Covid 19 image classification using hybrid averaging transfer learning model

Qamar Abbas a,*, Khalid Mahmood b, Saif ur Rehman c, Muhammad Imran d

a Department of Computer Science, Faculty of Computing and Information Technology, International Islamic University, Islamabad Pakistan
b Institute of Computing and Information Technology, Gomal University, D.I.Khan, KP Pakistan
c University Institute of Information Technology, PMAS Arid Agriculture University, Rawalpindi Pakistan
d Department of Robotics and Artificial Intelligence, SZABIST University Islamabad

* Corresponding author: Qamar Abbas, Email: qamar.abbas@iiu.edu.pk

Received: 01 August 2023, Accepted: 25 September 2023, Published: 01 October 2023

KEYWORDS
Image Classification
Covid-19
Transfer Learning
Hybrid Model
Averaging

ABSTRACT
The outbreak of Corona Virus 2019(Covid-19) is a great threat to the whole world. It is crucial to early detect patients infected with covid-19 and treat them to mitigate the rapid spread of this disease. It is an immediate priority to overcome the traditional screening and develop an accurate as well as speedy covid-19 automatic diagnosis system. Computer Tomography (CT) and Chest X-Ray imaging coupled with deep learning models to develop and test Computer Aided Screening (CAS) of covid-19 images from the normal images. In this paper classification and screening of covid-19 disease are performed by using pre-trained convolutional neural networks and a proposed hybrid model on an available standard dataset of chest X-Ray images. The proposed hybrid model employs the pre-trained Convolutional Neural Network models and Transfer Learning models. Our proposed model consists of three stages where extraction of features is performed in first stage by using pre-trained machine learning model. Deep features are extracted by using the infusion of the Transfer Learning Technique in the second stage of the model. The third stage uses Flatten and Classification layers to diagnose of Covid-19 patients. In order to assure the consistency of the proposed model, by considering standard dataset X-Ray images. Simulation results of performance metrics of Accuracy, F1 Score, Precision, Recall, ROC, and AUC curve, and training and testing loss are used to evaluate and compare the proposed model with existing models. Experimental result demonstrates that the hybrid model improves the screening process for Covid-19 disease by achieving higher accuracy.

1. Introduction
Novel Coronavirus 2019 pneumonia (COVID-19) is the seventh known Corona virus caused by severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) began its rapid propagation in human being within a shorter period of time [1]. A great challenge to human beings begin when a huge number of cases were reported in the in city of Wuhan, China during December 2019 because of unknown reasons [2].
After reporting of number of cases in Wuhan, China, the World health organization (WHO) started working to investigate the disease along with Chinese authorities and declared a pandemic on 11 March 2020 [3]. Till 5 December, 2021 total unfortunate deaths are 5,265,963, recovered cases are 239,465,982, active cases are 21,032,380 and total cases are 265,764,325 throughout the world [4]. Covid-19 badly affects lives of being as well as crushed the world economy badly. Significant results have been achieved by researchers for the classification of Covid-19 by applying machine learning algorithms. Some of the commonly used machine learning algorithms and models are AdaBoost [5], Regression Model [6], Decision Tree [7], support vector machines [8], Bayes Network [9], Random Forest [10] and convolutional neural networks [11]. Gene sequencing or reverse transcription polymerase chain reaction (RT-PCR) is considered to be one of the gold standards for the detection of Covid-19 but the major drawback with this detection method is slow screening and high false rate [12] [13]. Early screening and timely start of treatment for covid-19 patients can greatly help to fight against a number of variants and to reduce rapid transmission of this disease from human to human [14] [15].

The field of Machine learning can help to develop enriched computer-based models that can work with high dimensional data to extract patterns, learn these patterns and generalize them for the screening of COVID-19 without explicitly programming [16]. There is a dire need to develop a speedy and accurate decision support system for that is helpful in detecting COVID-19 at an early stage. It is evident from the ongoing research that various learning algorithms are significantly helping the medical practitioners to detect COVID-19 from radiology images [17]. Convolution neural networks are widely utilized by many researchers to extract features, detect, predict, and classify the covid-19 disease by using radiology images and reported promising results [18].

This research work presents a hybrid CNN model which is helpful to improve the classification performance metric. We have used a standard image dataset to measure the power of hybrid CNN and compared it with well-known pertained CNN and transfer learning models. During the analysis of simulation results of existing pre-trained models, it was observed that the performance of the pre-trained model has a great variation for different datasets and the proposed model is more consistent as compared to existing pre-trained models.

