



Rapid COVID-19 diagnosis using ensemble deep transfer learning models from chest radiographic images

Neha Gianchandani¹ · Aayush Jaiswal¹ · Dilbag Singh² · Vijay Kumar³ · Manjit Kaur²

Received: 11 June 2020 / Accepted: 3 November 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes novel coronavirus disease (COVID-19) outbreak in more than 200 countries around the world. The early diagnosis of infected patients is needed to discontinue this outbreak. The diagnosis of coronavirus infection from radiography images is the fastest method. In this paper, two different ensemble deep transfer learning models have been designed for COVID-19 diagnosis utilizing the chest X-rays. Both models have utilized pre-trained models for better performance. They are able to differentiate COVID-19, viral pneumonia, and bacterial pneumonia. Both models have been developed to improve the generalization capability of the classifier for binary and multi-class problems. The proposed models have been tested on two well-known datasets. Experimental results reveal that the proposed framework outperforms the existing techniques in terms of sensitivity, specificity, and accuracy.

Keywords COVID-19 · SARS-CoV-2 · Transfer learning · Chest X-ray · Ensemble models

1 Introduction

The outbreak of novel coronavirus surprised the whole world. The disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is also known as COVID-19. According to the World Health Organization (WHO), more than fifteen million peoples are infected from this virus across 215 countries (Coronavirus Disease (COVID-19) 2020). 0.63 million deaths with 9.43 million recovered cases are reported globally by 23 July 2020. USA, Brazil, and India are severely affected with 4.1 million, 2.2 million, and 1.3 million active corona cases, respectively.

Due to the communicable nature of this virus and inappropriate treatment, the early detection of an infected person

is required to discontinue the coronavirus spread (Oh et al. 2020). The main source for detecting the infected person is symptoms developed in the patients. The infected person may suffer from fever, cough, breathing problem, sore throat, diarrhoea, and headache (COVID-19 symptoms 2020). Loss of smell, tiredness, disappearing of taste, and aches may be found in some patients. However, the presence of COVID-19 symptoms may not be found in some infected persons (Loey et al. 2020). Due to this, the detection of an infected person is a very difficult task. The health facilities of many developed countries have been exhausted due to the rapid increase in corona affected persons. These countries are facing a shortage of corona testing kits and ventilators. Due to this, some countries announced the lockdown to break the chain of coronavirus and safe their population.

Besides the imposition of lockdown, screening of patient is required for isolation and treatment. Initially, real-time reverse transcription-polymerase chain reaction (RT-PCR) method is used for detecting the coronavirus infection in the patients (Corman et al. 2020). RT-PCR is performed on the clinical samples of patients and results are obtained within a few hours to two days. Due to the low sensitivity of RT-PCR and unavailability of kits, radiography imaging techniques can be utilized to detect COVID-19 in the patients (Xie et al. 2020). It has been observed from research articles that chest scans may be suitable for COVID-19 detection

✉ Manjit Kaur
Manjit.kaur@bennett.edu.in

¹ Department of Computer Science and Engineering, School of Computing and Information Technology, Manipal University Jaipur, Jaipur, Rajasthan 303007, India

² Computer Science Engineering, School of Engineering and Applied Sciences, Bennett University, Greater Noida 201310, India

³ Department of Computer Science and Engineering, National Institute of Technology Hamirpur, Hamirpur, Himachal Pradesh 177005, India

(Corman et al. 2020; Xie et al. 2020). Hence, the diagnosis and detection of coronavirus infection in COVID patients can be done through chest radiography techniques. The chest X-ray and chest CT scan are two well-known radiological techniques. Chest X-ray is preferred over the chest CT scan due to easy availability of X-ray machines in hospitals and low ionizing radiation exposure to patients (Yu et al. 2020).

The radiological expert is required to analyze chest X-ray for finding the COVID-19 imaging patterns. However, it is a time-consuming and error-prone task (Cellina et al. 2020). Therefore, the automatic analysis of chest x-ray is required. Recently, deep learning (DL) techniques are widely used in medical imaging for the diagnosis of different diseases (Shen et al. 2017). DL techniques automatically extract the features from the given image without any user intervention (Wang and Xia 2018). Due to this, these techniques are used in the classification of infections in chest X-ray images. Therefore, DL techniques are used for automatic analysis of chest X-ray for detecting the COVID-19 imaging patterns from patient's images.

The main contribution of this paper is as follows:

- (i) An ensemble deep transfer learning models are designed for COVID-19 diagnosis using chest X-ray images.
- (ii) The proposed model not only helpful in diagnosing the COVID-19 infected patients but also able to differentiate the COVID-19 from diagnosing pneumonia (i.e., viral and bacterial).
- (iii) The proposed framework extracts the features from chest X-ray images using an ensemble deep learning network. Thereafter, the extracted features are applied to classifier for further classification.
- (iv) The performance of the proposed framework has been tested on two well-known datasets.
- (v) The proposed framework has been compared with the competitive models in terms of various performance metrics such as accuracy, f-measure, area under curve, precision and recall.

The remainder of this paper is structured as follows. The related work is described in Sect. 2. Section 3 presents the proposed approach for COVID-19 detection in infected persons. Experimental results and discussions are given in Sect. 4. The concluding remarks are drawn in Sect. 5.

2 Related work

Recently, deep learning techniques such as convolutional neural networks (CNN) and transfer learning are widely used for classification of COVID-19 from chest X-rays. Researchers have done a lot of work in the short span of time.

Basu and Mitra (2020) proposed a domain extension transfer learning for detecting the abnormality instigated by COVID-19 in chest X-rays. They used Gradient Class Activation Map for detecting features from X-ray images and validated over NIH chest X-ray dataset. The accuracy obtained from this model is 95.3%. Luz et al. (2020) studied the deep learning architectures for identification of COVID-19. COVIDx dataset is used to demonstrate the efficiency of the proposed model. Flat EfficientNet model achieved 93.9% accuracy with 96.8% sensitivity. Hemdan et al. (2020) developed a novel automatic framework namely COVIDX-Net for identification of coronavirus infection in the patients' chest X-ray. COVIDX-Net utilized seven different deep learning architectures. COVIDX-Net is tested on only 50 chest X-ray images. Both VGG19 and DenseNet201 provided 90% accuracy on these images. Ozturk et al. (2020) presented an automatic COVID-19 detection model named as DarkCovidNet using chest X-ray. DarkCovidNet is trained on 125 COVID-19 chest images. The classification accuracies obtained from this model are 98.08% and 87.02% for binary and multi-class cases. The main drawback of this model is that the limited number of coronavirus infected X-ray images are used for validation.

Togaçar et al. (2020) used a fuzzy colour technique for converting the original chest X-ray dataset into a structured dataset. Thereafter, the structured dataset is converted into a stacked dataset using image stacking technique. MobileNetV2 and SqueezeNet are used for identification of COVID-19 patterns in the images. The classification accuracy obtained from MobileNetV2 and SqueezeNet is 98.25% and 97.81%, respectively. Tuncer et al. (2020) proposed a novel method for detection of COVID-19 pattern in chest X-ray images. Residual Exemplar Local Binary Pattern (ResExLBP) is used for feature extraction. After that, the important features are selected through iterative ReliefF. These selected features are applied to five well-known classifiers. This approach is tested on 321 chest X-ray images and achieved a 99% success rate for SVM classifier. Das et al. (2020a, b) presented an automatic detection of coronavirus infection from chest X-ray images using deep learning technique. The classification accuracy obtained from the proposed model is 97.40%. Wang and Wong (2020) developed a COVID-Net framework for detection of COVID-19. COVID-Net attained 92.4% accuracy in classifying normal, pneumonia, and COVID-19 classes. COVID-Net provides better performance than the VGG-19 and ResNet-50 architecture.

