A novel and efficient deep learning approach for COVID-19 detection using X-ray imaging modality

Prashant Bhardwaj | Amanpreet Kaur

Electronics and Communication, Thapar University, Patiala, India

Correspondence
Prashant Bhardwaj, Electronics and Communication, Thapar University, Patiala, Punjab, India.
Email: pbhardwaj_phd19@thapar.edu

Abstract
With the exponential growth of COVID-19 cases, medical practitioners are searching for accurate and quick automated detection methods to prevent Covid from spreading while trying to reduce the computational requirement of devices. In this research article, a deep learning Convolutional Neural Network (CNN) based accurate and efficient ensemble model using deep learning is being proposed with 2161 COVID-19, 2022 pneumonia, and 5863 normal chest X-ray images that has been collected from previous publications and other online resources. To improve the detection accuracy contrast enhancement and image normalization have been done to produce better quality images at the pre-processing level. Further data augmentation methods are used by creating modified versions of images in the dataset to train the four efficient CNN models (Inceptionv3, DenseNet121, Xception, InceptionResNetv2) Experimental results provide 98.33% accuracy for binary class and 92.36% for multiclass. The performance evaluation metrics reveal that this tool can be very helpful for early disease diagnosis.

KEYWORDS
deep learning models, Matthews correlation coefficients, simple averaging, weighted averaging

1 | INTRODUCTION
Severe Acute Respiratory Syndrome 2 (SARS-CoV-2), the name given by the International Committee on Taxonomy of Viruses (ICTV), instigates the novel coronavirus or COVID-19. Recently COVID-19 has become a global health emergency that has appeared in Wuhan, China in late December 2019. Coronavirus is one of the significant pathogens that fundamentally focus on the human respiratory framework and is highly infectious to the health of individuals. The virus later transformed into a worldwide pandemic, as declared by World Health Organization (WHO). In the primary stage of disease transmission, the number of people affected by the disease was minimum; it did not imitate threats of such huge capacity. With the slow progression of time, the virus spread with an extremely high-risk potential of affecting millions of lives in all countries, and become the principal factor of many people’s death globally. Due to high mortality rate by COVID-19 cases worldwide, countries with more number of active cases suggest people to stay indoors and announced a complete lockdown to stop disease transmission. COVID-19 infection is fatal because it is easily transmitted through direct or indirect contact with the affected individual and its symptoms include fever, dry cough, vomiting, diarrhea, and myalgia as indicated by WHO and Centre for Disease and Control (CDC). Till date, May 2021, the COVID-19 pandemic already contributed to over 3 170 882 deaths and more than 150 779 711 confirmed cases of COVID-19 infection. Researchers are...
actively participating to detect the COVID-19 positive cases and also finding the diagnosis procedure and medical treatment of affected patients rapidly. Daily increment of positive COVID-19 cases and wrong diagnosis is challenging in managing the pandemic. With large number of infected individuals and less number of test kits, practitioners totally rely upon automated detection system to combat the pandemic proficiently at early stage. These detection systems could also be important in identifying patients who need isolation to prevent the disease from spreading in the community. The laboratory tests used for COVID-19 detection including nucleic acid reagent detection, viral antigen detection were time-consuming and produce a higher false-negative detection rate. Reverse transcription-polymerase chain reaction (RT-PCR) test-kits have emerged as the main technique for diagnosing COVID-19. The current RT-PCR system is time-consuming and also it requires additional resources and approval to detect infected patients, which can be more costly in many developing societies. Due to the inaccessibility of test kits and the false-negative rate of virus and antibody tests, the medical authorities have temporarily used radiological examinations as a clinical investigation for COVID-19 detection but there are limited kits for detecting the virus efficiently. These issues have posed a great life threat, especially in societies with limited medical assets. However, noting the recent spread of COVID-19, the researchers need to look for other better options like X-ray and computed tomography (CT) scans. The information obtained from radiological images, that is, X-ray and CT scans is important for clinical diagnosis. Radiological images have high sensitivity in disease diagnosis and more accurate than RT-PCR. Thus, radiologists can notice the characteristics of a lung infected with COVID-19 by using chest X-ray and CT scan. CT scan causes multiple problems of healthcare services that are coupled with global based features for classifying COVID-19 subjects from other classes and reduce the misdiagnosis rate for COVID-19 and help with lower COVID-19 subjects, the severity of the disease is not properly diagnosed. A large medical image dataset is used in our work and the proposed model shows promising results in terms of accuracy and other evaluation metrics (precision, recall, and F-1 score).

- We develop an ensemble learning based system using deep convolutional neural network which is trained and evaluated on publically available chest X-ray image dataset having the knowledge to classify between Covid vs. normal vs. pneumonia subjects. The limited number of COVID-19 subjects used by many researchers in their work leads to degrading the model efficiency as with lower COVID-19 subjects, the severity of the disease is not properly diagnosed. A large medical image dataset is used in our work and the proposed model shows promising results in terms of accuracy and other evaluation metrics (precision, recall, and F-1 score).
- We have used four (Inceptionv3, DenseNet121, Xception, and InceptionResNetv2) powerful and efficient DL models with more number of training parameters that are coupled with global based features for classifying COVID-19 subjects from other classes and reduce the misdiagnosis rate for COVID-19 and help
the doctors, field specialists, and physician to know the severity of disease at early stage.

• Data augmentation methods are employed for COVID-19 detection to avoid overfitting problems. We have fine-tuned our model by using four pre-trained models in depth to compensate for the loss of valuable information.