Most of the deep learning models suffers in vanishing gradient problem during the learning of features for complex problems. Bigger number of extraction of features is another challenge to deep learning models. So there is a dire need to develop a model that should be capable to handle vanishing gradient problem and should increase those parameters which results high learning accuracy. In this research we have developed a hybrid averaging transfer learning model which is capable to handle vanishing gradient problem and increase the overall performance of the model due to extraction of high learning feature.

The rest of the paper contains literature review in section 2 along with existing pre-trained convolutional neural networks. Proposed hybrid averaging model is discussed in section 3 of this paper. Dataset and evaluation metrics used in this research work are presented in section 4 and experimental setup in section 5 of this paper. The results generated during simulation are presented in section 6 and conclusion and future work are given in the last section of this paper.

2. Literature Review

Several researchers have presented various aspects of Covid-19 in their research work. In this research work, we will present the overview of research contributions of researchers by using CNN. Hasan et al. [19] have presented a study that focuses on making it feasible for the 2DEMD-based modified CT-Scan and Chest-X-ray images to be utilized as performance amplification criteria using CNN to detect SARS-CoV-2 virus in patients by utilizing publically accessible radiology images of the three databases. Fang and wang [20] have introduced a particular region of lung infection-based deep network model. The localized ROI features of images are extracted to process images of Covid-19 by utilizing convolution and de-convolution techniques. Various continuous non-linear kernels of size 3x3 are used to enhance the capacity of learning depth and achieve accuracy up to 98 %. In order to detect COVID-19 in real-time and automatically, the research work introduces an AI-based solution by Anas et al. [21]. Generally, these strategies showed exceptional results for timely detection of disease. Furthermore, they also presented a powerful detailed framework for lung segmentation, identifying, limit & evaluate the infection of COVID-19 from the chest X-Ray images. They have utilized restricted CXR datasets containing fewer numbers of COVID-19 images for the assessment of COVID-19
A novel multi-layer attention and Dense GAN are utilized for lesion segmentation for COVID-19 images in this research [22]. The improved dense GAN was introduced to produce better simulations, which enhanced the small dataset from the medical domain and achieved better accuracy in diagnostic pulmonary scoring tests such as COVID-19 chest X-rays images. The HMB-HCF framework is a hybrid COVID-19 method that includes nine phases. One of these, lung segmentation with X-rays named LSAXI was used in the third phase and designed hypothetical CNN model called ‘HMB1-COVID19’. This final step predicts person recoveries from various injuries based on data gathered across 8 different sources which are then unified into one cohesive whole using filters while also pooling together respective units over time periods during treatment via three blocks. The HMB-LSAXI method provides the success ratio of 53.30% [23]. Various convolutional neural networks are used in radiography to explore some of the ways that deploy deep CNNs to COVID-19 for the screening of coronavirus [16].

A new approach to identify COVID-19 infections using various radiography images with dimensionality reduction techniques applied to chest radiography images [24]. The proposed method used an optimized set of synthetic features to separate non-infected individual cases with 94% of accuracy; it also extracts identifying characteristics present in whole CXR scans with no segmentation happening during training or testing periods. The authors found promising results of their model by identifying and extracting features without segmentation execution from chest X-Ray images. The presented method is an automated diagnosis tool that utilizes CNN infused with machine learning & DenseNet model to examine chest X-rays with high accuracy and precision [25]. The input data from these images are fed into CNNs, which predict whether or not they contain viruses related to COVID-19. This prediction utilizes DenseNet architecture; its functions include classification by comparing multiple optimizers, loss functions, and LR scheduler in each pixel’s neighborhood while also taking into account neighboring pixels’ responses. Keras based five various models such as ResNet50 (Image Recognition Network), InceptionResNetV2 (Inversional remembering nets), Xception transfer learning, and VGG-16 network with four different methods being benchmarked against each other in terms of performance metrics such as precision; recall; F1 scores are used for Covid-19 image data [26].