Mahmud et al. (2020) proposed a CovXNet for detecting COVID-19 patterns from the patient's chest X-rays. The predications obtained from CovXNet are further optimized using the stacking algorithm. Their model is tested over 915 images. The classification accuracies obtained from CovXNet are 97.4% and 89.6% for binary and three-class cases,

respectively. Apostolopoulos and Mpesiana (2020) presented an automatic detection of COVID-19 from X-ray images using five pre-trained deep learning architectures. The pre-trained architectures are VGG-19, MobileNet, Inception, Xception, and Inception_ResNet_V2. The accuracies obtained from VGG-19 are 98.75% and 93.48% for binary and multiclass cases, respectively.

Rahimzadeh and Attar (2020) presented the hybridization of Xception and ResNet50_V2 for identification of COVID-19 patterns in the chest X-ray images. Their model is tested on 6054 images and attained classification accuracy of 91.4%. Shelke et al. (2020) proposed a classification model to analyze and diagnosis of COVID-19 using chest X-ray images. The accuracy obtained from DenseNet-161 is 98.9%. However, they have tested on a very small dataset that consists of only twenty-two X-ray images. Narin et al. (2020) studied the three pre-trained deep learning architectures namely ResNet50, Inception_V3, and Inception-ResNet_V2 for detection of coronavirus infection in the chest X-ray images. The COVID-19 detection accuracy achieved from ResNet50 is 98%.

Abbas et al. (2020) proposed a DeTraC model for detection of COVID-19 from the chest X-ray images. They used a pre-trained CNN model for deep feature extraction. Thereafter, class decomposition is used to extract the local structure from the data distribution. Gradient descent method is used for classification. The classification results are further refined through a class composition layer of DeTraC. DeTraC achieved an accuracy of 95.12% over 105 chest X-ray images. Chouhan et al. (2020) proposed an ensemble approach for pre-trained deep learning architectures. They used AlexNet, DenseNet121, Inception_V3, ResNet18 and GoogleNet architectures for COVID-19 classification. Their ensemble model achieved the classification accuracy of 96.4%. Das et al. (2020a, 2020b) presented a truncated InceptionNet model for screening of COVID-19 from the chest X-ray images. The model is tested on three publicly available chest X-ray datasets.

3 Proposed ensemble deep transfer learning model

In this section, the detail description of the proposed ensemble deep learning model is presented.

3.1 Motivation

The availability of large-scale well-annotated datasets especially in image classification supports better feature extraction and results owing to a distinct boundary between various classes and enough data to train on. Hence, the models trained on such diverse data are generic enough to learn and

represent class boundaries even in the case of medical data where there is a contingency for high generalization error due to the unbalanced and limited availability of dataset. Transfer Learning, is thus, utilizing the feature extraction capabilities of pre-trained models on huge datasets like ImageNet (Deng et al. 2009) that can capture the class boundaries well. Transfer learning not only eliminates the prerequisite for a huge dataset but also provides quick results in this pandemic situation. This motivates us to utilize these techniques in our proposed deep learning framework for classification of COVID-19.

3.2 Ensemble learning

To reduce the errors in detection, an ensemble deep transfer learning model is proposed by combining the outputs of different independently trained neural networks (in our case, transfer learning models) as shown in Fig. 1. An ensemble of transfer learning networks may be a robust approach by reducing errors. It produces optimal results from the combined networks with the least possible errors. After data-preprocessing, the convolutional neural network architecture is built using pre-trained models. The core components of the employed models are described in the preceding subsections.

3.3 Layers of deep CNN

3.3.1 Pooling layers

Pooling layers are used for multi-scale analysis and reducing the input data (image) size for feature reduction. Max Pooling and Average Pooling layers are most widely utilized in CNN. The average pooling layer (l_{avg}) is used in the proposed architecture, which can be mathematically defined as (Wang et al. 2018):

$$l_{avg} = \frac{\sum A_x}{|A_x|} \quad (1)$$

where x and $|k|$ denote pooling region and k cardinal number, respectively. A denotes the activation set in x defined as:

$$A = a_y \quad (2)$$

Here, $y \in \text{poolingregion}(x)$.

3.3.2 Fully connected layer and Softmax

The fully connected layers (FC) have full connections to the neurons. The inputs to FC are multiplied with FC weight matrix to produce the multiplication result. In the proposed architecture, a fully connected dense layer is used with Softmax activation function for multi-class classification.

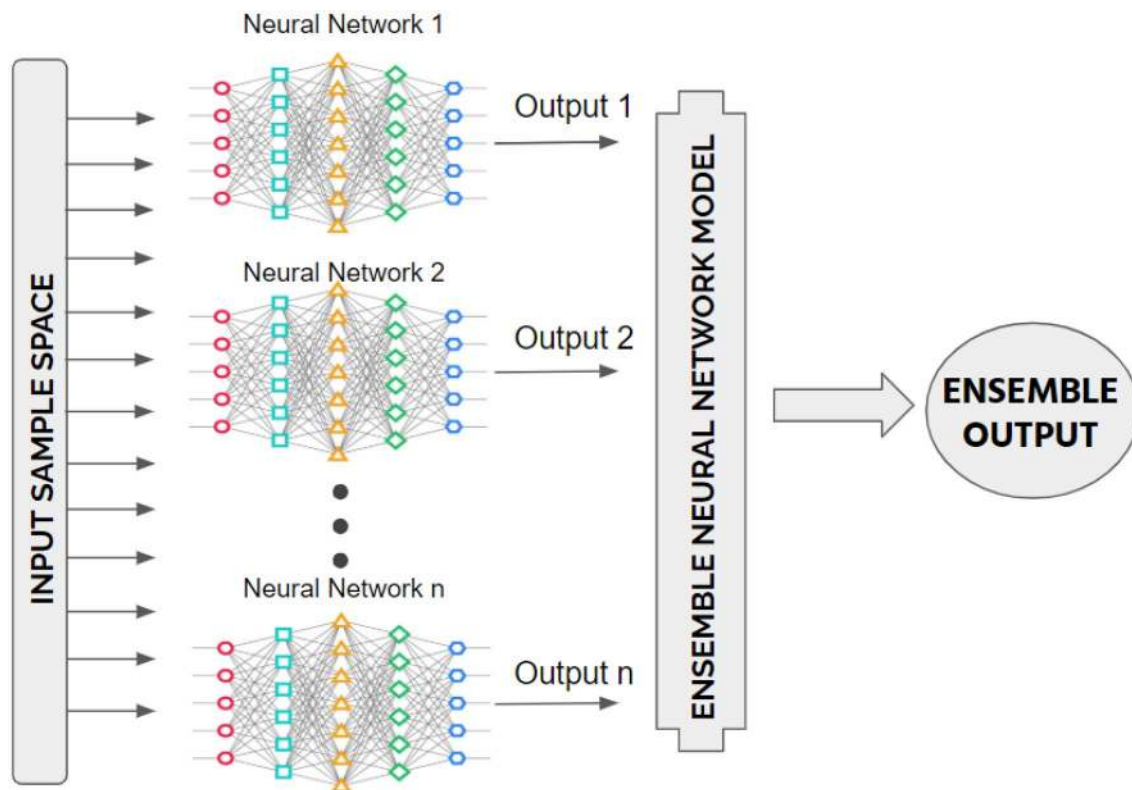


Fig. 1 Block diagram of an ensemble learning model by considering n number of artificial neural networks

