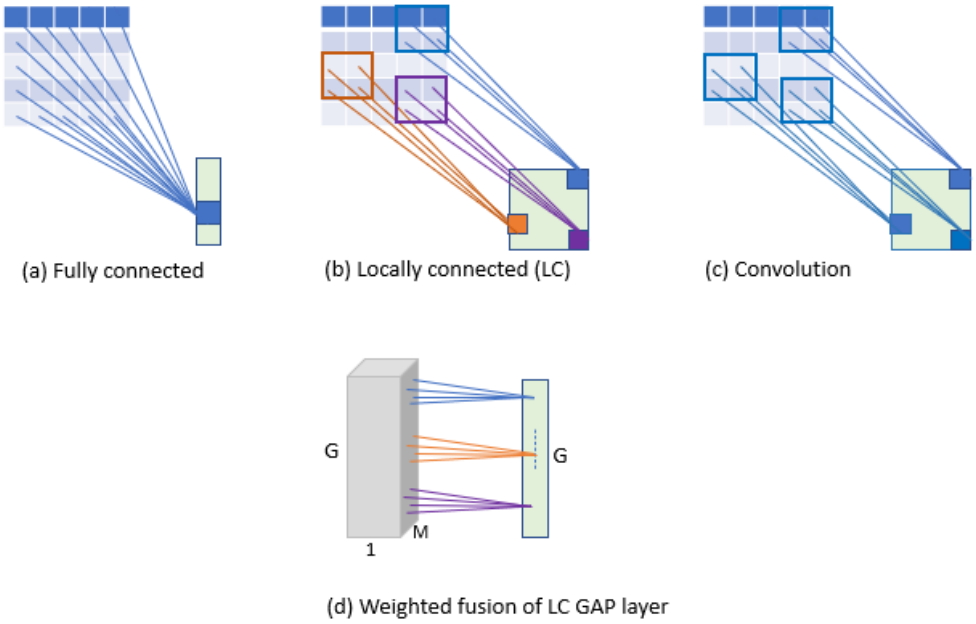


Fig. 4 Various types of fusion layers. (a) fully connected neuron without any weights, (b) locally connected neurons mapped to 2×2 regions, (c) convolutional neurons mapped to 2×2 regions, (d) locally connected GAP layer with non-sharing weights

As part of the fusion process, first, we combine all GAP layer features from M (i.e., 3) base CNN models coupled with attention layers and form a fusion GAP layer (G_{fusion}) as shown in Fig. 4 (d) of size $1\times G \times M$. For instance, the i-th feature map of G_{fusion} can be denoted as $z^{(i)}$ where $\mathbf{1} \leq \boldsymbol{i} \leq \boldsymbol{M}$. Then, we convolve a LC layer having G filters with no weight sharing over G_{fusion} where each filter has a kernel of size $1\times 1 \times M$. Thus, the LC layer is capable of producing an enhanced fused feature by learning adaptive weights for various feature maps of G_{fusion} representing the importance of individual base CNN model. Finally, we have a fused feature set called $z^{(f)}$ after convolved by LC layer which is a one-dimensional feature vector of length G. We can represent the n-th element of $z^{(f)}$ as

$$\boldsymbol{z_n}^{(f)} = \boldsymbol{\varphi}\left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{M}} \boldsymbol{w_{n,i}}^{(f)}.\boldsymbol{z_n}^{(i)} + \boldsymbol{b_n}^{(f)}\right), \qquad \mathbf{1} \leq \boldsymbol{n} \leq \boldsymbol{G} \tag{3}$$

where $\boldsymbol{\varphi}$ represents the activation function used such as ReLU. In addition, weights and bias for fusing the n-th elements of features in G_{fusion} which are from different base CNN models are represented by $\boldsymbol{w_{n,i}}^{(f)}$ and $\boldsymbol{b_n}^{(f)}$, respectively. Thus, we attain feature fusion with adaptive

weights without making any manual tuning at the cost of additional $G \times (M+1)$ parameters required by LC layer.

3.2.4 Training of Fusion Model

Training of our fusion model entails both forward pass and backward propagation similar to a traditional CNN. Given a set of CXR images, $\boldsymbol{\chi} = \{\boldsymbol{x_i} \in \boldsymbol{D}\}$ and a set of corresponding labels $\boldsymbol{Y} = \{\boldsymbol{y_i} \in \boldsymbol{L}\}$ and a training set, T containing N images $\{\boldsymbol{x_i}, \boldsymbol{y_i}\}$ where $\boldsymbol{x_i}$ and $\boldsymbol{y_i}$ represent i-th input image and its true class label. Additionally, let $\boldsymbol{W}$ denote the set of all parameters learned by the fusion network including the weights of locally connected layer for fusion. The objective is to minimize the cost for total loss, L where $\boldsymbol{p}(\boldsymbol{x_i})$ denotes the predicted label of the input image, $\boldsymbol{x_i}$:

$$\underbrace{\boldsymbol{argmin}}_{\boldsymbol{w}} \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{L}(\boldsymbol{p}(\boldsymbol{x_i}), \boldsymbol{y_i}) \qquad (\mathbf{4})$$

To compute L we have used the softmax loss function. During the backpropagation, a chain rule is used to recursively calculate the derivatives (partial) of L based on a weight to minimize it. Since the main contributors of the fusion network are the base CNN models we will include them in the detail computations of the partial derivates. We start with computing L's gradient with regard to the outputs obtained from the base models. For instance, the gradient of the loss, L with respect to the output from i-th base model, $z^{(i)}$) can be computed as below:

$$\frac{\boldsymbol{\partial L}}{\boldsymbol{\partial z^{(i)}}} = \frac{\boldsymbol{\partial L}}{\boldsymbol{\partial z^{(f)}}} * \frac{\boldsymbol{\partial z^{(f)}}}{\boldsymbol{\partial z^{(i)}}}, \qquad \mathbf{1} \leq \boldsymbol{i} \leq \boldsymbol{M} \qquad (\mathbf{5})$$