2 | RELATED WORK

Recently, there have been numerous AI learning-based methods that were used for COVID-19 detection using radiographic data. In the medical care field, DL which is a sub-branch of AI is in effect and gradually developed as a Computer-Aided Design (CAD) tool to help doctors/ radiological experts for better disease prediction. DL methods can be a guide for professionals to improve the quality of COVID-19 detection. Already many methods can be a guide for professionals to improve the quality of COVID-19 detection such as COVID-Net, CoroNet, and COVIDNet. A chest X-ray image dataset of 5907 images. Approximately 500 Covid images were used in their work. Their model achieved an accuracy of 98% for multiclass problem. Kesim et al. proposed a CNN-based model for the classification using chest X-ray image. A chest X-ray image dataset from 12 categories has used, with an accuracy rate of 86% reported in experimental results. Sedik et al. proposed a CNN and long short term memory (LSTM) based model named ConvLSTM. Two different datasets were used in their study. The proposed model was tested on two different imaging modalities X-ray and CT scans. The model has been evaluated on two different scenarios (Covid vs. normal and Covid vs. pneumonia) and achieved accuracy of 100% obtained.
Bhandary et al.\textsuperscript{49} modified the AlexNet model to detect lung abnormalities based on chest X-ray images. More specifically, the authors used a DL approach to screen for pneumonia. A new “threshold filter” has been presented and a feature ensemble strategy has also been defined that produced a 96% classification accuracy rate. Chouhan et al.\textsuperscript{7} presented five new deep-transfer-learning-based models applied as an ensemble to detect pneumonia in chest X-ray images. The authors reported an accuracy score of 96.4% using their developed ensemble deep model. Kumar et al.\textsuperscript{50} utilizes ResNet152 pre-trained model with XGBoost classifiers and evaluated it on a 3-class problem containing 1341 normal images, 1345 pneumonia images, and only 62 COVID-19 images with 30% holdout. Their model achieved an accuracy of 97.3%. Masud et al.\textsuperscript{51} proposed an Internet of Medical Things (IoMT) based lightweight security model to help the medical practitioners for COVID-19 detection in a more efficient way. In their work strength of Mutual Authentication and Secret Key (MASK) have been evaluated to prevent the physical attacks and increase the computational efficiency. Öksüz et al.\textsuperscript{52} proposed a DL model using three pre-trained CNNs (SqueezeNet, ShuffleNet, and EfficientNet-B0). The total of 2905 CXR samples, including 1345 samples for pneumonia, 219 for COVID-19, and 1341 for normal samples were used in the study. The classification accuracy of 98.30 has been achieved for multiclass problem. Li et al.\textsuperscript{22} proposed an automatic system named as COVNet using DL methods for COVID-19 detection. The pre-trained CNN architecture named ResNet50 was used in the study. A total of 4536 chest CT samples, including 1296 samples for COVID-19, 1735 for pneumonia, and 1325 for normal samples were used in the study. The dataset has been divided into two parts, that is, 90% for training and 10% for testing respectively. The experimental result revealed that the system obtained a sensitivity of 90%, AUC of 96%, and specificity of 96% for COVID-19 cases.

3. The need for GPU resources to train newly designed CNNs or pre-trained CNNs. Also, the deeper features of COVID-19 are not known to be separated, when a greater number of diseases are being considered.

Considering all those factors, we can say that there is a need for effective detection system for COVID-19 detection. As the spread of this deadly virus has increased demonically and taken the lives of many people residing in different countries. The current COVID-19 screening method is RT-PCR. This is the first method that is used by many practitioners/doctors to diagnose COVID-19. But the problem with this method is that it is very time-consuming and the results obtained from this method take few days to weeks. This creates a major challenge with less equipped clinics or hospitals. When compared to PCR techniques, chest radiographic imaging (CRI) methods have many advantages, that is, they are easily available and cheap. The use of radiographic imaging systems plays an important role in those areas where appropriate test kits are not available. Quality of radiographic images depends on the digital devices.\textsuperscript{53} Also chest radiograph technique plays a great role as an image retrieval model based on deep metric learning, where images of the same contents are pulled together.\textsuperscript{54} This technique has incredible clinical value for the treatment and management of COVID-19 patients. With the improvement of this model visual saliency-guided complex image retrieval model can be utilized to get the image patterns more clearly.\textsuperscript{55} CRI methods include X-ray and CT scans. In a small number of tests, CT has been used to analyze and detect features of COVID-19 with more clarity, that is, due to its high-resolution value for lung uniformity and ground-glass opacity. However, due to more cost and less availability of CT machines in rural areas, it may not be a good choice.\textsuperscript{19} Whereas, an X-ray test can be considered an ideal solution to detect COVID-19, as it is more available at a lower cost. But, it can be a challenging task for a radiologist for discovering X-ray images to differentiate between community-acquired pneumonia (CAP), COVID-19, and other lung-related diseases. Due to the increased rush of patients in hospital emergency rooms (ERs), accurate disclosure of radiographic data is mandatory, as it can save a lot of time. In the current section, we address the above-mentioned problem and present an approach to handle this problem more effectively.