The working of the dense layer with Softmax activation (Wang et al. 2018) can be described through the concept of conditional probability. Softmax function ($P(c, s)$) is the probability of a sample s belonging to class c and defined as:

$$P(c, s) = \frac{P(s, c) \times P(c)}{\sum_{n=1}^C P(n) \times P(s, n)} \tag{3}$$

In Eq. 3, $P(c)$ is the class probability and the total number of classes is given by C and the class probability is $P(c)$. Equation 3 can also be redefined as:

$$\text{Softmax} = P(c, s) = \frac{\exp(\beta^c[s])}{\sum_{n=1}^C \exp(\beta^n[s])} \tag{4}$$

Here,

$$\beta^c = \ln[P(s, c) \times P(c)] \tag{5}$$

3.3.3 Dropout

Dropouts are added to regularize the convolutional neural networks by randomly setting the outgoing neurons at the hidden layers 0 at each training iteration. The dropped neurons have no contribution to the forward pass or

back-propagation while training phase. This leads to different architecture being used for each forward pass step despite having shared weights. This assists in avoiding over-fitting in the model. In this paper, the dropouts are set to 0.2 and 0.3 for the three-class classification problem (Krizhevsky et al. 2012).

3.3.4 Rectified linear units (ReLU)

The activation function of ReLU can be defined as:

$$f_{relu}(a) = \max(0, a) \tag{6}$$

The training time of ReLU is significantly lesser than sigmoidal functions. It is also helpful to overcome gradient-based training and poor performance in neural networks due to the widespread saturation. In comparison to the sigmoid function, ReLU reduces the training time by fastening the convergence of stochastic gradient descent (Wang et al. 2018).

3.3.5 Sigmoid function

The sigmoid function (also referred to as logistic function) is simply used to predict the probability of a given output. The

sigmoid function gives desirable results for the probability of [0,1]. The sigmoid produces a flexible S-shaped curve or the ‘‘Sigmoid’’ curve. It can be mathematically defined as:

$$\text{Sigmoid} = \sigma(\theta) = \frac{1}{1 + e^{-\theta}} \quad (7)$$

Here, θ is the input.

3.4 Architectures of deep transfer learning models

Figure 2 shows the architectures of DenseNet201 (Huang et al. 2017), ResNet152V2 (He et al. 2016), Inception-ResNetV2 (Szegedy et al. 2017), and VGG16 (Simonyan and Zisserman 2014). The neurons for the first dense layer are 64 in place of 128 for binary class classification problem. The fine-tuned pre-trained model with several layers is used for feature extraction. In Dense Layer, Softmax activation function is introduced for three-class classification problem. However, Sigmoid activation is used for binary classification. The models are trained for 300 epochs with batch size set to 16. Adam optimizer is used for fine-tuning of the models. To prevent overfitting, regularization is achieved by using an early stopping criterion.

3.4.1 Ensemble deep transfer learning model for three-class classification problem

The above-mentioned transfer learning classifiers motivated us to propose the ensemble stacked classifier. The ensemble of deep CNNs may prove to be a powerful technique for better results as it works on the concept of combining the decisions obtained from various models (Siwek et al. 2009; Islam and Zhang 2018). Owing to the stochastic nature of deep CNNs, each neural network architecture learns some specific patterns from other neural networks. The ensemble technique provides better boosting accuracy and feature extraction capability (Alshazly et al. 2019).

Figure 3 shows the architecture of the proposed ensemble model for three-class classification problem. The ensemble is done through a concatenation of two deep learning models namely, VGG16 and DenseNet201. InceptionResNetV2 and ResNet152V2 are not considered for the ensemble process due to their relatively low performance for the three-class problem as reported in the literature. During the hyperparameter tuning, fully connected layers (64 neurons each with 0.2 and 0.1 dropouts) having Softmax activation function have been added for three-class classification.

The base neural networks have been frozen to preserve ImageNet weights during the training phase. The architecture added after ensemble is trained on Dataset 2. The ensemble neural network has been trained for 10 epochs with a batch size of 64. The learning rate is set to 0.00001.

3.4.2 Ensemble deep transfer learning model for two-class classification

The four modified architectures of pre-trained models as depicted in Fig. 2 (using Sigmoid in the last Dense Layer instead of Softmax) are trained on the first dataset. Similar to the three-class ensemble model, all the four transfer learning models, i.e., modified DenseNet201, Inception ResNetV2, ResNet152V2, and VGG16 are initially trained and evaluated individually. The best two of four models are further ensembled in the proposed architecture. A similar architecture is followed post ensemble with two dense layers (64 neurons each and dropouts) for further classification. However, a dense layer with Sigmoid activation is introduced for final binary classification to obtain probabilities in the range of [0,1] for two classes namely, COVID positive and COVID negative. Figure 4 shows the architecture of the proposed model for binary classification.

The proposed model is trained for 20 epochs while keeping the base models are frozen. The batch size and learning rate are set to 64 and 0.00001, respectively. The entire pre-processing, augmentation, and model training is performed through TensorFlow framework.

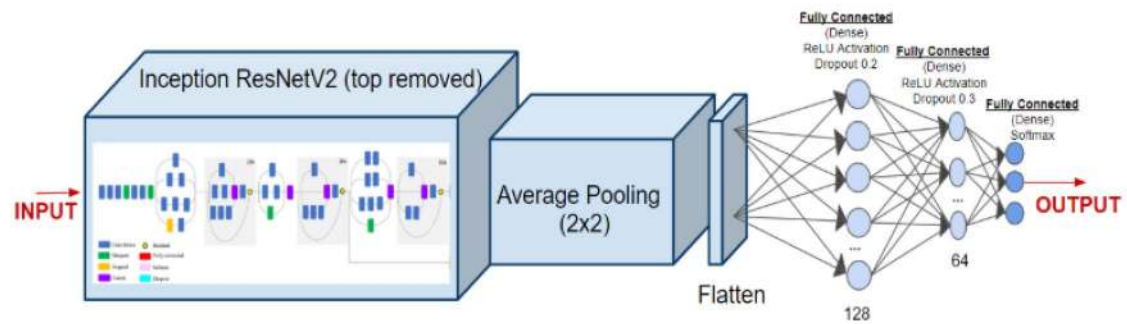
4 Performance analysis

In order to validate the performance of the proposed model, they have been compared with recently developed transfer learning models such as VGG16, ResNet152V2, Inception-ResnetV2, and DenseNet201. Experimental results and discussion are mentioned in the preceding subsections.