A novel deep neural network (D-NN) model namely COVIDetectNet for the screening of COVID-19 is proposed in the research work [27]. The presented method has three stages: pre-learned feature extractor convolutional layers and fully connected ones from AlexNet architecture were used. In step 1, CNN & dense layers of pre-trained AlexNet models are used for the extraction of features. Secondly, the best features are selected from deep features that are obtained by utilizing the SVM method. Lung images are used in an automated system for image classification for positive/negative covid. The feature reduction method along with the q-transform model is used for the extraction of optimal features giving accurate performance by using SVM, decision tree algorithms, and K-Nearest Neighbor (kNN). When evaluated with a dataset comprising 276 patients data set to measure their accuracy rates; it achieved high levels using SVM [28]. Historical power consumption data was utilized to detect electricity theft by using a robust CNN-LSTM. The dataset first had many missing data points which were filled in by new algorithms and techniques to address class imbalance problems that occurred when generating synthetic minority over-sampling (SMOTE) datasets from historic periods where there was not enough information available about individual users. The performance of the model has been improved after including the training with augmented images and obtained higher accuracy than SVN and also applied logistic regression [29]. The Multiclass CNN model models could not spot only those images that were unclear because they both have almost identical symptoms, findings on X-rays etc., making it difficult to differentiate between Pneumonia and Covid 19 correctly with this method alone. However, proposed Network methods make use of large datasets as compared CT (or other existing) recall systems without extracting manual features from data – increasing accuracy in performance in multiclass and binary classification using available datasets to predict accurate performance. Due to multiple disease symbols that cannot be classified accurately along with COVID-19 [30].

Deep and machine learning methods are used to differentiate COVID-19 from non-COVID-19 pneumonia relying on chest images [31]. These approaches have high accuracy, but lack transparency because you can’t identify exactly what feature is being applied by the algorithm for evaluating the
output. This makes it difficult when applying these approaches to clinical data sets where there might be many features involved. The accessibility of open databases about X-ray and CT images of patients using machine learning approaches on huge numbers of clinical images. The proposed deep learning method can detect COVID-19 for Chest X-ray images with high sensitivity and specificity by fine-tuning four pre-trained CNN models that were already trained on available training sets. Two datasets were combined COVID-Xray-5k and images were used with the assistance of a radiologist to verify the COVID-19 labels. The study in the research work executes a comprehensive simulation analysis to evaluate the efficiency of each of the 4 models using ROC, AUC, sensitivity, and specificity performance metrics for COVID Xray-5k Dataset [32]. A fusion-based model based on attention-based selection is presented to perform aggregation and feature selection [33]. Availability of public COVID-19 dataset that has 20 COVID-19 CT scans and non-COVID-19 lung lesions to COVID-19 infections dataset. The automatic COVID screening (ACOS) system proposed in the research work employs a hierarchical classification technique to identify patients with different diseases. It uses conventional machine learning algorithms and radiomic texture descriptors for its separation of normal, pneumonia-infected individuals from those who are suffering nCOVID - 19 illness. The main advantage of the proposed strategy is that it only requires limited annotated images as well as being easily model-able using any number of available computers or servers allowing deployment even if you didn’t have enough resources before! This majority vote-based classifier ensemble strategy reduces misclassification chances [34]. The pre-trained weights and proposed algorithm are used to learn a pattern from extracted features of COVID-19 cases that are acquired from Covid patients. The algorithm uses five extra layers to do this well, just as other algorithms can be applied with input datasets that include CXR or CT Scan images taken at different times during treatment for Pneumonia when it occurs on chest x-rays (COVID), SARS coronavirus 2 which causes pneumonia but produces no splenic signs seen only by X ray imaging. The three radiography image datasets have been considered alongside normal patients’ scans done without any diagnosing indications in the algorithm [17].

A pre-trained transfer learning-based VGG-16 model is used with an aim to detect whether it is a healthy chest X-ray patient or pulmonary disease and when it’s marked with COVID 19 then go to identify areas which discriminate/highlight specific symptoms that may occur in patients suffering from such condition so they can be accurately diagnosed further development of treatments for these diseases [35]. A novel CNN model was introduced in the screening of COVID-19 among chest X-ray and CT images in the research work and named as CoroDet. The presented model was capable of successfully diagnosing COVID or Normal class. Both types with different clinical symptoms like pneumonia having high accuracy levels in comparison to ten existing techniques were tested on the data from patients’ records without sacrificing much time spent comparing their results against gold standard diagnosis methods that could have spared physicians’ precious minutes out there waiting around at hospitals everywhere observing what happens next. Yet another contribution made possible by using big whopping datasets during the training process as well: the largest ever assembled. The performance of the presented strategy shows the high performance of the presented model as compared to existing well-known methods [36].

