



Automatic detection of COVID-19 from chest CT scan and chest X-Rays images using deep learning, transfer learning and stacking

Ebenezer Jangam^{1,2} · Aaron Antonio Dias Barreto³ · Chandra Sekhara Rao Annavarapu²

Accepted: 27 March 2021 / Published online: 7 June 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

One of the promising methods for early detection of Coronavirus Disease 2019 (COVID-19) among symptomatic patients is to analyze chest Computed Tomography (CT) scans or chest x-rays images of individuals using Deep Learning (DL) techniques. This paper proposes a novel stacked ensemble to detect COVID-19 either from chest CT scans or chest x-ray images of an individual. The proposed model is a stacked ensemble of heterogeneous pre-trained computer vision models. Four pre-trained DL models were considered: Visual Geometry Group (VGG 19), Residual Network (ResNet 101), Densely Connected Convolutional Networks (DenseNet 169) and Wide Residual Network (WideResNet 50 2). From each pre-trained model, the potential candidates for base classifiers were obtained by varying the number of additional fully-connected layers. After an exhaustive search, three best-performing diverse models were selected to design a weighted average-based heterogeneous stacked ensemble. Five different chest CT scans and chest x-ray images were used to train and evaluate the proposed model. The performance of the proposed model was compared with two other ensemble models, baseline pre-trained computer vision models and existing models for COVID-19 detection. The proposed model achieved uniformly good performance on five different datasets, consisting of chest CT scans and chest x-rays images. In relevance to COVID-19, as the recall is more important than precision, the trade-offs between recall and precision at different thresholds were explored. Recommended threshold values which yielded a high recall and accuracy were obtained for each dataset.

1 Introduction

COVID-19, short for Coronavirus Disease 2019, is a life-threatening disease caused by Severe Acute Respiratory Syndrome Corona Virus (SARS-CoV-2). It is a disease

spreading like wildfire throughout the world for which currently there is no cure. This virus was first detected in Wuhan, China, from where it spread to the rest of the world. Among the many countries of the world, the United States of America, Brazil, and India are the worst hit countries. As per the statistics on January 23rd, 2021, the total number of cases worldwide was 98,809,162, and the number of deaths were 2,118,030 [1].

The most common symptoms of COVID-19 include cough, fever, breathing difficulties, and loss of taste and smell. The dangerous aspect of this virus is the unknowing spread of the virus from most of the infected people to the uninfected people around them. Therefore, the quick detection of infected people especially in the early stages is of paramount importance. The current diagnosis is done by a real-time reverse transcription-polymerase chain reaction (RRT-PCR) method from a nasopharyngeal swab. However, these test results take about a day's time to arrive, during which the infected person could spread this disease. Therefore, developing quick and efficient testing methods is the need of the hour. Another drawback of RRT-PCR is its low sensitivity in detecting COVID-19 [2–4] resulting in more false negatives ultimately leading to more spread of the disease.

This article belongs to the Topical Collection: *Artificial Intelligence Applications for COVID-19, Detection, Control, Prediction, and Diagnosis*

✉ Chandra Sekhara Rao Annavarapu
acsrao@iitism.ac.in

Ebenezer Jangam
jebenezer@vrsiddhartha.ac.in

Aaron Antonio Dias Barreto
aaron.18je0003@ece.iitism.ac.in

¹ Department of Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India

² Department of Computer Science and Engineering, Indian Institute of Technology (ISM), Dhanbad, Jharkhand, India

³ Department of Electronics Engineering, Indian Institute of Technology (ISM) Dhanbad, Jharkhand, India

One of the promising methods to detect COVID-19 quickly and early in symptomatic people is by performing chest CT scans or taking chest x-ray images of patients [5]. To illustrate the difference between the COVID-19 positive CT scan and COVID-19 negative CT scan, COVID-19 negative CT scan images and COVID-19 positive CT scan images are shown in Figs. 5 and 6 respectively. CT has demonstrated high sensitivity in the detection of COVID-19 during the initial screening of the patient [6, 7]. It can be useful in rectifying false negatives obtained with rRT-PCR in symptomatic cases [8]. However, this is also a time-consuming process as an expert is needed to read these CT scans and x-ray images to determine if a person is COVID-19 positive.