Furthermore, the gradient of L is calculated with respect to the inputs of the base models. Given, we have M different base models, let $h^{(i)}$ be the input to the i-th base model. Then, the gradient of L with respect to the input $h^{(i)}$ can be formulated as:

$$\frac{\boldsymbol{\partial L}}{\boldsymbol{\partial h^{(i)}}} = \frac{\boldsymbol{\partial L}}{\boldsymbol{\partial z^{(i)}}} * \frac{\boldsymbol{\partial z^{(i)}}}{\boldsymbol{\partial h^{(i)}}}, \qquad \mathbf{1} \leq \boldsymbol{i} \leq \boldsymbol{M} \qquad (\mathbf{6})$$

As commonly used, we have leveraged Adam optimizer with mini-batch to train the entire fusion model. In summary, our fusion model works as a single deep learning model which is multi-headed to accept identical input images from training data. Intermediate feature vectors generated from the base models coupled with attention networks are fused and fed through a locally connected layer for weighted contribution of features from distinct base CNN models in final classification of the input images. The algorithm below summarizes the complete fusion

mechanism:

Algorithm 1: Fusion of CNN models in classification of CXR images

Input: Training data $T = \{x_i, y_i\}_{1 \le i \le N}$, Test data T_{test}, base CNN models

Output: Classification results from a fusion model

T_{train}, $T_{validation}$ = split dataset, T

for k = 1 to *k-fold* **do**

Generate feature vectors from base CNN models (M) coupled with attention network:

for m = 1 to M **do**

Extract feature vectors $z^{(m)}$ based on T_{train}

end for

$z^{(f)}$ = weighted concatenation([$z^{(1)}$, $z^{(2)}$, …, $z^{(M)}$] with fusion weights $w_1, w_2, \ldots, w_M$

Construct a new dataset, G_{fusion} containing the weighted features and target label:

for i = 1 to N **do**

$G_{fusion} = \{z_i^{(f)}, y_i\}$

end for

Learn a final classifier, $F^{(k)}$ based on the newly constructed dataset, G_{fusion}

Validate $F^{(k)}$ with $T_{validation}$

end for

Classification with test data:

results = classify (F, T_{test})

return results

4. Experiments and Results Analysis

To demonstrate the effectiveness of fusion model in screening COVID-19 cases, we extensively evaluate and compare the performance of our proposed model with finetuned pre-trained CNN models namely ResNet50V2, VGG-16, and InceptionV3 using two publicly available CXR datasets. In the subsequent sections, we will present the curated dataset with preprocessing steps, experimental settings, and results with discussion.

4.1. Dataset Description and Preprocessing

In this study, we use COVID-19 patients' CXR images that are acquired from an open source database created by Dr. Joseph Cohen [17] in his GitHub repository. Dr. Joseph Cohen is continuously uploading CXR pictures of COVID-19 patients having acute respiratory distress syndrome (ARDS), COVID-19, pneumonia, Middle East respiratory syndrome (MERS), and severe acute respiratory syndrome (SARS). At present, there are about 616 CXR images of COVID-19 patients. To create a classifier from CXR images it is also necessary to have related

CXR images of patients who do not have COVID-19. Fortunately, Kaggle [38] has a repository of CXR images of pneumonia and healthy patients. Hence, we have considered both image sources as a dataset to build, train, and test the proposed model. In general, the CXR images of confirmed COVID-19 cases show various shapes of "pure ground glass" also referred to as hazy lung opacity during the disease development. We have gathered 616 CXR images from all three categories such as normal, non-COVID pneumonia, and COVID-19 positive to create the curated dataset and thus a total 1848 CXR images constitute the final dataset. To assess the performance of the used models, we have used a 5-fold cross validation technique where the entire dataset is divided into 5 equal parts in patient level. Table 1 shows the distribution of images at a ratio of 60:20:20 in each fold for training, validation, and test datasets. During training we use the training and validation sets while the hold-out test set is used for the performance evaluation of the models.

Table 1. Number of samples in training, validation, and test sets from various CXR image categories

Class	**No. of Samples**			
	Training (60%)	Validation (20%)	Testing (20%)	Total (100%)
Normal	368	124	124	616
Pneumonia	368	124	124	616
COVID-19	368	124	124	616

4.1.1 Pre-processing

Due to the fact that the images in the dataset were collected from different locations with various clinical settings, the intensity and quality of images vary considerably. Nevertheless, we avoid extensive pre-processing of our CXR images in the dataset to gain improved generalization ability of our proposed fusion network model. This in turn makes our model further robust to artifacts and noises present in the images while extracting feature embeddings from the input images. Thus, we only used few standard pre-processing tasks including image resizing, normalization, image augmentation to optimize the model training method. The size of the CXR images in the dataset varies from 365 × 465 to 1125 × 859 pixels. Hence, we re-scale all images to a size of 224 × 224 pixels to get a consistent image dimension for the entire dataset. Additionally, we perform intensity normalization also called scaling which is an important pre-processing task to expedite model convergence by eliminating feature biases and

attaining a uniform distribution for the dataset. We convert the image pixel values from [0, 255] to [0, 1] to obtain a standard normal distribution by using min-max normalization technique. Finally, we apply image augmentation to tackle the problem of small dataset and to increase training efficiency while preventing the models from overfitting. A summary of augmentation features used for preparing the training dataset is given Table 2.

4.2. Experimental Settings

The fine-tuned CNN base models and the proposed fusion model are implemented using TensorFlow. More specifically, we have used Keras functional API to build the fusion model which can handle models with shared layers, non-liner topology, and multiple inputs or outputs. In addition, we use a special form of convolutional layer API from Keras called 2D "locally connected layer" with unshared weights for the implementation of weighted concatenation of base models. We use Google Colab notebook environment for model training and testing which provides free GPU access. It currently offers NVIDIA Tesla P100 GPU with 16GB RAM and is equipped with Python 3.x packages and Keras API with backend TensorFlow.

Table 2: Parameters and functions used for model training

Training Parameters	**Values/Types**
No. of Epochs	50
Batch Size	8
Optimizer	Adam
Loss Function	Categorial
Initial Learning Rate	0.001
Rotation Range	15
Shear Range	10%
Zoom Range	10%
Width and Height Shifting	10%
Horizontal Flip	Yes
Fill Mode	Nearest
Re-scaling	1/255

Towards the end of the fusion model after the locally connected layer, we add a dense layer with 256 neurons with ReLU activation function. Finally, we add a dense layer with a softmax unit to generate the classification scores. We have used categorical cross-entropy loss functions for model learning. In addition, Adam optimizer is used for model training and optimizing with

an initial learning rate of 0.001. Subsequently, the learning rate is updated with a decay computed from the ratio of initial learning rate and the number of training epochs. Furthermore, we have used a callback from Keras called *ModelCheckpoint* to monitor the performance metrics and periodically save the model based on some monitoring criteria such as validation loss or accuracy.