3. Proposed Hybrid Averaging Model

The proposed model is shown in Fig. 1 of this paper. It starts with reading an X-Ray image from a data set, and then a convolution layer is added to extract channel-wise data from the source image. To extract features from the source image we have applied two different pre-training transfer learning models VGG-16 and DenseNet-201 model. The two different models have different feature extraction mechanisms and both models can extract varying features and their hybrid will be useful in getting improved research results. After extracting features using model-1 and model-2, the features are averaged and then forwarded to flattened layer and dense layers for more robust learning. Flatten layer is placed between the feature extraction layer of CNN and ANN classification layer. It flattens the output of CNN layers and gives it as an input for the ANN. A Softmax layer is added to get the classification results into Covid-19 positive or negative class. The proposed hybrid model is presented in Fig. 1.
The data of Covid-19 images is considered as a complex data and it may face the vanishing gradient issue. We have used hybrid averaging model consisting of Densenet-201 that is helpful to handle the vanishing gradient problem. It also strengthens the features that becomes helpful to reduce number of training parameters during the extraction of features. The advantage of VGG is that it extracts those features which contributes in high accuracy. Therefore, the vanishing gradient handling and extraction of high accuracy features makes the proposed hybrid averaging model as a better performing model.

### 3.1 Dataset and Evaluation Metrics

The detail of used datasets and evaluation metrics is given in this section.

#### 3.1.1 Dataset

In this research work, we have used a standard dataset to evaluate the performance of developed approach. This dataset is an open dataset consisting of Chest X-Ray images of 5831 images. The dataset is available at [https://github.com/ieee8023/covid-chestxray-dataset/](https://github.com/ieee8023/covid-chestxray-dataset/)

#### 3.1.2 Evaluation metrics

Various standard and commonly used machine learning evaluation metrics for classification tasks are used in the experimentation of our model which includes F1 Score, Accuracy, Recall, Precision, Specificity and area under curve (AUC) score [42]. More details about these evaluation metrics is given the formulas written in the seven equations (1)-(5) where False Positive, False Negative, True Positives, and True Negative are referred to as FP, FN, TP, TN respectively. It is important to mention here that FP shows incorrectly classified images, FN represents images that belong to a class detected as un belonging class, TP shows correctly classified images to the belonging class and TN refers to a number of those images which does not belong to a class but classified as belonging to the class [43].

#### 3.1.2.1 Accuracy

Accuracy is an important measurement in classification which shows the ratio of correct classified images and total images used to measure the experimental results.
\[
\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}
\]  \hspace{1cm} (1)

3.1.2.2 Recall/Sensitivity

It actually calculates the ratio between the total correctly predicted positive classified images and the sum of all FN and TP images.

\[
\text{Recall/Sensitivity} = \frac{TP}{TP+FN}
\]  \hspace{1cm} (2)

3.1.2.3 Precision

Precision is used to calculate the ratio between the total correctly classified images among all positive sampled classified as positive. The formula to calculate precision is as follows.

\[
\text{Precision} = \frac{TP}{TP+FP}
\]  \hspace{1cm} (3)

3.1.2.4 F1-Score

F1 Score is another important Performance metric which is the harmonic average showing the weighted average of precision and sensitivity. It is used to show the trade-off between sensitivity and precision. The formula to calculate F1 Score is as follows.

\[
\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]  \hspace{1cm} (4)

3.1.2.5 Specificity

Specificity (also known as False Positive Rate(FPR)) is the sensitivity of the negative class samples. Specificity and Sensitivity are the inverse of each other when one increases then the other decreases across various thresholds. It can be calculated by taking the ratio between TN and the sum of all TN and FP images.

\[
\text{Specificity} = \frac{TN}{TN+FP}
\]  \hspace{1cm} (5)

3.1.2.6 Area under the Curve (AUC) and Receiver Operating Characteristic (ROC)

ROC and AUC curve are used to evaluate the discriminatory ability and find optimal cut-off points in comparing the efficiency of classifiers. ROC curve is basically a plot showing the False Positive Rate (Specificity) against the True Positive Rate (Sensitivity).

\[
\text{TPR} = \frac{TP}{TP+FN}
\]  \hspace{1cm} (6)

\[
\text{FPR} = \frac{FP}{FP+TN}
\]  \hspace{1cm} (7)

4. Experimental Setup

This section contains details of experiments that are used to generate results by using the considered data set. Learning rate is basically step size and is used to control the learning adaption of the problem by the deep learning model. The small value of learning rate required more number of training epochs in the model however its bigger value results in bigger change in the learning weights of the model. We have used dataset of COVID-19 images so we have used value of learning rate as 0.03. We have divided the data into three sets where 70% is considered as training data, and validation and testing data are used as 10% and 20% respectively. The same hyper-parameters are used in the experimentation of all used models. The experimental results are generated using Google-Colab environment. The performance metric of Accuracy, Precision, Recall, F1-Score, Specificity and AUC/ROC curve is considered throughout all experiments for the proposed hybrid model, VGG-16 and Resnet-201 models. The experiment that we have considered Chest-XRay (discussed as a dataset) standard dataset to assess the performance of proposed and other used models.