4. MATERIALS AND METHODOLOGY

Ensemble-based DL model is proposed for COVID-19 detection using four pre-trained architecture as
Ensemble learning\textsuperscript{56} is a newly emerging field of AI including ML and DL.\textsuperscript{57,58} From the last few decades, ensemble system also known as multiple classifier system has gained everyone's attention in the field of AI. As these systems have proven to be very effective to solve many real-world problems like in healthcare field and many computer vision problems. These systems combine the features of multiple models to boost the overall efficiency of the system by reducing system error. The different models adapt diverse features of each other and are grouped to make more accurate predictions using classifiers. Ensembling is a two-way process. In the first step, deep features of a network are extracted using pre-trained architectures. In the next step, classifiers are used to make accurate predictions. The ensemble network always gives more accurate results as compared to a single model.\textsuperscript{56} Several classifiers like decision tree, K-nearest neighbor, SVM, auto encoders, Boltzmann Machine (BM), etc. used in many ML and DL models to improve model efficiency. The main motivation for using the classifier is to extract deep features from the image that may lead to improve the model performance. It uses various algorithms to train the data set and make the final prediction based on the clustering. The concept of voting technique was used in this research work. The voting group collects the decisions of many classifiers and performs a particular classification task; provides flexibility in clustering strategies so that maximum possible classification accuracy is obtained.\textsuperscript{59} Voting techniques can be divided into two categories: hard and soft voting.\textsuperscript{60} In hard voting method the class labels of the test samples are absolute by the majority voting method. Each base classifier independently gives a class label to a given test sample during the testing phase. The final grouping of the test sample is determined by the maximum number of times a particular class label is assigned to that test sample.\textsuperscript{61} On the other hand, soft voting methods calculate the average probability of all classes, and the final prediction is made on the basis that which class is having the highest probability.\textsuperscript{60} As these techniques do not use any algorithm for combining predictions from base classifiers as required in the stacking set,\textsuperscript{61} this makes them a good choice to be used in ensemble models.

The layout of the proposed methodology is shown in Figure 1. As seen from Figure 1 powerful CNN model (inceptionv3, inceptionresnetv2, densenet121, and xception) with more number of trainable parameters is used to fine-tune the proposed model and simply extracts
more features in depth. The detailed summary of the proposed method is given in the following section.

4.1 Dataset collection

In this study, CRI based dataset is used. The dataset consists of 10 000 X-ray images in Portable Network graphics (PNG) format. The resolution of each image is set to $224 \times 224 \times 3$. The different chest radiographs are combined into one dataset which can be used for classification purposes. Pneumonia dataset was taken from kaggle repository, whereas COVID-19 and non-COVID dataset was taken from previous publications and online available resources. The final dataset contains overall of 10 000 CXR images in which 2022 sample belongs to pneumonia class, 2161 to COVID-19 and 5863 to non-COVID class. In this study, to develop a robust and deep efficient model to perform classification task, we employed the use of 5-cross validation. To perform this cross-validation the overall samples are distributed into three sets; training (80%), testing (10%), and validation (10%). The dataset distribution for different classes is shown in Table 1.

4.2 Data pre-processing and augmentation

Pre-processing methods can be valuable for eliminating undesirable noise present in the given image. In the present work, contrast enhancement and image normalization method is applied as shown in Figure 2A used to change pixel intensity value for acquiring a better-enhanced image. By changing the pixel intensity, hidden information that exists within the low range of gray level image is revealed. On the other hand, data augmentation methods were also employed to augment training samples by creating diverse data without losing useful information. The main reason to use different image augmentation methods is that it increase the overall performance of the system by adding more diverse data to the limited dataset. Data augmentation methods including Image rotation,

| Class label | Number of samples | Training | Testing | Validation | Modality |
|-------------|-------------------|----------|---------|------------|----------|
| Pneumonia   | 2022              | 1617     | 203     | 202        | X-ray    |
| COVID-19    | 2161              | 1728     | 216     | 217        |          |
| Non-COVID   | 5863              | 4690     | 586     | 587        |          |

TABLE 1 Distribution of data for all classes

FIGURE 2 (A) Data pre-processing and (B) data augmentation
image scaling, flipping, and translation were applied to the original dataset as shown in Figure 2B. These acts as a stabilizer and reduce over-fitting problems while training our DL model.

4.3 | Neural networks

In the previous many years, rise of deep convolutional networks has brought a big breakthrough in many filed of image processing including computer vision and machine vision tasks. Deep neural networks are used to extract deep features or the most valuable information from an input data. These types of networks are used to solve big data problem and usually trained on a higher dataset. Recently, these networks made a great impact in many fields including industries and healthcare.62 Most preferred and usable CNN are DenseNet,63 Xception,26 Inception,64 and Resnet.43 These are the pre-trained neural networks and already trained in ImageNet dataset. The features extracted from them are transferrable to a newly designed model. These networks are very useful while training a new architecture from scratch, as the weights used in these pre-trained architectures can further be used in a newly designed architecture. There are many inputs, hidden and output layers present in a network. Input layers are used to give input data to any model and hidden layers are used to extract the important features, whereas output layers are used to make the final classification. In our work, we employed four different pre-trained models, InceptionV3, DenseNet121, InceptionResNetV2, and Xception for COVID-19 detection.

4.3.1 | Inception v3 architecture

A CNN architecture that is widely used for image recognition problems. Inception v3 has 24 million parameters and achieved a good accuracy on ImageNet dataset. This network used a factorization method for computation and provides more accurate results. Keras is the prime host of Inceptionv3 network and plays an important role in its structure development. The main motive behind the use of this network is to avoid representational bottlenecks, that is, reducing the input dimensions of the next layer. Different convolution and pooling layer used in the inceptionv3 architecture is shown in Figure 3.