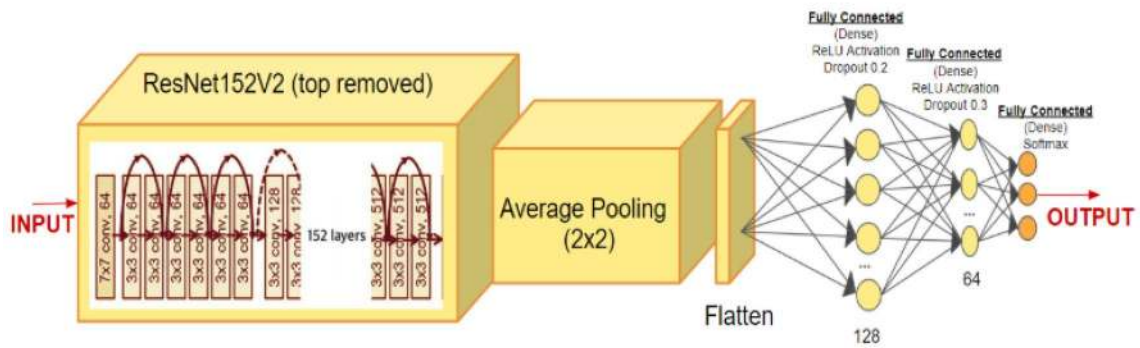
4.1 Dataset

Two different datasets are used for validation of the proposed models. The first dataset has been obtained from Kaggle datasets resource (Asraf 2020). It consists of X-Ray-scans that have COVID positive, COVID negative and Pneumonia subjects. The COVID positive and COVID negative from the mentioned dataset is utilized for binary classification.

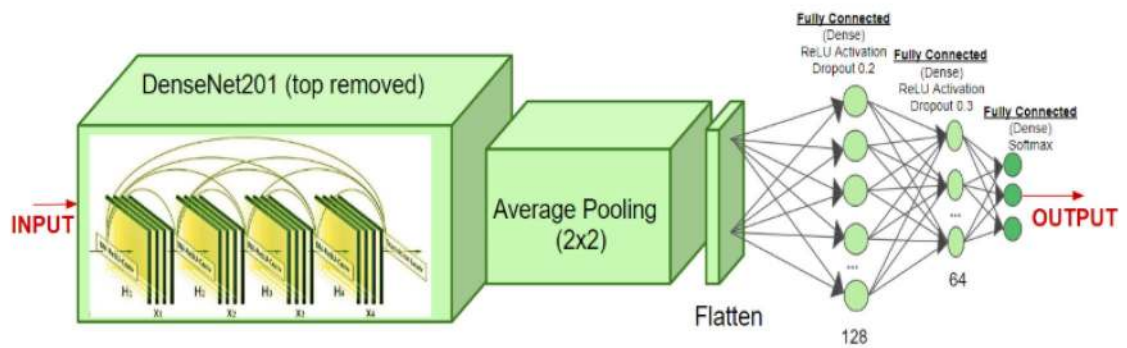
The second dataset is created by researchers from the University of Dhaka and Qatar University along with medical practitioners and collaborators (Chowdhury et al. 2020). It consists of 1203 Chest X-ray scans on COVID, normal, and pneumonia subjects (viral as well as bacterial pneumonia). This dataset is used for multi-class classification. Due to the lack of data corresponding to the COVID class, image data from (Darshan 2020) is also combined for making it 401 images corresponding to each class. Samples images taken from two datasets are shown in Fig. 5.



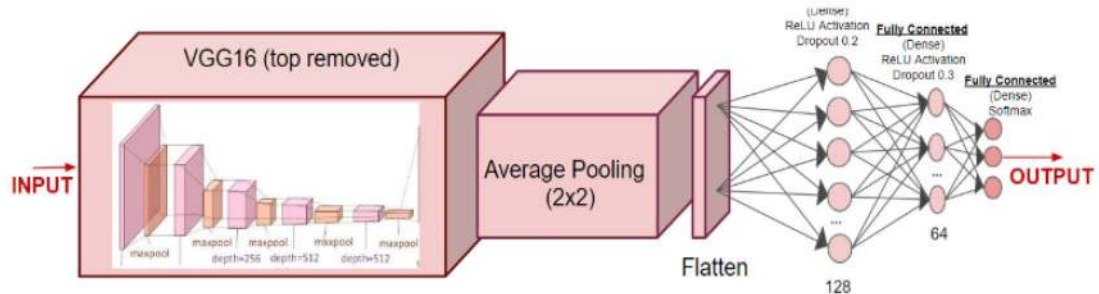
(a) Modified Inception ResNetV2



(b) Modified ResNet152V2



(c) Modified DenseNet201



(d) Modified VGG16

Fig. 2 Architectures of modified deep transfer learning models a Inception ResNetV2, b ResNet152V2, c DenseNet201, d VGG16

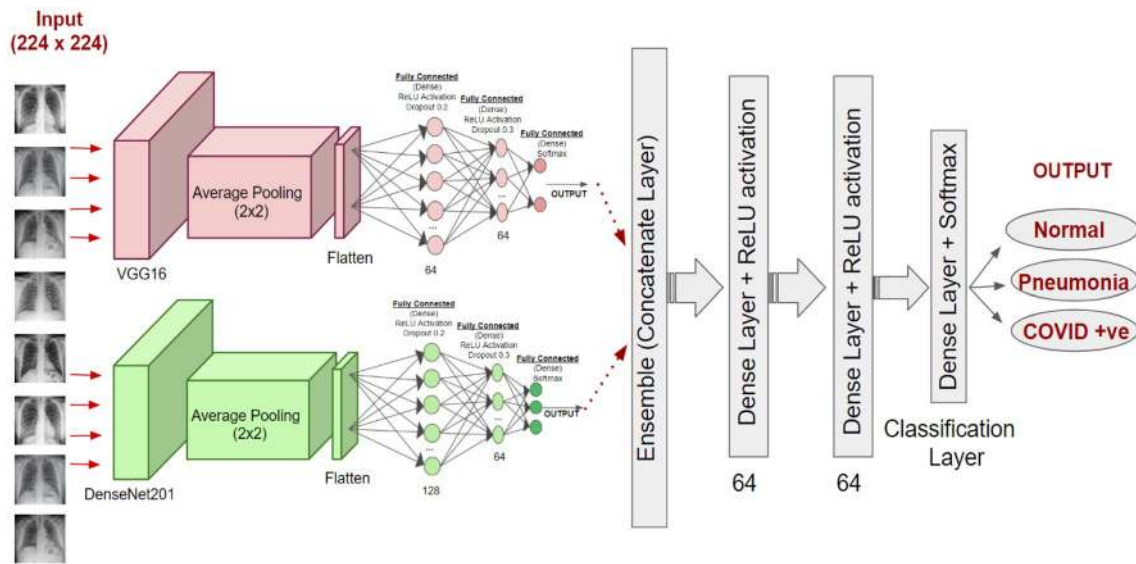


Fig. 3 Proposed ensemble deep transfer learning model by using the modified VGG16 and DenseNet201 models for multi-classification