We evaluate our model using the following six metrics: accuracy, recall, precision, specificity, F1-score, and AUC (Area Under Curve). Here, accuracy calculates the proportion of predictions that precisely matches with the real values. Precision also called positive predictive value (PPV) is the fraction of true positive cases over all positive predictions. Recall or sensitivity also called true positive rate (TPR) which is very critical in medical applications refers to the fraction of all COVID-19 positive incidents that are accurately categorized as positive. Specificity also called true negative rate (TNR) refers to the proportion of all negative cases that are accurately categorized as negative. For example, the percentage of healthy people who are correctly diagnosed as not carrying the virus. F1-score denotes the harmonic mean which is calculated from precision and recall by taking their weighted average. Finally, AUC refers to the area under the receive operating characteristic (ROC) curve that demonstrates how TPR increases with the decrease in false positive rate (FPR).

4.3. Evaluation Results and Discussion

To evaluate the efficacy of our proposed fusion model in diagnosing COVID-19 cases we took the following approach. We start by evaluating our model quantitatively and a comparison of performance results is also done with different fine-tuned base CNN models. Finally, we present the results from a qualitative perspective using feature representation and model interpretation.

Table 3. Classification results for ResNet50V2, VGG-16, InceptionV3, and the Fusion model on holdout test dataset

Model	**Accuracy**	**Precision**	**Recall**	**Specificity**	**F1-score**	**AUC**
ResNet50V2	95.49	96.85	99.19	98.27	98.00	95.94
VGG-16	92.70	97.50	94.35	98.69	95.89	95.73
InceptionV3	92.97	97.60	98.39	98.67	97.99	94.72
Fusion Model	96.75	98.38	98.38	99.15	98.38	97.23

Table 4. Performance results for each class using all evaluated models on test dataset

Model	Class	Accuracy	Precision	Recall	F1-score
ResNet50V2	COVID-19	99.19	97.0	99.0	98.0
	Normal	94.30	94.0	94.0	94.0
	Pneumonia	90.24	93.0	90.0	92.0
VGG-16	COVID-19	94.35	97.0	94.0	96.0
	Normal	94.30	91.0	94.0	92.0
	Pneumonia	89.43	90.0	89.0	90.0
InceptionV3	COVID-19	98.38	98.0	98.0	98.0
	Normal	93.49	90.0	93.0	92.0
	Pneumonia	86.99	91.0	87.0	89.0
Fusion Model	COVID-19	98.38	98.0	98.0	98.0
	Normal	98.37	94.0	98.0	96.0
	Pneumonia	93.49	98.0	93.0	95.0

We carry out a number of experiments using holdout test dataset to compare the performance results of various base CNN models with our fusion model. In addition to overall performance (as shown in Table 3), class-wise performance is also evaluated using the same metrics (shown in Table 4). It is noticed that the fusion model shows consistently better results than the base models in performance metrics including accuracy, precision, specificity, and area under curve (AUC). The fusion model produces impressive values of accuracy (96.75%) and specificity (99.15%) on holdout test dataset which are deemed to be very critical performance estimates for applications in medical settings. It clearly shows the benefit of using weighted combination of features extracted from different base models in the fusion network in classifying COVID-19 patients over other contemporary methods. The specificity result indicates that our model can accurately identify negative COVID-19 cases with an accuracy of 99.15% from all the cases that are negative. However, among the base models ResNet50V2 outperforms other CNN models by accuracy and sensitivity. It is worthwhile to mention that ResNet50V2 shows slightly better sensitivity or recall value (99.19%) than the fusion model (98.38%).

The impact of model fusion over individual base models is also elucidated by class-specific results in Table 4. All the base models tend to show relatively poor performance in classifying pneumonia images while show moderate performance in identifying healthy cases. As per classifying the COVID-19 positive subjects, ResNet50V2 shows superior performance in terms

of accuracy and sensitivity. Remarkably, the fusion model exploits the benefit of a weighted