4.3.2 | DenseNet121 architecture

DenseNets are called the densely connected CNN’s. These networks generally alleviate the vanishing gradient problem and decrease the parameter usage to extract more features. In our study we use DenseNet121 architecture. It is a CNN architecture with 121 deep layers. More number of layers used in the network makes the network very deep but at the same time makes it more effective by using shorter connections between layers. Maximum information can be shared through these connections. It works in a hierarchical manner as each layer of this network is connected to its alternate layers, that is, the first layer $j_1$ is connected to $j_2, j_3, j_4$, and the second layer $j_2$ to $j_3, j_4, j_5$, and so on. $(j - 1)$th layer of the activation map is considered as the input of $j$th layer. But one thing

![Inception v3 architecture](image)

FIGURE 3 Inception v3 architecture
needs to be taken out the input convolutional layer must be equal to output convolutional layer as shown in Figure 4. Mathematically, it can be given as:

\[ Y_j = W_j ([Y_0, Y_1 \ldots Y(j - 1)]) \]

where \( W_j \) represents the input \( j \)th layer and \( Y_j \) represents the output of \( j \)th layer. These networks use a concatenation process for combining the features of a network. Transition layer and batch normalization functions are used to stabilize the learning process.

4.3.3 | Inceptionresnet v2 architecture

Inceptionresnet v2 is a pre-trained CNN architecture that builds on the inception family but includes residual connections. This network consists of 56 million parameters with 164 deep layers and obtained better results for ImageNet dataset. They are generally used for image classification, image segmentation, and object detection. The layout of the inception resnetv2 architecture used in this study is shown in Figure 5.

4.3.4 | Xception architecture

Xception is a CNN architecture that was firstly used in 2017. Xception network is the modified version of inception and ranked third on ImageNet dataset challenge. Instead of conventional convolution, this network uses depth wise convolution layers which involves mapping of cross-channel and spatial correlations. This network consists of 22.9 million parameters and used for many image classification and object detection problems. The main motive to design this architecture is to create a network with more parameters that can be used to solve any computer network (CN) problem. The layout of the xception architecture used in this study is shown in Figure 6.

5 | EXPERIMENTAL ENVIRONMENT

Three-class scenarios are employed for COVID-19 detection from CRIs. For each scenario, four CNN pre-trained models are used for making final predictions. All experimental illustrations were carried out on a terminal
having specifications: Intel Core i3, 5005U cpu@2.00ghz, 8 GB RAM. Google collaboratory is an online clouding software that is used for simulation by using different Python's libraries. In this research work, two python libraries keras and tensorflow are used. Graphs for experimental results are obtained using python's matplotlib library. While training our model we have resized the image samples to [224, 224] so that all the samples are consistent in terms of size.

The value for test dataset was kept changing with new input images while the training and validation dataset was kept constant. As mentioned in Table 2, 5-cross validation is performed on training and testing dataset using stochastic gradient descent (SGD) as an optimizer function with a learning rate of 0.001. All the pre-trained CNNs are trained with input (CXR) images for many epochs to stabilize the loss function. The test dataset was used to get the final classification accuracy of the proposed model.

5.1 | Evaluation metrics

The proposed model performance is illustrated in the form of confusion metrics (CM) and receiver operating characteristics (ROC). Accuracy, precision, recall, MCC, F-score are also used to evaluate the performance of network and are given as:

Accuracy: It is the most important performance measure. It can be calculated as a ratio of true predicted observation with respect to the total observations. Accuracy of the system can be shown mathematically by using the formula:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP},
\]

where TP denotes the true positive labels, TN denotes the true negative labels, FN denotes the false negative labels, and FP denotes the false positive labels.

Precision: It is the ratio of true predicted positive observations with respect to the total predicted positive observations

\[
\text{Precision} = \frac{TP}{TP + FP}.
\]

Recall: It is the ratio of true predicted positive observations with respect to the total observations in both positive and negative class. Mathematically, it can be given as:

\[
\text{Recall} = \frac{TP}{TP + FN}.
\]

F-1 score: It can take weighted average value of precision-recall. Mathematically it can be calculated using the formula:

\[
F-1\text{ score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{Precision} + \text{recall}}.
\]

MCC: It stands for Matthew's correlation coefficient. It is a single value metric that reviews the confusion matrix. It consists of four different entries: true negatives (TN), true positive (TP), false negatives (FN), and false positives (FP). It defines a relationship between actual and predicted classes. Mathematically, MCC can be given as:

\[
\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}.
\]
PROPOSED SYSTEM RESULTS

This section provides testing and performance analysis to verify the efficiency suggested technologies to detect COVID-19. The calculation of different evaluation parameters for three-class scenarios is presented in Table 3. The table shows the value of Input image, number of trainable parameters, MCC values for all the pre-trained CNN architecture used in this study. For binary class xception model has the higher MCC value of 96.4%, whereas for multiclass classification inceptionresnetv2 models show a high value of 91.2%. The proposed model achieved the highest precision, recall, and F-score value for Covid class. It is indicated in Table 3 that for binary and multiclass problem inceptionresnetv2 model shows a promising accuracy of 96.28% and 88.19% respectively using ensembling methods. Five-fold cross-validation method is employed to examine the deep CNN architectures to boost the overall performance of the proposed model. The complete CXR image dataset is split into three sets as mentioned above to avoid the problem of overfitting. The value of each individual model is calculated for 5-folds is averaged and the mean value is used to calculate other evaluation metrics. After performing 5-fold cross-validation, the performance evaluation metrics for binary and multiclass is achieved as in Table 4. The proposed ensemble system for COVID-19 detection achieved highest overall accuracy of 98.45% and 92.36 for binary and multiclass respectively using 5-fold cross-validation.