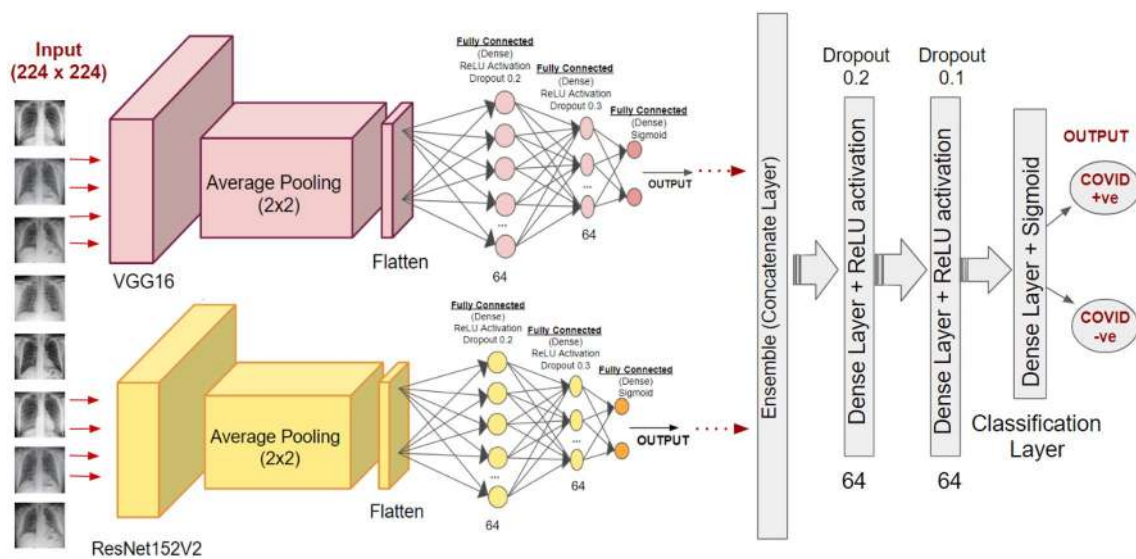


Fig. 4 Proposed ensemble framework (architecture) for binary classification

4.2 Data preparation and preprocessing

For better performance of the proposed system, X-ray images are resized to $224 \times 224 \times 3$ (RGB). It is essential to augment data for better generalization capability. As neural networks have millions of parameters, the data must be in proportional orders for a good learning capacity. To make up for the lack of data available, the training and validation data is augmented by horizontal flipping, vertical flipping, rotation by 45° and slant angle (0.2) for sheer transformation.

Augmentation is also carried out on the validation dataset as our model may be validated through the variety of inputs.

The models not only learn from synthetically modified data but are also validated on augmented data. To ensure a uniform data distribution, image normalization has been carried out by dividing the images by the number of channels, i.e., 255 to achieve the normalized data in the range of [0, 1]. This will ensure better convergence during the training of the neural network.

The last step is data splitting for training, validation, and testing. For both classification problems, the split is initially done through 15% of data for testing and remaining data (85%) is again split. Among this data, 68% for model training and 17% for model validation. The model learns

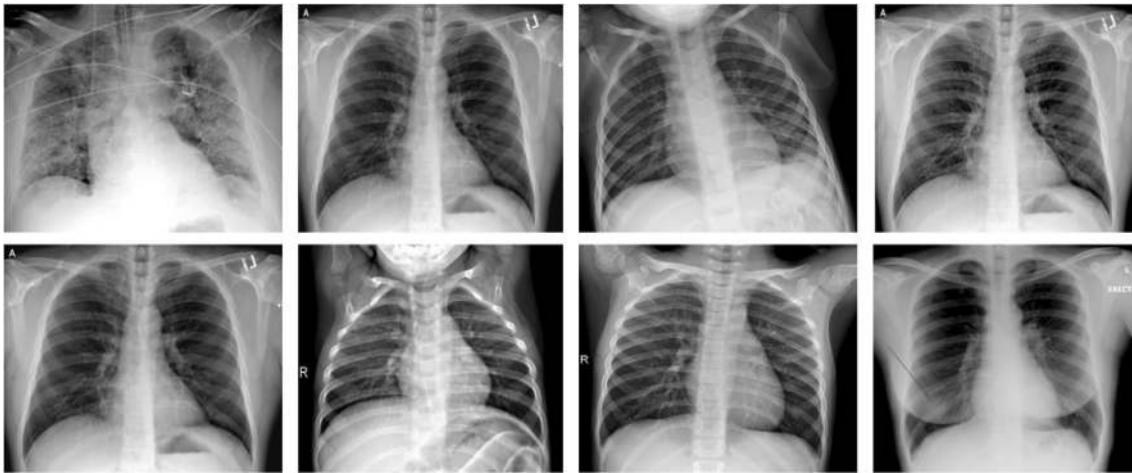


Fig. 5 Sample X-ray images for binary and multi-classification

and assigns weights from the training data due to this more weightage has been assigned from the total dataset. However, a slight variation in results may be noticed if the training proportion of data is decreased or increased.

4.3 Performance metrics

The performance of the proposed models has been tested for classification. The confusion-matrix based quantitative metrics are Accuracy (*ACC*), Sensitivity (*SNS*), Specificity (*SPS*), Precision (*P*), Recall (*R*), and F1-Score (*F1*). Accuracy (*ACC*) can be defined as the overall performance of the model by quantifying the correctly predicted labels. *ACC* can be mathematically explained as (Basavegowda and Dagnev 2020, Osterland and Weber 2019, and Ghosh et al. 2020):

$$Accuracy = ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

SNS and *SPS* denote how accurately a classifier predicts the positive and negative labels quantifying the true positive rate and the true negative rate respectively. *SNS* and *SPS* can be explained as in Eq. 9 and Eq. 10, respectively (Gupta et al. 2019):

$$Sensitivity = SNS = \frac{TP}{TP + FN} \quad (9)$$

$$Specificity = SPS = \frac{TN}{TN + FP} \quad (10)$$

The exactness of results can be evaluated using Precision (*P*). Precision gives the agreement of data labels with the positive labels given by the classifier (Please refer to Eq. 11). Similarly, Recall (*R*), also known as Sensitivity (as defined

in Eq. 9) evaluates the completeness of the classifier. Higher the *R*-value lower would be false negatives as can be defined as (Wang et al. 2019 and Wiens 2019):

$$Precision = P = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = R = \frac{TP}{TP + FN} \quad (12)$$

F1-Score quantifies the data distributions by giving a combined score of *SNS* and *SPS* as in Eq. 13

$$F1Score = F1 = \frac{2 \times P \times R}{P + R} \quad (13)$$

In Eqs. 8–13, *TP* defines the true positive subject, i.e., a subject who is COVID-19 +ve and the model has correctly classified the same as COVID-19 +ve. Similarly, *TN* is the truly predicted negative case. Whereas, *FP* AND *FN* are the incorrect predictions by the model corresponding to COVID-19 +ve and COVID-19 -ve subjects. The proposed model has also been evaluated using a multiclass confusion matrix for three classes that are COVID-19 (+ve), Normal and Pneumonia (see Fig. 6).

4.4 Experimentation 1: three class-classification

Table 1 shows performance evaluation of the proposed approach with VGG16, ResNet152V2, InceptionResnetV2, andDenseNet201. It is observed from Table 1 that the above-mentioned models achieve comparable results. DenseNet201 attained the best testing accuracy of 96.68%, which is better than VGG16, ResNet152V2, and Inception-ResnetV2. The proposed ensemble technique outperforms the pre-trained models with a testing accuracy of 99.21%.