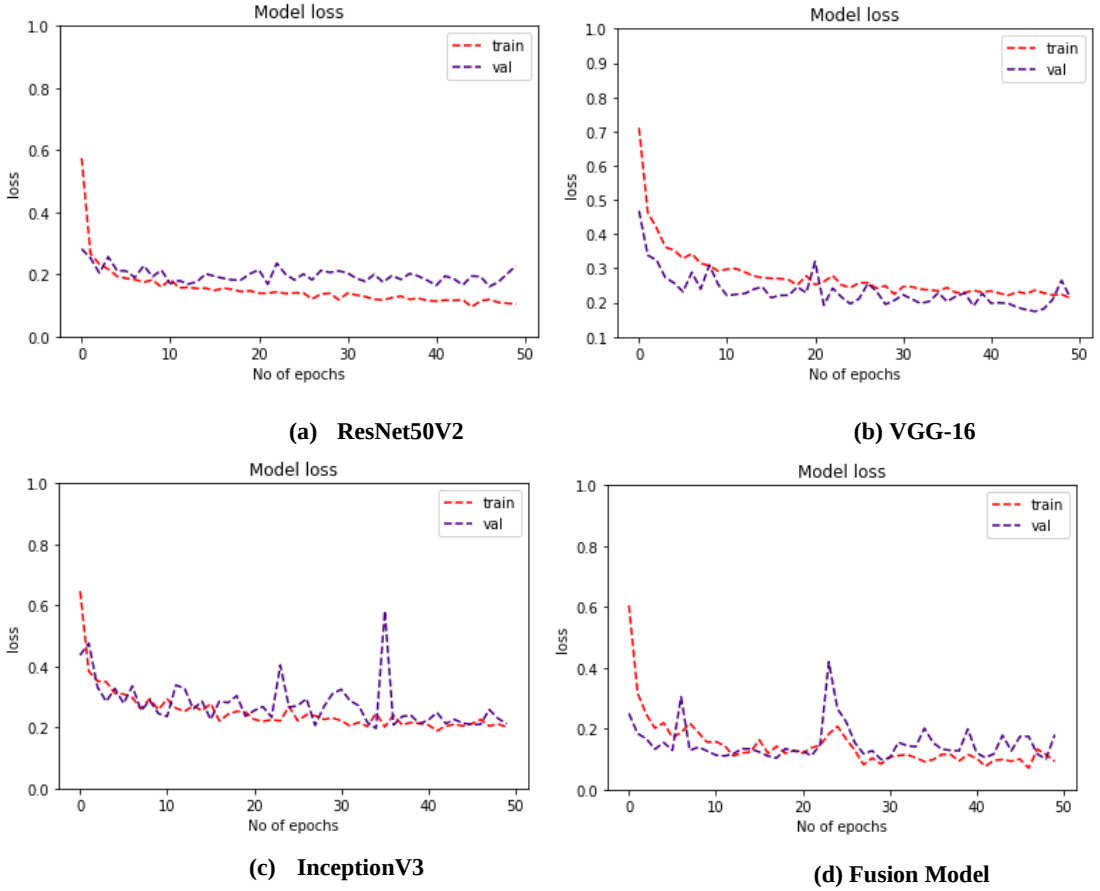
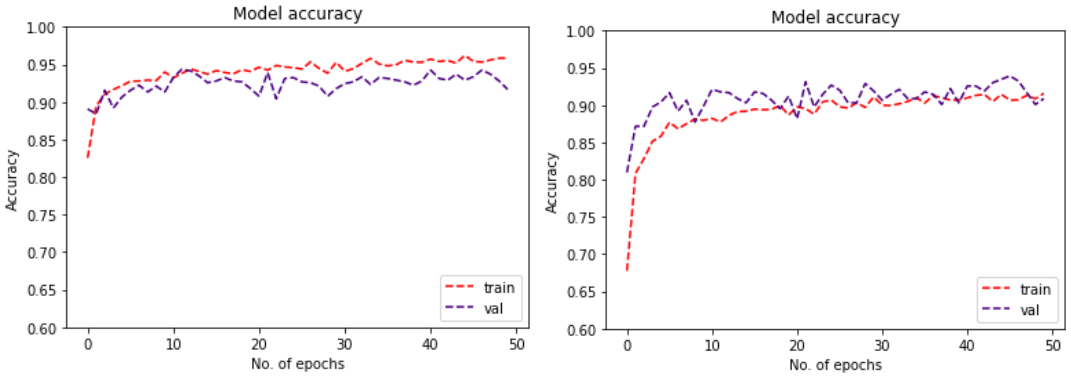


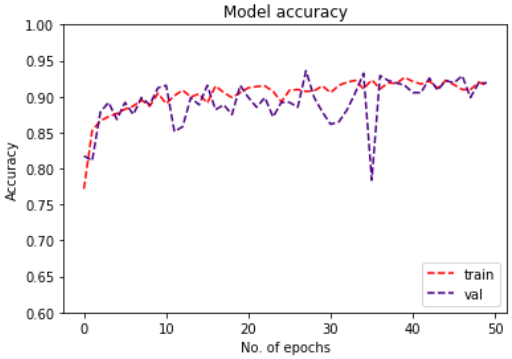
Fig. 5. Training and validation loss for various models using CXR images from COVID-19, normal, and other pneumonia cases

combination of multiple base CNN models by preserving the strength of ResNet50V2 to compensate for the weaknesses of VGG-16 and InceptionV3 in improving the performance results of all metrics for all three classes. This result is hypothetically meaningful since accurately classifying CXR images for all three subject groups (COVID-19, normal, pneumonia) are very important for an effective diagnostic tool. All the models show a modest learning progression during the training period by incurring a stable decline in both types of losses. Moreover, the fusion model training and validation (in Fig 5 and 6) seem to better converge using the same number of training epochs like other models. Even though the curated dataset contains limited samples, the learning curves show that the models are not prone to overfitting. This is essentially attributed to the fusion model's generalizability, data augmentation applied to the training set, and the use of regularization technique such as dropout applied to the fusion model.