DISCUSSION

The trials to recognize and classify COVID-19 cases using X-ray imaging modality are divided into two different setups. At first, the ensemble model is trained to identify binary class problems including COVID-19 and non-COVID sets. Secondly, the ensemble DL model is trained to classify multiclass problems including COVID-19, non-COVID, and pneumonia. The effectiveness of the offered model is calculated using different ensembling
techniques for both binary and multiclass classification problems. The dataset has been divided into three categories: 80% of X-ray image data utilized for training, 10% for testing, and 10% data for validation. Figure 7A,C represents the confusion matrix for binary and multiclass respectively using simple averaging approach. Whereas Figure 7B,D represents the confusion matrix for multiclass problem using weighted ensembling technique. The confusion matrix is an \( \frac{N}{C \times N} \) matrix that is used to calculate the overall proficiency of the suggested model, where \( N \) represents the number of target groups. The matrix compares the actual target values (True label) with the predicted values. Horizontal axis denotes the value of true label and vertical axis denotes the value of predicted level for different classes: 0 (COVID), 1 (normal), and 2 (pneumonia). True label are the values which already sets to true, that is, there will no change in these during the classification task, but predicated label are used to make predictions. The numerical values in the confusion matrix indicate the value for TP, FP, FN, and FP. These labels are used to calculate different evaluation metrics of the proposed model. On the other hand, Figure 7E,G shows the precision-recall curve for binary class problem and also the relationship between different threshold values of precision and recall. Similarly Figure 7F,H represents the precision-recall curve for multi class problem. Different colors were used to represent the area under the curve for different classes. Maroon color denotes the area under the curve for class 0 (COVID), blue color denotes class 1 (normal), and green color denotes class 2 (pneumonia). The maximum area under the curve represents higher precision and recall value. The maximum precision value shows low false positives, whereas the maximum recall value specifies low false negatives. It can be seen from the graphs that there is a substantial growth in loss values at the initial stage of training. In the advanced stage of training, these loss values decrease substantially. However, the deep model inspects all the X-ray images

| Model            | Input size | No. of layers | No. of parameters in million | Class | MCC  | MSE  | Mean squared log error |
|------------------|------------|---------------|------------------------------|-------|------|-----|------------------------|
| Xception         | 224 × 224 × 3 | 48            | 24                           | 2     | 0.964 | 0.017 | 0.0085                 |
|                  |            |               |                              | 3     | 0.889 | 0.269 | 0.0825                 |
| DenseNet 121     | 224 × 224 × 3 | 121           | 1                            | 2     | 0.754 | 0.128 | 0.0617                 |
|                  |            |               |                              | 3     | 0.853 | 0.267 | 0.0771                 |
| Inception v3     | 299 × 299 × 3 | 164           | 56                           | 2     | 0.962 | 0.019 | 0.0091                 |
|                  |            |               |                              | 3     | 0.870 | 0.299 | 0.0915                 |
| Inception ResNet v2 | 299 × 299 × 3 | 170           | 22.9                         | 2     | 0.952 | 0.023 | 0.0114                 |
|                  |            |               |                              | 3     | 0.912 | 0.138 | 0.0403                 |

| Table 3 Calculation of evaluation metrics |
|-------------------------------------------|

| Model         | Input size | No. of layers | No. of parameters in million | Class | MCC  | MSE  | Mean squared log error |
|---------------|------------|---------------|------------------------------|-------|------|-----|------------------------|
| Xception      | 224 × 224 × 3 | 48            | 24                           | 2     | 0.964 | 0.017 | 0.0085                 |
| DenseNet 121  | 224 × 224 × 3 | 121           | 1                            | 2     | 0.754 | 0.128 | 0.0617                 |
| Inception v3  | 299 × 299 × 3 | 164           | 56                           | 2     | 0.962 | 0.019 | 0.0091                 |
| Inception ResNet v2 | 299 × 299 × 3 | 170           | 22.9                         | 2     | 0.952 | 0.023 | 0.0114                 |

| Table 4 Classification results for binary and multiclass after 5-fold cross-validation |
|-------------------------------------------|

| CNN model       | Class     | Accuracy (%) | Precision | Recall   | F-1 score |
|-----------------|-----------|--------------|-----------|----------|-----------|
| Xception        | Binary    | 95.05        | 0.9949    | 0.9680   | 0.9813    |
|                 | Multiclass| 85.30        | 0.8312    | 0.9704   | 0.8955    |
| DenseNet 121    | Binary    | 94.53        | 0.8130    | 0.9532   | 0.8776    |
|                 | Multiclass| 86.05        | 0.8457    | 0.9852   | 0.9101    |
| Inception v3    | Binary    | 95.13        | 0.9924    | 0.9680   | 0.9800    |
|                 | Multiclass| 82.33        | 0.8098    | 0.9754   | 0.8849    |
| Inception ResNet v2 | Binary    | 96.28        | 0.9949    | 0.9557   | 0.9749    |
|                 | Multiclass| 88.19        | 0.9207    | 0.9729   | 0.9461    |
| Ensemble 1 (simple averaging) | Binary | 98.45 | 0.9975 | 0.9704 | 0.9838 |
|                 | Multiclass| 91.74        | 0.8725    | 0.9606   | 0.9144    |
| Ensemble 2 (weighted averaging) | Binary | 98.33% | 0.9975 | 0.9680 | 0.9825 |
|                 | Multiclass| 92.36%       | 0.8772    | 0.9680   | 0.9204    |
present in the given dataset again and again during training. Our experiments are designed to assess the effect of ensembling techniques on the accuracy of COVID-19. Therefore the execution is carried out by different ensembling techniques, that is, simple averaging and weighted averaging detection.
7.1 | Receiver operating curve

The ROC curve is used as another rating scale to provide an accurate visualization of simulation results. Within the ROC curve, the TP labels are symbolized as a utility of the FP labels at distinct cut-off points. No intersection in the given two distributions indicates that the ROC curve crosses through the upper left corner. Hence, from Figure 8 it can be the closest the ROC curve is to the upper left corner, the greater the efficiency of the given system.