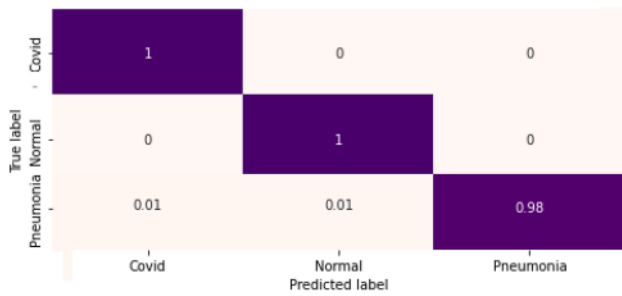


Fig. 6 Confusion matrix for three-class classification

The proposed architecture provides better generalization capability than the other models. The ensemble technique takes a minimum of 6 mi for the training phase and a few seconds for the testing phase. Each class namely COVID-19 (+ve), normal, and pneumonia achieve high precision values. It means that the correct positive predictions out of all the subjects can be classified as positive.

The precision for pneumonia subjects is one indicating that there are zero false-positive predictions. The macro average (unweighted mean per label) precision, recall, and F1-Score for the proposed framework are 0.99, which justifies the validity of the model.

Figure 6 shows the computed confusion matrix for three-class classification. Figure 7 shows the training and validation loss for the proposed model. The model converges well, reaching peak accuracy while keeping the training as well as validation losses minimal. It is found that the training and validation losses are not far from approximately similar values. Our model is not over-fitted for this dataset.

Figure 8 depicts the Area Under the receiver curve (AUC) for the proposed ensemble model. It is observed that a perfect AUC = 1 is achieved for COVID-19 (+ve) subject and normal subjects, indicating the model has a perfect sense of separability of the said, two classes. For pneumonia, the score is 0.99 is achieved, which is appreciable, however, for a critical situation like COVID-19, efforts to increase it can be made.

4.5 Experimentation 2: two class-classification

For the binary classification, the four learning models are individually trained and achieved accuracies of 95.12% (VGG16), 94.61% (ResNet152V2), 93.07% (Inception-ResNetV2) and 94.10% (DenseNet201). The training time is calculated to be approximately 10 min for the proposed architecture. Both VGG16 and ResNet152V2 achieved the highest accuracy. However, there is still scope of improvement (precision and sensitivity scores) by considering the

Table 1 Test evaluation of three class classifiers

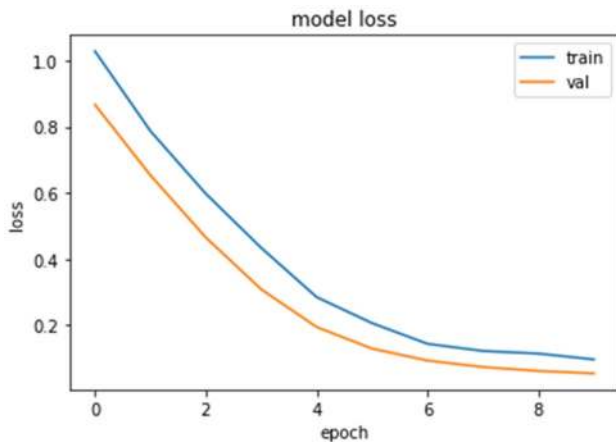
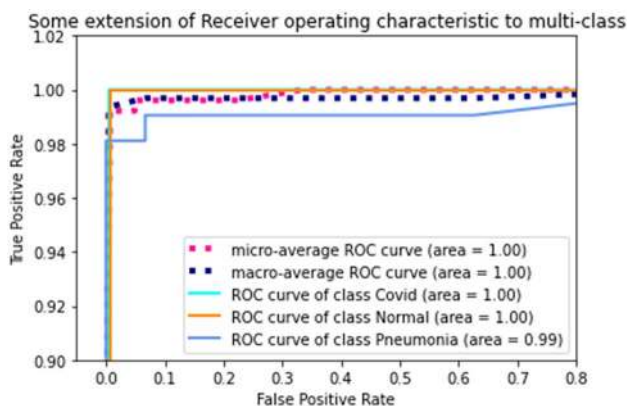
Model/Metric	Label	Precision	Recall	F1-Score	Accuracy(overall) (%)
VGG16	Covid19	0.95	0.98	0.97	95.70
	Normal	0.94	1.00	0.97	
	Pneumonia	0.98	0.88	0.93	
	[Macro average]	0.96	0.96	0.96	
ResNet152V2	Covid19	0.92	1.00	0.96	95.58
	Normal	0.95	0.98	0.96	
	Pneumonia	0.99	0.92	0.95	
	[Macro average]	0.95	0.96	0.96	
InceptionResnetV2	Covid19	0.95	1.00	0.98	95.02
	Normal	0.95	0.95	0.95	
	Pneumonia	0.95	0.90	0.92	
	[Macro average]	0.95	0.95	0.95	
DenseNet201	Covid19	0.95	1.00	0.98	96.68
	Normal	0.95	1.00	0.98	
	Pneumonia	1.00	0.90	0.95	
	[Macro average]	0.97	0.97	0.97	
Proposed Ensemble (VGG16 + DenseNet)	Covid19	0.98	1.00	0.99	99.21
	Normal	0.99	1.00	0.99	
	Pneumonia	1.00	0.98	0.99	
	[Macro average]	0.99	0.99	0.99	

Final precision, recall, F1-score and accuracy (macro average values) of the proposed ensemble model for three-class classification are highlighted in bold

Table 2 Test evaluation of two-class classifiers

Model/Metric	Precision	Recall	F1-Score	Sensitivity	Specificity	Accuracy (%)
VGG16	0.945	0.969	0.952	0.969	0.933	95.12
ResNet152V2	0.939	0.953	0.946	0.953	0.938	94.61
InceptionresNetV2	0.924	0.938	0.931	0.938	0.923	93.07
DenseNet201	0.901	0.989	0.943	0.989	0.892	94.10
Proposed ensemble	0.959	0.964	0.961	0.964	0.958	96.15

Final precision, recall, F1-score, sensitivity, specificity and accuracy of the proposed ensemble model for binary classification are highlighted in bold

**Fig. 7** Loss plot for the proposed ensemble (three class)**Fig. 8** Area under ROC (AUC) for the proposed ensemble, depicting the specific AUC for each class along with the micro-average and macro-average overall AUC

sensitive problem. For this, the ensemble model is introduced, which is utilized in the multi-class classification problem. The proposed ensemble architecture consists of VGG16 and ResNet152V2 that surpassed the basing of models by achieved a 1.2% (roughly) higher accuracy. The proposed architecture attained 0.959 precision and

accuracy of 96.15% indicating the correctness of predicted COVID and normal subjects.

The experimental results reveal that the proposed ensemble technique attains high specificity rates. It means that there would be no false-positive predictions, which directly implies a lesser obligation on the healthcare system. This will aid in the correct usage of testing kits and health-care facilities by reaching those who genuinely require it.

To evaluate the proposed binary classifier, the confusion matrix and the loss plot is given in Figs. 9 and 10, respectively. 4% false predictions are observed (positive as well as negative) with a 96% true predication for COVID-19(+ve) as well as normal subjects. The model fits well, as can be seen from the loss plot, avoiding any overfitting, as clear from the minimum (~ 0) difference between the training and validation error.