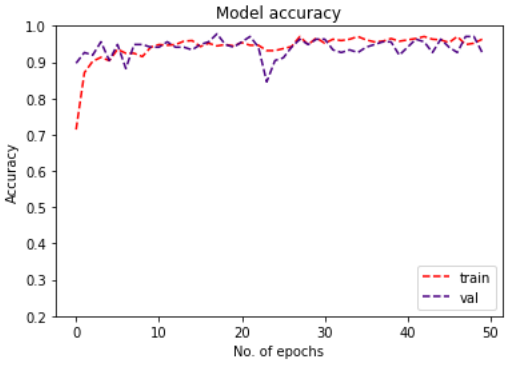


(a) ResNet50V2

(b) VGG-16



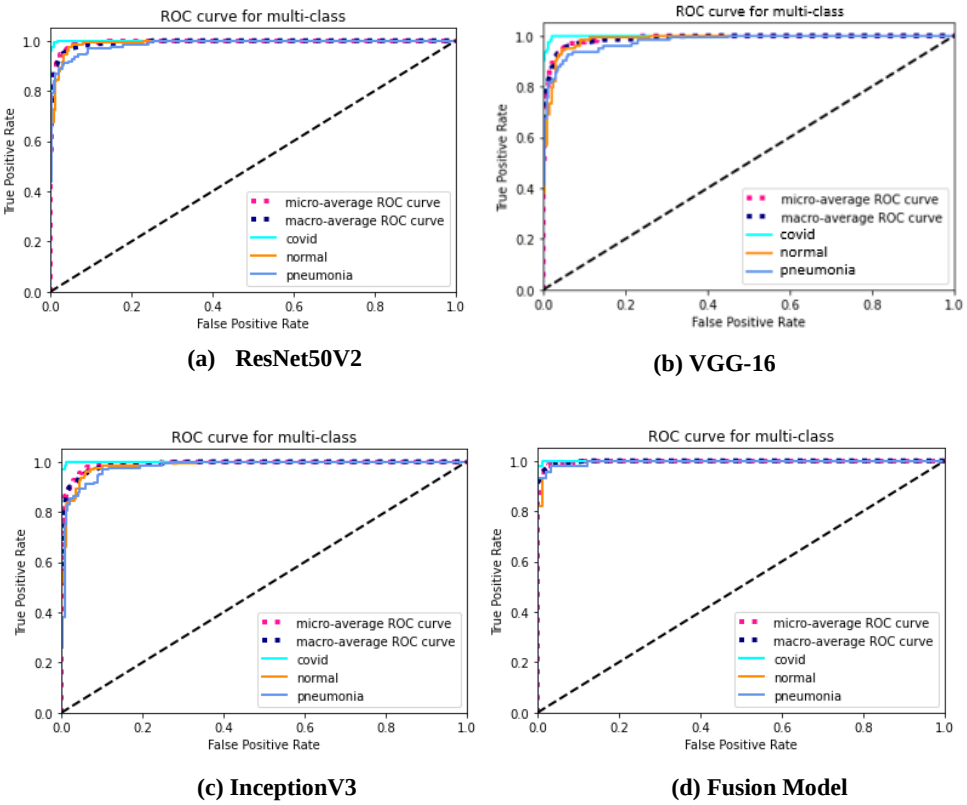
(c) InceptionV3



(d) Fusion Model

Fig. 6. Training and validation accuracy for various models using CXR images from COVID-19, normal, and other pneumonia cases

For a deeper understanding of the performances of the evaluated models, receiver operating characteristic (ROC) curve and confusion matrix for each model are shown in Fig 7 and 8, respectively. A ROC curve plots true positive rate (TPR) against false positive rate (FPR). The ROC curve (in Fig 7) shows the stability of the fusion model where we observe that the area under the curve for all classes is very similar and the model achieve an average AUC score (as shown in Table 3) of 0.9723 for all classes. The other base CNN models show superior discrimination abilities for COVID-19 classes but exhibit poor classification performance for non-COVID classes. This is substantiated by the fact that the curve for COVID-19 is slightly higher than the other classes. We observe from the confusion matrix (in Fig. 8 and Table 5) that the fusion model generates very less false positive (FP) and false negative (FN) cases (2 and 2 respectively) for COVID-19 classes compared to other classes (e.g., 10 cases of FP and FN for both normal and pneumonia cases) and other base models as well. This is vital for our model to minimize wrong diagnosis. The low value of FP indicates that the number of cases



(a) ResNet50V2 **(b) VGG-16**

(c) InceptionV3 **(d) Fusion Model**

Fig. 7. ROC curves for multiple classes (COVID-19, normal, and other pneumonia) using various models

that are misidentified as COVID-19 positive is less and it positively contributes to increased values of precision and specificity. To limit the count of FP cases is very important to reduce unwanted financial liabilities on health providers. Likewise, the reduced count of FN cases implies that the number of COVID-19 cases that are missed is less and it contributes to an increased value of sensitivity. In reality, it is very vital to keep the count of FN cases to low ensure that the model does not identify somebody contracted with the virus as healthy and thus the patient's line of treatment is hampered. On the other hand, base CNN model such as ResNet50V2 produces very low FN count for COVID-19 cases but generates higher FP and FN counts for other classes of images. The proposed fusion model presents a balanced classification performance by reducing FP and FN counts for all classes of images in the dataset. Based on these obtained results, the proposed fusion model appears to be the best performing among all evaluated models.

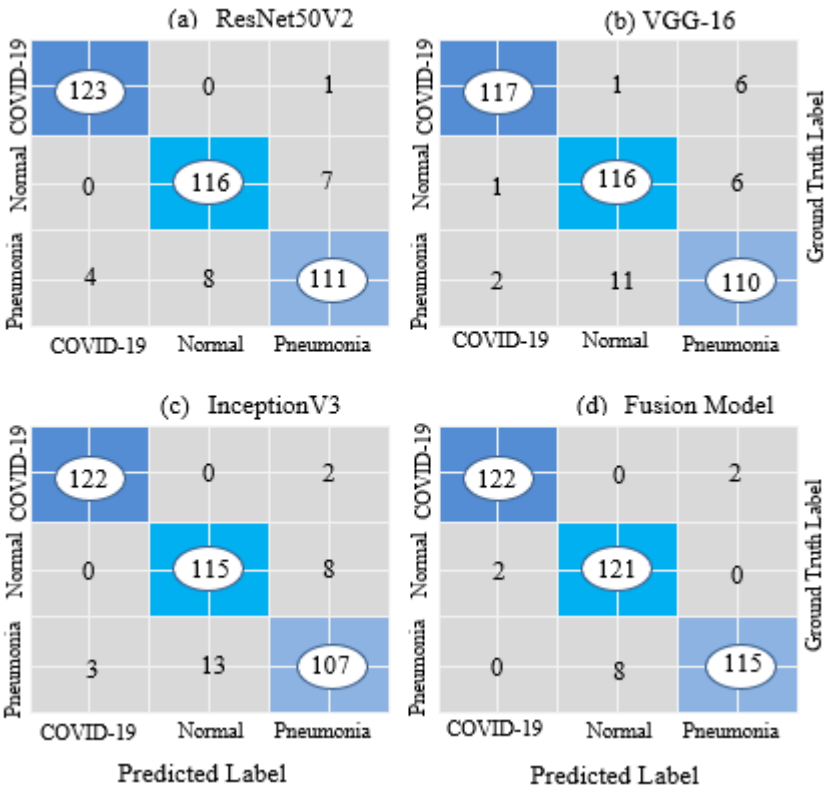


Fig. 8. Confusion matrix for various evaluated models with test dataset containing CXR images from all three classes

Table 5. Confusion matrix for COVID-19 class using test dataset

Model	TP	FP	TN	FN
ResNet50V2	123	4	227	1
VGG-16	117	3	226	7
InceptionV3	122	3	222	2
Fusion Model	122	2	236	2

4.4 Feature Representation and Model Interpretation

As part of the qualitative analysis, we investigate how well the features are distributed in the feature space to understand the class separability. Since convolutional layers produce high dimensional output, we need to adopt a dimensionality reduction technique to visualize them in 2D space. To achieve this, we use t-SNE (t-Distributed Stochastic Neighbor Embedding) [39] which is a popular technique for exploring and reducing high dimensional data. Unlike PCA (Principal Component Analysis), t-SNE is a non-linear technique to generate a low-dimensional mapping by using correlation between data points.