Figure 8A,C represents the value of ROC curve using simple averaging technique and Figure 8B,D represents the value of ROC curve using weighted averaging approach. Different color notations: purple for class 0 (COVID-19), sky blue for class 1 (normal), and yellow for class 2 (pneumonia) were used to mark the curve lines. And it is clear from Figure 8A,B that the area value for binary classes is more as compared to multiple classes. The more the area of the ROC curve, the better the model performance.

8 | COMPARISON WITH STATE-OF-THE-ART METHODS

In this section, we make assessments of the pre-trained DL architectures that have been proposed to detect COVID-19 to date against our proposed ensemble model. In Table 5 some of the latest approaches used for COVID-19 detection are considered. Ozturk et al. presented a DL model for COVID-19 detection using X-ray imaging modality. From the experimental results, it was seen that accuracies of 98.08% and 87.02% were achieved respectively for binary and multiclass problems. Bhandary et al. improved the AlexNet model for the detection of lung anomalies using chest X-ray images. The authors used a DL approach for the detection of pneumonia. A new “threshold filter” has
also been introduced with ensemble learning which produced a 96% classification accuracy rate. Sethy and Behera proposed a model for multiclass problem including 127 (COVID-19, pneumonia [viral and bacterial], and normal) images. The dataset has been divided into two parts, that is, 80% for training and 20% for testing. ResNet-50 model was used in their work with SVM classifier and an accuracy of 95.33% was obtained. Gianchandani et al. proposed a modified DL model using four pre-trained CNNs (DenseNet201, ResNet152V2, InceptionResNetV2, and VGG16). Two different kinds of image datasets D1 and D2 have been employed for multiclass problem. For binary and multiclass, accuracies of 96% and 99% have been obtained respectively. Rahimzadeh and Attar proposed a DL model based on two CNN pre-trained architectures Xception and resnet50v2. The dataset used in the study includes 15 085 (COVID-19 = 180, normal = 8851, pneumonia = 6054) images. Multiple features were extracted to boost the overall performance of the proposed model. The classification accuracy of 91% has been obtained for multiclass problem. From Table 5, it is clear that the projected approach has worked well in terms of accuracy and other evaluation metrics on a higher chest X-ray image dataset when compared with other DL approaches.

### 9 | CONCLUSION

The recent epidemic has changed human lives to an extraordinary extent and has become a global health problem. Though the attempt of the academic community has been tremendous through different fronts, the virus progression occurs at a rapid rate. Different DL and ML algorithm has been developed to diagnose the virus at an early stage. Because COVID-19 is highly infectious, it is important to control its transmission path effectively to prevent the spread of the disease. In the proposed work, we present a deep ensemble learning architecture for COVID-19 detection using four different pre-trained deep neural network architectures (inceptionv3, densenet121, inceptionresnetv2, and xception). The model has been trained on CXR images to check the robustness of the proposed model. Data augmentation methods have been employed, in order to avoid the problem of limited dataset. We also validate our model using 5-fold cross-validation for binary and multiclass problems including three scenarios: COVID-19, pneumonia, and normal and obtained the best results. At last, we can conclude that this technique can be a valid tool that can be helpful for doctors and researchers to detect COVID-19 efficiently.

### ACKNOWLEDGMENTS

I would like to express my special thanks of gratitude to my research supervisor, Dr. Amanpreet Kaur, Assistant Professor, Department of Electronics and Communication Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, for giving me the opportunity to do research and providing invaluable guidance throughout this research. Her sincerity and motivation have deeply inspired me. It was a great privilege and honor to work and study...
under her guidance. I would also like to thank all the teachers and head of department and Dr. Alpana Aggarwal for guiding me in my research. At last I am grateful to my parents for their love, sacrifices for educating and preparing for my future.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Prashant Bhardwaj https://orcid.org/0000-0002-9496-5326