A study revealed that the chest CT of an individual with COVID-19 pneumonia is more likely to have a peripheral distribution, ground glass opacities, fine reticular opacity, and vascular thickening [4, 9]. In another finding, chest CT of a COVID-19 infected individual had a peripheral distribution, a lesion range > 10 cm, the involvement of 5 lobes, and no pleural effusion [10]. However, detection of COVID-19 using CT scan is a challenging aspect because the CT characteristics vary with the disease progression [11, 12]. Initially, ground glass opaque shadows [9, 13] and crazy paving patterns [11] are visible. After few days, the density of lesions gradually increases, and the halo and reverse halo signs appear [14]. As the disease progresses further, the lung lesions resemble the white lungs [9]. During the final stage, the density of lesions decreases, and the area of lesions also narrows [15].

To classify COVID-19 images from other images, suitable features are to be extracted. Feature extraction from COVID-19 images is a complex task due to variation of characteristics on a day-to-day basis and variation from one case to another [16, 17]. Moreover, we need to distinguish it from other types of pneumonia [18]. The hand crafted features, however, with limitations varying according to tasks, were not capable of providing adequate significant features [19]. However, Deep Learning, in particular, the textitConvolutional Neural Networks (CNNs) revealed their self-potential to extract useful features in image classification tasks [20]. This feature-extraction process requires the transfer learning techniques in which pre-trained CNN models capture the generic features of large-scale datasets such as ImageNet that are later transferred to the task required. Hence, the availability of pre-trained CNN models such as VGG 19 [21], ResNet [22] and DenseNet [23] is highly supportive in this process and seem to be quite promising for the detection of COVID-19 from chest CT scans and chest x-ray images.

The contributions of the paper are as follows:

- We proposed a novel stacked ensemble to detect COVID-19 from both chest x-rays and CT scans to achieve high accuracy. Pre-trained models and additional fully connected layers were used to design the base classifiers. After an exhaustive search, heterogeneous base classifiers with high accuracy were combined to form a weighted averaging based heterogeneous stacked ensemble. The diversity is based on the set of false positives and false negatives resulted from the base classifiers.
- We used five different public datasets which are collected from different locations to evaluate the performance of the proposed model. Among the five datasets, three datasets consist of the CT scans while two datasets consist of the chest x-rays.
- We explored the trade-offs between recall and precision to select the optimum threshold which results in a high recall. In the context of the COVID-19 pandemic, it is important for the model to reduce the number of false negatives as they may result in a widespread of the virus. Hence, we experimented with varying thresholds to get optimal recall without affecting accuracy.
- We compared the performance of the proposed stacked ensemble with baseline models and existing models.

The organization of the paper is as follows. Section 2 explicates the basic terms used in this paper. Section 3 contains a review of the literature related to the theme of this paper. Section 4 details the proposed model to detect COVID-19 from chest CT scans and chest x-rays images. Section 5 gives the details of the datasets, experimental methodology, and evaluation metrics. Section 6 gives an insight into the results and analysis of the experiments conducted. Section 7 presents our model's performance compared to other existing models, and Section 8 is the conclusion of the paper.

2 Preliminaries

In this section, the basic terms and concepts used in this paper are explained.

2.1 Deep learning

Deep learning [24] is a part of machine learning that deals with algorithms based on the human brain structure called neural networks. Neural networks can be hundreds of layers deep and can contain about a million parameters. Neural networks can achieve almost human-level performance on many tasks, and their performance only increases with the amount of data fed into them. Various deep

learning models are used, specifically for image detection, called Convolutional Neural Networks (CNNs) [25]. The advantages of these CNNs over traditional neural networks are that they use parameter sharing and use relatively fewer parameters to get the same or even better performance than a traditional neural network, thus saving both space and time.