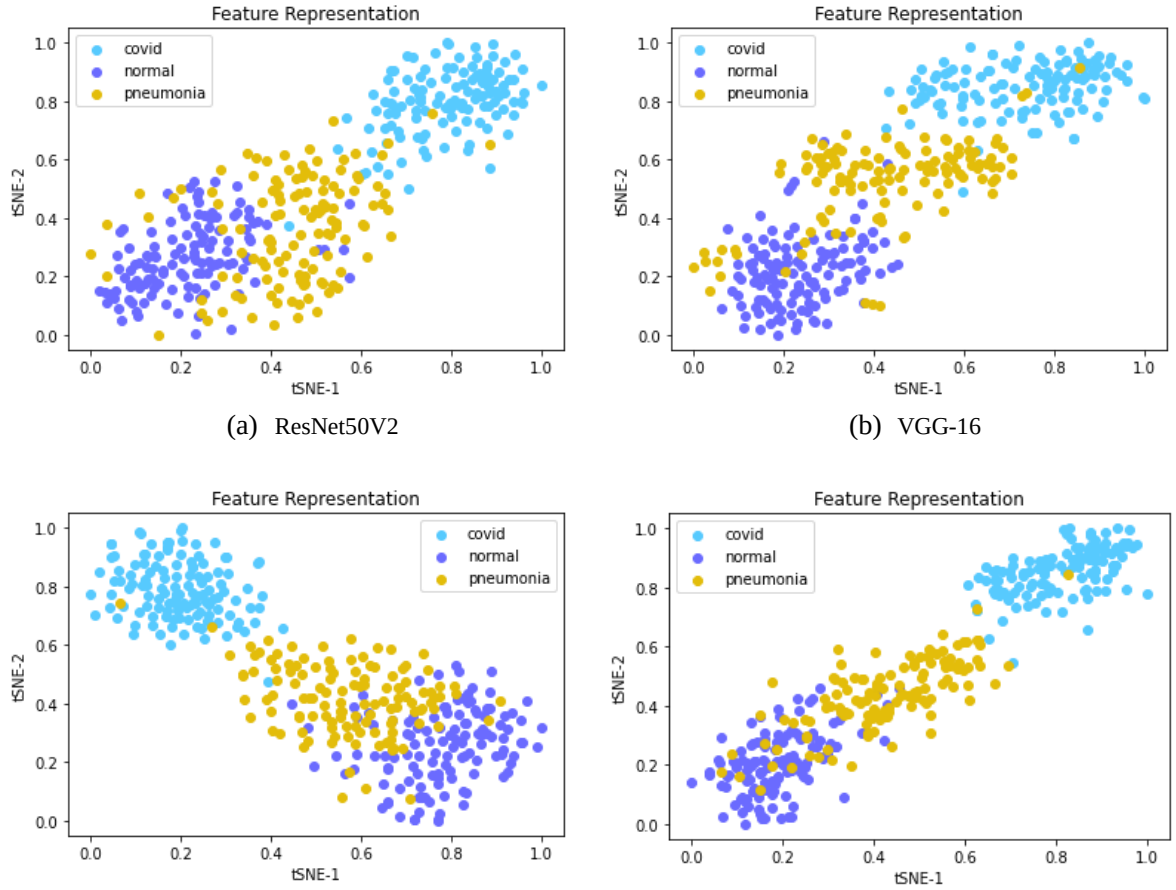


Fig. 9. Feature representation based on predicted labels with t-SNE plot for multilabel classification using fusion and other CNN models. (a) ResNet50V2, (b) VGG-16, (c) InceptionV3, (d) Fusion Model

Fig 9 shows the visualization of three class balanced features extracted from various base CNN models and the proposed fusion model using CXR images from holdout test dataset. It is observed that the fusion model exhibits better feature representation compared to other models. The t-SNE of the fusion model seems to be well plotted in a relatively compact space and shows a clear separation of COVID-19 classes compared to that of normal and pneumonia classes. However, similar to other models, the fusion model shows an area of overlap between normal and pneumonia classes.

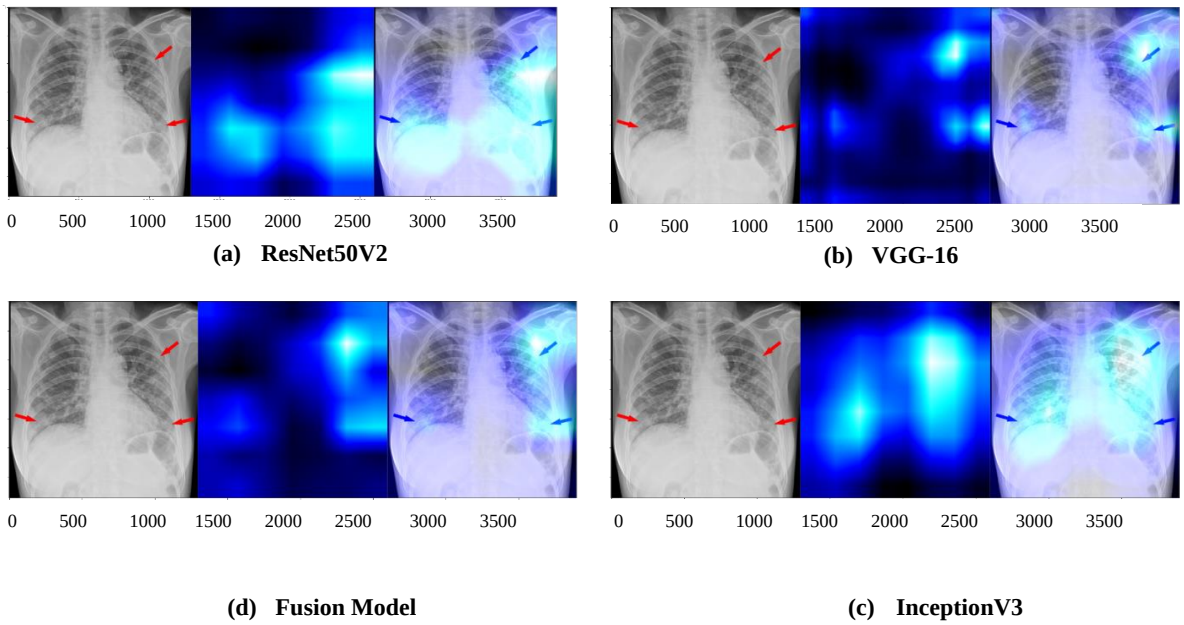


Fig. 10. Illustration of heatmap visualization with GradCAM for a COVID-19 patient. Original image (focus areas marked with red arrows) followed by a heatmap and an overlay of heatmap onto original image. (a) ResNet50V2, (b) VGG-16, (c) InceptionV3, (d) Fusion Model

We also investigate how the models are making decisions regarding COVID-19 prediction. More precisely, it is crucial to comprehend what the models are learning from the input images during training and validation. Class activation maps (CAMs) [19] were introduced as a technique to indicate a specific image region used by the CNN to determine output class while making predictions. This is achieved by projecting the output layer weights back to the last CNN part. We use Grad-CAM [34] which is a generalization of CAM for multiple end use cases which works based on the gradients with respect to the target class flowing into the last convolutional layer to generate a localization map pinpointing the crucial regions in the image for correct prediction.