REFERENCES
1. Burki TK. Coronavirus in China. Lancet. 2020;8(3):238. https://doi.org/10.1016/S2213-2600(20)30056-4
2. Bhattacharya S, Reddy Maddikunta PK, Pham QV, et al. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey. Sustain Cities Soc. 2020;65:102589. https://doi.org/10.1016/j.j.scs.2020.102589
3. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. J Med Virol. 2020;92(4):418-423. https://doi.org/10.1002/jmv.25681
4. Nicola M, O’Neill N, Sohrabi C, Khan M, Agha M, Agha R. Evidence based management guideline for the COVID-19 pandemic. Int J Surg. 2020;77:206-216. https://doi.org/10.1016/j.ijsu.2020.04.001
5. Coronavirus Update (Live): 135,337,870 Cases and 2,929,304 Deaths from COVID-19 Virus Pandemic – Worldometer. (n.d.). https://www.worldometers.info/coronavirus/
6. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA. 2020;323(11):1061-1069. https://doi.org/10.1001/jama.2020.1585
7. Chouhan V, Singh SK, Khamparia A, et al. A novel transfer learning based approach for pneumonia detection in chest X-ray images. Appl Sci. 2020;10(2):10-559. https://doi.org/10.3390/app10020559
8. Norouzifard M, Nemati A, Gholamhosseini H, Klette R, Nouri-Mahdavi K, Yousefi S. Automated glaucoma diagnosis using deep and transfer learning: proposal of a system for clinical testing. In: International Conference Image and Vision Computing New Zealand; November 2019:1-6. https://doi.org/10.1109/IVCNZ.2018.8634671
9. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks. Comput Biol Med. 2020;121:103795. https://doi.org/10.1016/j.compbiomed.2020.103795
10. Pathak Y, Shukla PK, Tiwari A, Stalin S, Singh S. Deep transfer learning based classification model for COVID-19 disease. IRBM. 2020;1:1-6. https://doi.org/10.1016/j.irbm.2020.05.003
11. Lee EYP, Ng MY, Khong PL. COVID-19 pneumonia: what has CT taught us? Lancet. 2020;20(4):384-385. https://doi.org/10.1016/S1473-3099(20)30134-1
12. Zhang Q, Yang LT, Chen Z, Li P. A survey on deep learning for big data. Inf Fusion. 2018;42:146-157. https://doi.org/10.1016/j.inffus.2017.10.006
13. Alam Khan F, Asif M, Ahmad A, Alharbi M, Aljuaid H. Blockchain technology, improvement suggestions, security challenges on smart grid and its application in healthcare for sustainable development. Sustain Cities Soc. 2020;55:102018. https://doi.org/10.1016/j.scs.2020.102018
14. Yu C, Li J, Li X, Ren X, Gupta BB. Four-image encryption scheme based on quaternion Fresnel transform, chaos and computer generated hologram. Multimed Tools Appl. 2018;77(4):4585-4608. https://doi.org/10.1007/s11042-017-4657-6
15. Reddy S. Radiology Perspective of Coronavirus Disease 2019 (COVID-19): Lessons From Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome Melina; October 2020:1-5.
16. Rubin GD. Covid 19 Fleischner society guidelines. Radiology RSNA. 2019;296:1-24.
17. Ai T, Yang Z, Hou H, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology. 2020;296(2):E32-E40. https://doi.org/10.1148/radiol.2020200642
18. Anthony Dangis, Marc Gillis. Accuracy and reproducibility of low dose submillisievert chest CT for the diagnosis of COVID-19. 2010;44:1-11.
19. Jacobi A, Chung M, Bernheim A, Eber C. Portable chest X-ray in coronavirus disease-19 (COVID-19): a pictorial review. Clin Imaging. 2020;64:35-42. https://doi.org/10.1016/j.clinimag.2020.04.001
20. Wang L, Lin ZQ, Wong A. COVID-net: a tailored deep convolutional neural network for detection of COVID-19 cases from chest X-ray images. Sci Rep. 2020;10(1):1-12. https://doi.org/10.1038/s41598-020-76550-z
21. Gunraj H, Wang L, Wong A. COVIDNet-CT: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest CT images. Inf Fusion. 2020;4:1-12.
22. Li L, Qin L, Xu Z, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. Radiology. 2020;296(2):E65-E71. https://doi.org/10.1148/radiol.2020200905
23. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 using deep neural networks with X-ray images. Comput Biol Med. 2020;121:103792. https://doi.org/10.1016/j.compbiomed.2020.103792
24. Ravi D, Wong C, Deligianni F, et al. Deep learning for health informatics. IEEE J Biomed Health Inform. 2017;21(1):4-21. https://doi.org/10.1109/JBHI.2016.2656665
25. Chowdhury MEH, Rahman T, Khandakar A, et al. Can AI help in screening viral and COVID-19 pneumonia? IEEE Access. 2020;8:132665-132676. https://doi.org/10.1109/ACCESS.2020.3010287
26. Khan AI, Shah JI, Bhat MM. CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Comput Methods Programs Biomed. 2020;196;105581. https://doi.org/10.1016/j.cmpb.2020.105581
27. Alzubaidi L, Fadhel MA, Oleiwi SR, Al-Shamma O, Zhang J. DFU_QUTNet: diabetic foot ulcer classification using novel deep convolutional neural network. Multimed Tools Appl.
Hammoudi K, Benhabiles H, Melkemi M, et al. Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19. *arXiv*. 2020;1-6. http://arxiv.org/abs/2004.03399

Hemdan EE-D, Shouman MA, Karar ME. COVIDX-Net: a framework of deep learning classifiers to diagnose COVID-19 in X-ray images. *arXiv*. 2020. http://arxiv.org/abs/2003.11055

Luz E, Silva PL, Silva R, Silva L, Moreira G, Menotti D. Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images. *arXiv*. 2020. http://arxiv.org/abs/2004.05717

Minaei S, Kafieh R, Sonka M, Yazdani S, Jamalipour Soufi G. Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning. *Med Image Anal*. 2020;65:101794. https://doi.org/10.1016/j.media.2020.101794

Ucar F, Korkmaz D. Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information; January 2020.

Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Singh V. Application of deep learning for fast detection of COVID-19 in X-rays using nCOVNet. *Chaos Solitons Fractals*. 2020;138:109944. https://doi.org/10.1016/j.chaos.2020.109944

Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings – 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017; January 2017: 2261-2269. https://doi.org/10.1109/CVPR.2017.243

Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2018:4510-4520. https://doi.org/10.1109/CVPR.2018.00474

Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115(3):211-252. https://doi.org/10.1007/s11263-015-0816-y

Jaeger S, Karargyris A, Candemir S, et al. Automatic tuberculosi screening using chest radiographs. *IEEE*. 2013;33(2):1-13.