The Area Under the receiver curve (AUC) is also computed as shown in Fig. 11 for all binary classifiers implemented. It can be observed that the proposed ensemble performs better than the individual base models giving good separability between the two classes. This also supports the fact that the ensemble process results in better-generalized models resulting in efficient frameworks. Although, a 0.1% difference is seen between the AUC of both base models (VGG16 and ResNet152V2) and the proposed ensemble, in medical research, 0.1% increase in the model quality is critical.

4.6 Comparative analysis

It can be established from the results that deep learning framework can significantly assist in the detection of COVID-19 subjects using X-ray images by giving a low cost and rapid solution for the diagnosis of COVID-19. Table 3 shows the performance evaluation of the proposed binary classification technique with state-of-the-art techniques. Waheed et al. (2020) proposed a framework an appreciable accuracy of 95% however, the false negative is the rate was 10%, which is menacing for the specific problem in hand. The proposed architectures by Hemandan et al. (2020) and Sethy and Behera (2020) achieved good results with high sensitivity and specificity values,

Fig. 9 Confusion Matrix for binary classification

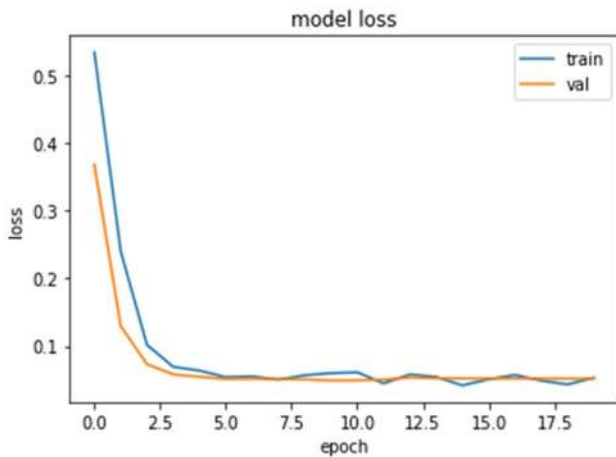
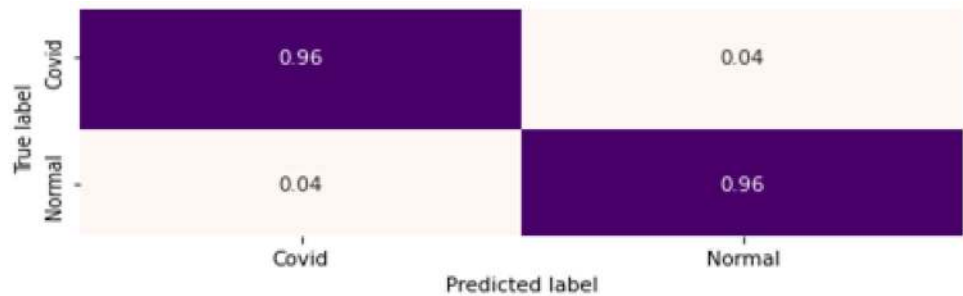


Fig. 10 Loss plot for binary classification (ensemble)

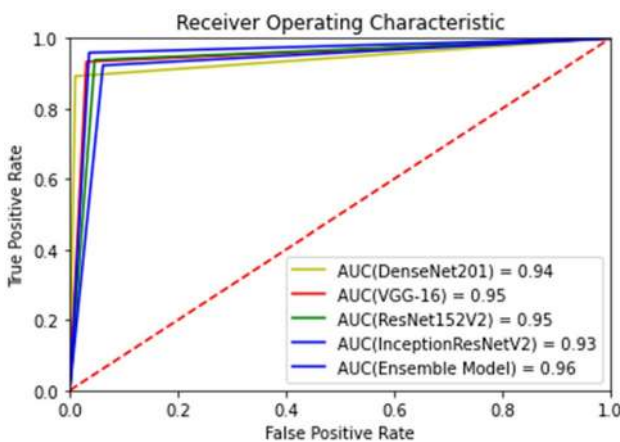


Fig. 11 Area under ROC (AUC) for the 5 models implemented for binary classification

the data utilized was extremely limited, which may indicate reduced generalization. To overcome these possible inhibitions, we utilized a larger data set with augmentation and well as tried to reduce the false predictions. Considering, the statistical metrics in Table 2, the proposed ensemble surpassed the state-of-the-art concerning

Table 3 Performance evaluation of different deep learning techniques on the binary dataset

State-of-the-art	Accuracy (%)
ResNet50 + SVM (Sethy and Behera 2020)	95.38
CovidGAN (Waheed et al. 2020)	95
COVIDX-Net (Hemdan et al. 2020)	90
CAAD (Zhang et al. 2020)	72.77
CNN trained from scratch (Maghdid et al. 2020)	94
Proposed ensemble model	96.15

Accuracy of the proposed ensemble model is highlighted to emphasize on the highest accuracy value with respect to the other techniques that have been utilized for binary classification

Table 4 Performance evaluation of different deep learning techniques on the multi-class dataset

State-of-the-art	Accuracy (%)
Dark COVID-Net (Ozturk et al. 2020)	87.02
MobileNetV2 (Apostolopoulos and Mpesiana 2020)	94.72
CORO-Net (Khan et al. 2020)	95
CNN + LSTM (Islam et al. 2020)	97
Covid-Net (Wang and Wong 2020)	93.3
Proposed ensemble model	99.21

Accuracy of the proposed ensemble model is highlighted to emphasize on the highest accuracy value with respect to the other techniques that have been utilized for three-class classification

better generalization as well as by minimizing the false predictions achieving an overall classification accuracy of 96.15%.

Table 4 shows the performance comparison of the proposed technique for multi-classification with five state-of-the-art techniques. Although the false predictions in MobileNetV2 (Apostolopoulos and Mpesiana 2020) and Dark COVID-Net (Ozturk et al. 2020) were slightly lower than the proposed model, however, the model lacked with an approximately 5% and 12% lower accuracy. Islam et al. (2020) proposed a novel approach using CNN as a feature extractor and LSTM for detection and their approach closely follows behind our proposed ensemble framework with an accuracy of 97%. Hence, the proposed ensemble

for the multi-classification task outperformed various other approaches with a final classification accuracy of 99.21%.

5 Conclusions and future scope

In this paper, the ensemble models using deep convolutional neural networks are proposed for the detection of COVID-19 from chest X-ray images. The proposed model not only detect COVID-19 infection from infected persons but is also able to detect COVID-19 from normal and pneumonia (Viral and Bacterial) persons. The proposed models are tested over two well-known datasets. They achieved the accuracies of 99.21% and 96.15% for three class and binary classification, respectively. The proposed model may be leveraged by health officials and researchers to accelerate the prediction of COVID-19 in this global pandemic situation.

There are some limitations of the proposed framework, which can be improved in the future. A limited patient dataset is available that eventually impacts the training and learning capacity of the proposed models. In future, more patient data can be leveraged specifically COVID-19 subjects that can improve the feature extraction capabilities of the proposed model. In the current pandemic situation, a deeper analysis is required for differentiating the COVID-19, viral pneumonia and bacterial pneumonia. The work can also be extended by adding risk and survival prediction of prospective/confirmed COVID-19 patients that can be helpful in the healthcare planning and management strategies.