Fig 10 shows the heatmap visualization of an example CXR image of a COVID-19 patient with GradCAM. The illustration includes the original image (focus areas marked with red arrows) followed by a heatmap highlighting the important regions within the lungs and an overlay of heatmap onto original image using the fusion and other evaluated base CNN models. The fusion model and VGG-16 seem to focus on both sides of the respiratory track to classify this particular case. However, ResNet50V2 and InceptionV3 seem to focus on relatively larger areas including lower respiratory track even though both the models have produced correct

predictions for the image. This needs to be verified clinically with an experienced radiologist through careful testing. The bottom line is to make sure that our model is depending on proper knowledge from the images to make the classification decisions.

4.5 Discussion

We have presented the performance comparison (as shown in Table 6) between the proposed fusion model in this study and some of the state-of-the-art methods from the literature. It is worthwhile to mention that the size of the CXR dataset used in these studies containing COVID-19 positive cases is very limited since COVID-19 is a new pandemic and the initiatives to collect and release large publicly available COVID-19 datasets are still in early stages. Some of the relatively earlier studies (e.g., [43, 46]) have used even less than 100 images of COVID-19 for model training. COVID-Net proposed by Wang and Wong [18] is one of the earliest

Table 6: Comparison of performance results from the proposed fusion model and state-of-the art methods

Authors	**Model**	**CXR Dataset Description**	**Classification Modes**	**Performance Evaluation**
Proposed Method	CNN Fusion Model	COVID-19 - 616 Normal- 616 Pneumonia- 616	Multiclass: COVID-19, Normal, Pneumonia	Acc (96.75%) Precision (98.38%) Sensitivity (98.38%)
Wang and Wong [18]	COVID-Net	COVID-19 - 183 Normal- 8066 Non-COVID19- 5538	Multiclass: COVID-19, Normal, Non-COVID19	Acc (92.4%) Precision (91.33%) Sensitivity (88.67%)
Narin et al. [31]	Pre-trained CNN	COVID-19 50 Normal- 50	Binary: COVID-19, Normal	Acc (97%)
Oh et al. [30]	ResNet18	Normal-191 Bacterial- 54 Tuberculosis-57 Viral-20 COVID-19-180	Multiclass: Normal, Bacterial, Tuberculosis, Viral, COVID-19	Acc (88.9%)
Ozturk et al. [27]	Darknet	COVID-19 - 127 No-findings - 500 Pneumonia- 500	Both binary and multiclass. Binary: COVID-19, No-findings Multiclass: COVID-19, No-findings, Pneumonia	Binary Acc (98.08%) Multiclass Acc (87.02%)
Ucar and Korkmaz [25]	SqueezeNet CNN	Normal- 1583, Pneumonia- 4290, COVID-19 - 76	Multiclass: COVID-19, Normal, Pneumonia.	Acc (95.7%) Sensitivity (90%)

efforts for COVID-19 detection from CXR images. They have proposed a custom deep learning model for the prediction of COVID-19 cases. However, the dataset used for model training contains less than 100 COVID-19 samples and a relatively large number of samples from healthy and non-COVID19 categories. This makes their dataset rather imbalance which may adversely affect the performance of the model. Furthermore, the size of parameters (111.6 million) used in COVID-Net is almost double than our proposed fusion model which contains about 62 million parameters. Thus, our fusion model offers a substantial computational savings in addition to performance gain as reported in Table 6. Another subsequent effort was done by Ucar and Korkmaz [25] who have used a Squeezenet CNN model for multilabel classification of CXR images. They have reported a state-of-the-art comparable accuracy of 95.7% on their dataset but the model seems to suffer from producing higher false negative (FN) cases which lowers down their sensitivity score (90%). Overall, the proposed fusion model shows superior performance results as compared to state-of-the-art methods while using a relatively higher number of COVID-19 samples (616) for model training.

Based on the amount of work done so far for the automatic detection of COVID-19 cases using deep learning models we can realize the role of AI in assisting radiologists for proper and faster diagnosis of potential coronavirus infection. However, we would safely argue that our fusion model is by no means a replacement for a human radiologist but rather we expect that our current findings offer a valuable contribution towards a growing recognition and adoption of AI-aided applications in clinical settings. Even if it is not sufficient to rely on the results obtained from CXR images to prescribe the line of treatment for a patient, an early screening can assist health professionals to quarantine positive encounters till a comprehensive test is prescribed.

5. Conclusions

In this paper, we introduce a new CNN fusion model for non-invasive screening of COVID-19 cases from CXR images. This study is one step towards better understanding of the dynamics of COVID-19 pandemic and proposes a state-of-the-art AI based solution for efficient and fast diagnosis system for COVID-19 infections The proposed research aims to achieve this through the development of a fusion network model with attention mechanism which consists of a weighted combination of different base CNN models. Specifically, we use fine-tuned pre-

trained ResNet50V3, VGG-16, and InceptionV3 models with attention network to extract salient features that are robust to overfitting and leverage a locally connected layer for a weighted contribution of these networks for final classification of COVID-19 cases. Experimental results demonstrate that our proposed fusion model with attention mechanism offer a highly accurate yet practical solution for automatically screening COVID-19 cases to accelerate line of treatment for patients.

References

[1] M. Abdel-Basset, W, Ding, L. Abdel-Fatah, "The fusion of Internet of Intelligent Things (IoIT) in Remote Diagnosis of Obstructive Sleep Apnea: A Survey & A New Model," Information Fusion. 10.1016/j.inffus.2020.03.010, 2020.