Kermanshahi DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018;172(5):1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010

Sethy PK, Behera SK. Detection of coronavirus disease (covid-19) based on deep features. *Preprints*. 2020;2020030050: 2020. https://doi.org/10.20944/preprints202003.0300.v1

Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med*. 2020;43(2):635-640. https://doi.org/10.1016/s1324-020-00865-4

Brunese L, Mercaldo F, Reginelli A, Santone A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. *Comput Methods Programs Biomed*. 2020;196:105608. https://doi.org/10.1016/j.cmpb.2020.105608

Chowdhury A, Shankaran R, Kavakli M, Haque MM. Sensor applications and physiological features in drivers’ drowsiness detection: a review. *IEEE Sens J*. 2018;18(8):3055-3067. https://doi.org/10.1109/JSEN.2018.2807245

Rahimzadeh M, Attar A. 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. *Inf Med Unlocked*. 2020;19:100360. https://doi.org/10.1016/j.imu.2020.100360

Tang S, Wang C, Nie J, et al. EDL-COVID: ensemble deep learning for COVID-19 cases detection from chest X-ray images. *IEEE Trans Inf. Inf*. 2021;17:6539-6549. https://doi.org/10.1109/TII.2021.3057683

Afifi A, Hafsa NE, Ali MAS, Alhumam A, Alsalman S. An ensemble of global and local-attention based convolutional neural networks for COVID-19 diagnosis on chest X-ray images. *Symmetry*. 2021;13(1):1-25. https://doi.org/10.3390/sym13010113

Rafi TH. An ensemble deep transfer-learning approach to identify COVID-19 cases from chest X-ray images. In: 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB; 2020. https://doi.org/10.1109/CIBCB48159.2020.9277695

Kesim E, Dokur Z, Olmez T. X-ray chest image classification by a small-sized convolutional neural network. In: 2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT; 2019:1-5. https://doi.org/10.1109/EBBT.2019.8742050

Sedik A, Hammad M, Abd El-Samie FE, Gupta BB, Abd El-Latif AA. Efficient deep learning approach for augmented detection of Coronavirus disease. *Neural Comput Appl*. 2021;1: 8. https://doi.org/10.1007/s00521-020-05410-8

Bhandary A, Prabhu GA, Rajinikanth V, et al. Deep-learning framework to detect lung abnormality – a study with chest X-ray and lung CT scan images. *Pattern Recognit Lett*. 2020;129:271-278. https://doi.org/10.1016/j.patrec.2019.11.013

Kumar R, Arora R, Bansal V, et al. Accurate prediction of COVID-19 using chest X-ray images through deep feature learning model with SMOTE and machine learning classifiers. *MedRxiv*. 2020:1-10. https://doi.org/10.1101/2020.04.13.20063461

Masud M, Gaba GS, Alqahtani S, et al. A lightweight and robust secure key establishment protocol for internet of medical things in COVID-19 patients care. *IEEE Internet Things J*. 2020;1:9. https://doi.org/10.1109/JIOT.2020.3047662

Öksüz C, Urban O, Güllü MK. Ensemble-CVDNet: a deep learning based end-to-end classification framework for COVID-19 detection using ensembles of networks. *ArXiv*. December 2020.

Spencer BF, Hoskere V, Narazaki Y, et al. Impact of digital fingerprint image quality on the fingerprint recognition accuracy. *Multimed Tools Appl*. 2019;78:3649-3688. https://doi.org/10.1007/s11042-017-5537-5

Zhong A, Li X, Wu D, et al. Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19. *Med Image Anal*. 2021;70:101993. https://doi.org/10.1016/j.media.2021.101993

Wang H, Li Z, Li Y, Gupta BB, Choi C. Visual saliency guided complex image retrieval. *Pattern Recognit Lett*. 2020;130:64-72. https://doi.org/10.1016/j.patrec.2018.08.010

Dietterich TG, Oregon. Ensemble. *Methods Mach Learn*. 2000; 1857(2):1-15.
57. Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. *Artif Intell Rev*. 2020;53:5455-5516. https://doi.org/10.1007/s10462-020-09825-6

58. Krizhevsky BA, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. 2012.

59. Kittler J, Hater M, Duin RPW. Combining classifiers. *Proc Int Conf Pattern Recognit*. 1996;2(3):897-901. https://doi.org/10.1109/ICPR.1996.547205

60. Kumthekar A, Ramachandra Reddy G. Ensemble learning technique for cloud classification. *Int J Eng Adv Technol*. 2019;9(2):2582-2587. https://doi.org/10.35940/ijeat.b3957.129219

61. Tewari S, Dwivedi UD. A comparative study of heterogeneous ensemble methods for the identification of geological lithofacies. *J Pet Explor Prod Technol*. 2020;10(5):1849-1868. https://doi.org/10.1007/s13202-020-00839-y

62. Liang Q, Zhu W, Sun W, Yu Z, Wang Y, Zhang D. In-line inspection solution for codes on complex backgrounds for the plastic container industry. *Measurement*. 2019;148:106965. https://doi.org/10.1016/j.measurement.2019.106965

63. Sarker L, Islam M, Hannan T, Zakaria A, Ahmed Z, Zakaria A. COVID-DenseNet: a deep learning architecture to detect COVID-19 from chest radiology images. *Preprints*; May 2020. https://www.preprints.org/manuscript/202005.0151/v1

64. Das D, Santosh KC, Pal U. Truncated inception net: COVID-19 outbreak screening using chest X-rays. *Phys Eng Sci Med*. 2020;43(3):915-925. https://doi.org/10.1007/s13246-020-00888-x

65. Fei-Fei L, Deng J, Li K. ImageNet: constructing a large-scale image database. *J Vis*. 2010;9(8):1037. https://doi.org/10.1167/9.8.1037

66. Gianchandani N, Jaiswal A, Singh D, Kumar V, Kaur M. Rapid COVID-19 diagnosis using ensemble deep transfer learning models from chest radiographic images. *J Ambient Intell Hum Comput*. 2020;1-13. https://doi.org/10.1007/s12652-020-02669-6

67. Nour M, Çömert Z, Polat K. A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization. *Appl Soft Comput*. 2020;97:106580. https://doi.org/10.1016/j.asoc.2020.106580

**How to cite this article:** Bhardwaj P, Kaur A. A novel and efficient deep learning approach for COVID-19 detection using X-ray imaging modality. *Int J Imaging Syst Technol*. 2021;31(4):1775-1791. https://doi.org/10.1002/ima.22627