References

- Abbas A, Abdelsamea MM, Gaber MM (2020) Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. arXiv preprint [arXiv:2003.13815](https://arxiv.org/abs/2003.13815)
- Alshazly H, Linse C, Barth E, Martinetz T (2019) Ensembles of deep learning models and transfer learning for ear recognition. *Sensors* 19(19):4139
- Apostolopoulos ID, Mpesiana TA (2020) Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med* 3:1
- Asraf A (2020) COVID19_Pneumonia_Normal_Chest_Xray(PA)_Dataset, June 29, 2020. <https://www.kaggle.com/amanullahasraf/covid19-pneumonia-normal-chest-xray-pa-dataset>
- Basavegowda HS, Dagnev G (2020) Deep learning approach for microarray cancer data classification. *CAAI Trans Intell Technol* 5(1):22–33
- Basu S, Mitra S (2020) Deep Learning for screening COVID-19 using chest X-Ray Images. arXiv preprint [arXiv:2004.10507](https://arxiv.org/abs/2004.10507)
- Cellina M, Orsi M, Toluian T, Valenti Pittino C, Oliva G (2020) False negative chest X-Rays in patients affected by COVID-19 pneumonia and corresponding chest CT findings. *Radiography* 26(3):e189–e194
- Chouhan V, Singh SK, Khamparia A, Gupta D, Tiwari P, Moreira C, Damaševičius R, De Albuquerque VHC (2020) A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Appl Sci* 10(2):559
- Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam KR et al (2020) Can AI help in screening viral and COVID-19 pneumonia? arXiv preprint [arXiv:2003.13145](https://arxiv.org/abs/2003.13145)
- Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T et al (2020) Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 25(3):2000045
- Coronavirus Disease (COVID-19) (2020) Events as They Happen (WHO). <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>. Accessed 15 July 2020
- COVID-19 symptoms (2020) https://www.who.int/health-topics/coronavirus#tab=tab_3. Accessed 15 July 2020
- Darshan (2020) COVID-19 detection X-ray dataset. March 22, 2020. <https://www.kaggle.com/darshan1504/covid19-detection-xray-dataset>
- Das D, Santosh KC, Pal U (2020a) Truncated inception net: COVID-19 outbreak screening using chest X-rays. *Phys Eng Sci Med* 43:915–925
- Das NN, Kumar N, Kaur M, Kumar V, Singh D (2020b) Automated deep transfer learning-based approach for detection of COVID-19 infection in chest X-rays. *IRBM*. <https://doi.org/10.1016/j.irbm.2020.07.001>
- Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. Ieee, pp 248–255
- Ghosh S, Shivakumara P, Roy P, Pal U, Lu T (2020) Graphology based handwritten character analysis for human behaviour identification. *CAAI Trans Intell Technol* 5(1):55–65
- Gupta B, Tiwari M, Lamba SS (2019) Visibility improvement and mass segmentation of mammogram images using quantile separated histogram equalisation with local contrast enhancement. *CAAI Trans Intell Technol* 4(2):73–79
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
- Hemdan E, Shouman M, Karar M (2020) COVIDX-Net: a framework of deep learning classifiers to diagnose COVID-19 in X-ray images. <https://arxiv.org/abs/2003.11055>
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
- Islam J, Zhang Y (2018) Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Informat* 5(2):2
- Islam MZ, Islam MM, Asraf A (2020) A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Inform Med Unlocked* 20:100412
- Khan AI, Shah JL, Bhat MM (2020) Coronet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Comput Methods Progr Biomed* 196:105581
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
- Loey M, Smarandache F, Khalifa NEM (2020) Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning. *Symmetry* 12(4):651
- Luz EJS, Silva PL, Silva R, Silva L, Moreira G, Menotti D (2020) Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray Images. *CoRR*. [arXiv:2004.05717](https://arxiv.org/abs/2004.05717)
- Maghdid HS, Asaad AT, Ghafoor KZ, Sadiq AS, Khan MK (2020) Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms. arXiv preprint [arXiv:2004.00038](https://arxiv.org/abs/2004.00038)

- Mahmud T, Rahman MA, Fattah SA (2020) CovXNet: a multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. *Comput Biol Med* 122:103869
- Narin A, Kaya C, Pamuk Z (2020) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv preprint [arXiv:2003.10849](https://arxiv.org/abs/2003.10849)
- Oh Y, Park S, Ye JC (2020) Deep learning covid-19 features on cxr using limited training data sets. *IEEE Trans Med Imaging* 39(8)
- Osterland S, Weber J (2019) Analytical analysis of single-stage pressure relief valves. *Int J Hydromech* 2(1):32–53
- Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR (2020) Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput Biol Med* 121:103792
- Rahimzadeh M, Attar A (2020) A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. *Informat Med Unlocked* 19:100360
- Sethy PK, Behera SK (2020) Detection of coronavirus disease (covid-19) based on deep features. Preprints 2020030300:2020
- Shelke A, Inamdar M, Shah V, Tiwari A, Hussain A, Chafekar T, Mehendale N (2020) Chest X-ray classification using Deep learning for automated COVID-19 screening. *Radiol Imaging*. <https://doi.org/10.1101/2020.06.21.20136598>
- Shen D, Guorong W, Suk H-I (2017) Deep learning in medical image analysis. *Annu Rev Biomed Eng* 19:221–248
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
- Siwek K, Osowski S, Szupiluk R (2009) Ensemble neural network approach for accurate load forecasting in a power system. *Int J Appl Math Comput Sci* 19(2):303–315
- Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence
- Toğaçar M, Ergen B, Cömert Z (2020) COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. *Comput Biol Med* 121:103805
- Tuncer T, Dogan S, Ozyurt F (2020) An automated residual exemplar local binary pattern and iterative ReliefF based corona detection method using lung X-ray image. *Chemometr Intell Lab Syst* 203:104054
- Waheed A, Goyal M, Gupta D, Khanna A, Al-Turjman F, Pinheiro PR (2020) Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection. *IEEE Access* 8:91916–91923
- Wang L, Wong A (2020) COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. arXiv preprint [arXiv:2003.09871](https://arxiv.org/abs/2003.09871)
- Wang H, Xia Y (2018) Chestnet: A deep neural network for classification of thoracic diseases on chest radiography. arXiv preprint [arXiv:1807.03058](https://arxiv.org/abs/1807.03058)
- Wang S-H, Phillips P, Sui Y, Liu B, Yang M, Cheng H (2018) Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *J Med Syst* 42(5):85
- Wang R, Yu H, Wang G, Zhang G, Wang W (2019) Study on the dynamic and static characteristics of gas static thrust bearing with micro-hole restrictors. *Int J Hydromech* 2(3):189–202
- Wiens T (2019) Engine speed reduction for hydraulic machinery using predictive algorithms. *Int J Hydromech* 2(1):16–31
- Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J (2020) Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. *Radiology* 1:200343–200343
- Yu X, Kang C, Guttery DS, Kadry S, Chen Y, Zhang YD (2020) ResNet-SCDA-50 for breast abnormality classification. In: *IEEE/ACM transactions on computational biology and bioinformatics*
- Zhang J, Xie Y, Li Y, Shen C, Xia Y (2020) Covid-19 screening on chest x-ray images using deep learning based anomaly detection. arXiv preprint [arXiv:2003.12338](https://arxiv.org/abs/2003.12338)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.