4.1 Experimental Results

Several pre-trained CNN and transfer learning models are available in the literature. Overall we have used Resnet-50, DenseNet, Inception, Xception, and Alexnet for comparison purpose. The hybrid of VGG-16 and DenseNet201 is used in this research work which utilizes the learning parameters of VGG-16 and DenseNet-201 models. We have used the Performance Metric given in section 4.2 to assess the proposed hybrid averaging model. From Table 2, it is evident that the proposed hybrid averaging model has better performance in terms of all parameters as compared to all other well-known transfer learning and deep learning models for the considered dataset. One can depict from graphs of training loss and validation loss which is illustrated in Fig. 2, 3, that the performance of the proposed hybrid averaging model has better performance as compared to other considered models. Training loss and validation are gradually and quickly decreasing as compared to other models. The same is confirmed from the bar graph reported in Fig. 4 of this paper. It can be observed by the confusion matrix shown in Fig. 5 that the proposed hybrid averaging model outperforms all other considered models as reported in the matrix. The ROC and AUC curve for the proposed hybrid averaging model cover more area than other transfer learning and deep learning models as shown in Fig. 6 of this paper.

It can be observed from Fig. 2(a) that performance of proposed hybrid averaging transfer learning model is similar to DenseNet 201, DenseNet 169 and
Xception in the start of training epochs however as number of training epochs increases the error of proposed model is more gradually decreasing as compared to all other and it continues till the last epoch. It can be concluded that learning capability of proposed hybrid averaging transfer learning model is better as compared to all other models. Validation loss graph illustrates the similar performance of proposed hybrid averaging transfer learning model in the early epochs but as number of epochs increases the validation loss decreases in the later iterations.

Table 2
Performance Metric showing precision, recall, F1Score and accuracy of considered CNN’s and Proposed Hybrid Model.

<table>
<thead>
<tr>
<th>Transfer Learning Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG16</td>
<td>0.88</td>
<td>0.91</td>
<td>0.89</td>
<td>0.91</td>
</tr>
<tr>
<td>Xception</td>
<td>0.93</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>DenseNet201</td>
<td>0.93</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>Resnet50</td>
<td>0.82</td>
<td>0.81</td>
<td>0.82</td>
<td>0.84</td>
</tr>
<tr>
<td>Densenet169</td>
<td>0.93</td>
<td>0.95</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>VGG19</td>
<td>0.88</td>
<td>0.89</td>
<td>0.89</td>
<td>0.9</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Resnet152</td>
<td>0.78</td>
<td>0.79</td>
<td>0.79</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Fig. 2. Logarithmic graphs of a) Training Loss and b) Validation Loss, showing validation loss vertically and epochs horizontally for various pre-trained and proposed hybrid CNN.

Fig. 3. Logarithmic graphs of a) Training accuracy and b) Validation accuracy, showing accuracy vertically and epochs horizontally for various pre-trained and proposed hybrid CNN.
Fig. 4. Bar Graph showing Performance Metric vertically and CNN model horizontally.

Fig. 5. Confusion Matrix for (a) VGG16, (b) Xception, (c) DenseNet201, (d) Resnet50, (e) Densenet169, (f) VGG19, (g) Proposed Hybrid Model, (h) Resnet152 Respectively.
5. Conclusion and Future Work

Early detection of COVID-19 is considered an effective way to screen the COVID-19 disease from one individual to another individual. To overcome the slow conventional screening techniques of Covid-19, it is important to detect the initial stage of COVID-19 with acceptable accuracy from CXR/CT images. In this research work, open dataset of Chest X-Ray images is used. We have presented a novel hybrid model based on transfer learning and deep learning models. The proposed model utilizes features of VGG-16 and Resnet-50 and pre-trained models with a few more layers. The three phases of the proposed model consist of the extraction of features in the first phase, deep feature selection and dropouts in the second phase, and classification in the third phase of the proposed model. Research results are generated by using commonly used performance metrics for the used dataset discussed in section 4 of this paper. Simulation result generated using Google-Colab shows that the proposed hybrid model has dominating performance. The future work of this research work is to optimize the hyper-parameters of proposed model by using evolutionary computing optimization algorithms to improve the convergence power of the proposed model.

6. References