[2] C. Szegedy, W. Liu, Y. Jia, et al., "Going Deeper with Convolutions," in Proc. of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[3] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in Proc. of 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016.

[4] A. Krizhevsky, I. Sutskever, G.E. Hinton, "ImageNet classification with deep convolutional neural networks," Communication of the ACM, vol. 60, pp. 84–90, 2017.

[5] Wang S., Kang B., Ma J., Zeng X., Xiao M., Guo J., et al., "A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)," medRxiv, 2.

[6] A. Zhao, G. Balakrishnan, F. Durand, J.V. Guttag, A.V. Dalca. Data augmentation using learned transformations for one-shot medical image segmentation. In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. 2019. p. 8543–53.

[7] S. Pereira, A. Pinto, V. Alves, and C. A. Silva. Brain tumor segmentation using convolutional neural networks in mri images. IEEE transactions on medical imaging 2016;35(5):1240–51.

[8] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[9] W. Ding, "SVM-Based Feature Selection for Differential Space Fusion and Its Application to Diabetic Fundus Image Classification," in IEEE Access, vol. 7, pp. 149493-149502, 2019, doi: 10.1109/ACCESS.2019.2944899.

[10] J. Bullock, A. Luccioni, K. H. Pham, C. S. Lam, M. Luengo-Oroz, Mapping the landscape of Artificial Intelligence applications against COVID-19, arXiv:2003.11336v2 [cs.CY] 23 Apr 2020.

[11] Chen, J., Wu, L., Zhang, J., Zhang, L., Gong, D., Zhao, Y., Hu, S., Wang, Y., Hu, X., Zheng, B., Zhang, K., Wu, H., Dong, Z., Xu, Y., Zhu, Y., Chen, X., Yu, L., and Yu, H., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: A prospective study. medRxiv, 2020.

[12] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. 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, pp. 3-11, 2018.

[13] Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., Yu, L., Chen, Y., Su, J., Lang, G., Li, Y., Zhao, H., Xu, K., Ruan, L., and Wu, W. Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia. arXiv preprint arXiv:2002.09334, 1-29, 2020

[14] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp. 1-9.
[15] COVID-19 Dashboard, CoronaBoard, URL: https://coronaboard.com/. Retrieved Aug 10, 2020
[16] "Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)". World Health Organization. 30 January 2020. Archived from the original on 31 January 2020. Retrieved Aug 10, 2020.
[17] J. Cohen. COVID-19 image data collection, 2020. https://github.com/ieee8023/covid-chestxray-dataset
[18] Wang L, Wong A., COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. arXiv preprint, Mar 2020.
[19] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning Deep Features for Discriminative Localization," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2921-2929, doi: 10.1109/CVPR.2016.319.
[20] Chakraborty S., An attempt- Detection of COVID-19 presence from Chest X-ray scans using CNN & Class Activation Maps. https://towardsdatascience.com/detection-of-covid-19-presence-from-chest-x-ray-scans-using-cnn-class-activation-maps-c1ab0d7c294b
[21] Delft Imaging, CAD4COVID: Triage for COVID-19 Using Artificial Intelligence on Chest X-rays, https://www.delft.care/cad4covid/
[22] Abbas A, Abdelsamea MM, Gaber MM. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. arXiv preprint arXiv:200313815. 2020.
[23] Bukhari SUK, Bukhari SSK, Syed A, SHAH SSH. The diagnostic evaluation of Convolutional Neural Network (CNN) for the assessment of chest X-ray of patients infected with COVID-19. medRxiv. 2020.
[24] Hammoudi K, Benhabiles H, Melkemi M, Dornaika F, Arganda-Carreras I, Collard D, et al. Deep Learning on Chest X-ray Images to Detect and Evaluate Pneumonia Cases at the Era of COVID-19. arXiv preprint arXiv:200403399. 2020.
[25] F. Ucar, D. Korkmaz, "Covidiagnosis-Net: Deep Bayes-Squeezenet based diagnostic of the coronavirus disease 2019 (COVID-19) from X-ray images," Medical Hypotheses (2020) 109761, 2020.
[26] Karim M, Dohmen T, Rebholz-Schuhmann D, Decker S, Cochez M, Beyan O, et al. Deep COVID Explainer: Explainable COVID-19 Predictions Based on Chest X-ray Images. arXiv preprint arXiv:200404582. 2020.
[27] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," Computers in Biology and Medicine (2020) 103792, 2020.
[28] J. Liu, Artificial Intelligence Assisted Radiology Technologies Aid COVID-19 Fight in China, Mar 27, 2020. https://www.itnonline.com/article/artificial-intelligence-assisted-radiology-technologies-aid-covid-19-fight-china
[29] C. Ross and R. Robbins, Debate flares over using AI to detect COVID-19 in lung scans, Mar 30 2020. https://www.statnews.com/2020/03/30/debate-over-artificial-intelligence-to-detect-covid-19-in-lung-scans/
[30] Y. Oh, S. Park, J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," IEEE Transactions on Medical Imaging, vol. 39, issue 8, pp. 2688 – 2700, 2020.

[31] A. Narin, C. Kaya, Z. Pamuk, “Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks,” arXiv preprint arXiv:2003.10849, 2020.
[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.
[33] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li et al.,"ImageNet: A large-scale hierarchical image database," International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255, 2009.
[34] Selvaraju, R. R.; Abhishek D.; Ramakrishna V.; Michael C.; Devi P. et al. “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” International Journal of Computer Vision, vol. 128, no. 2, pp. 336–359, 2019.
[35] S. Jetley, N.A. Lord, N. Lee, and P.H. Torr, “Learn to pay attention,” CoRR, abs/1804.02391, 2018.
[36] K. Gregor and Y. LeCun, "Emergence of complex-like cells in a temporal product network with local receptive fields," CoRR arXiv:1006.0448, 2010.
[37] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1701-1708, doi: 10.1109/CVPR.2014.220.
[38] Chest X-Ray Images (Pneumonia), https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia, Last Accessed August 10, 2020.
[39] L. van der Maaten, G. Hinton, “Visualizing Data using t-SNE,” Journal of Machine Learning Research, 9(2605):2579-2605, 2008.