2.2 Transfer learning

Transfer learning [26] is a deep learning technique that uses a deep learning model trained to do a specific task to perform another related task. The parameters of the original model are fine tuned to the second task. One advantage of using transfer learning is that it saves much time, as training a model from scratch would take a longer time than using a pre-trained model and just fine-tuning to the given task. Another advantage is that the data required to fine-tune a model is far less than the data used to train a model from scratch. Transfer learning is used from task A to task B when A and B have the same input type, images, in this case, and the amount of data available for task B is less as compared to task A.

2.3 Stacking

Stacking is an ensemble machine learning algorithm that learns how to combine two or more models to produce the correct output. Stacking improves the overall performance of the model. The outputs of the base models become inputs to the meta-model. The meta-model uses these inputs to produce the correct output.

For heterogeneous stacking, the base models have different network architectures. If the models had the same network architecture with little variation in their hyper-parameters, their predictions would also be relatively similar, and stacking might not achieve better performance than the base models.

Given a data set D , stacking initially splits D into subsets of equal size D_1, D_2, \dots, D_N . One of the subsets D_i is kept aside for future use. The remaining subsets generate K base classifiers using K learning algorithms. D_i is the training set, and D_i is the testing set of the i^{th} fold. After generating base classifiers, the D_i set generates the meta classifier.

The meta classifier's training set consists of predictions from K base classifiers over the instances in D_i . Meta classifier data has K -attributes whose values are the predictions from K base classifiers for each instance in D_i . The process is repeated for N folds $i = 1, 2, \dots, N$. At the end of the cross-validation process, each example of the training data for meta classifier has K -attributes and a target label. Once the data is available for a meta classifier from

all the instances of D , any learning algorithm can generate the meta classifier model. For the classification of a new example, the base classifier produces a vector of predictions used by the meta classifier to predict the class [27].

2.4 VGG 19

Visual Geometry Group (VGG 19) [21] is a deep learning model designed by the Visual Geometry Group (VGG), Department of Engineering Science, the University of Oxford for image classification. It consists of 19 layers (16 convolutional layers, 3 fully connected layers, 5 MaxPool layers, and 1 SoftMax layer) and has almost 144 million parameters.

2.5 ResNet 101

Residual Network (ResNet 101) [22] is a deep learning model used for image classification. ResNet is short for the residual network, and 101 signifies the number of layers with trainable parameters in the model. ResNet 101 consists of 101 layers and almost 45 million parameters. The distinctive feature of ResNet architecture is that it has a shortcut or skip connections present in a residual and many of such residual blocks are connected in series.

2.6 DenseNet 169

Densely Connected Convolutional Networks (DenseNet 169) [23] is a deep learning model used for image classification consisting of 169 layers. The DenseNet model consists of Dense blocks and each layer in a Dense block has a connection with all the subsequent layers in that block.

2.7 Wide ResNet 50 2

Wide Residual Network (Wide ResNet 50 2) [28] is a deep learning model used for image classification. It is a modified version of the ResNet model, which has a depth of 50 and a width of 2. It has almost 69 million parameters.

3 Literature review

Researchers have proposed different automatic Deep Learning based methods [29–39] which can assist the medical practitioners to detect this virus quickly and efficiently. This section reviews the existing deep learning models to detect COVID-19 from chest x-rays and CT Scans.

Since the coronavirus outbreak, significant research on a quick and efficient detection method has become worth of

attention. Work by Xiaowei Xu et al. [40] showed a piece of evidence that chest CT scans and chest x-rays can diagnose the coronavirus disease, as the lungs of the infected people are affected by this virus. Their proposed model achieved an accuracy of 0.87 on the dataset that they used. However, due to the privacy reasons, the chest CT scans, and chest x-rays images of COVID-19 positive patients are not publicly available, and therefore building models for the detection of COVID-19 was not an easy task.

One of the first publicly available datasets was built by Xingyi Yang et al. [41]. They proposed a deep learning model with an accuracy of 0.89 and an F1 score of 0.90. Another dataset that was made publicly available was the COVIDx dataset was created by Linda Wang et al. [42]. When they published the COVID-Net paper, their proposed model was able to achieve an accuracy of 0.93. Muhammad Farooq and Abdul Hafeez [43] in their paper, COVID-ResNet has improved further and their proposed model achieved an accuracy of 0.96. He et al. [29] built a publicly available CT scan image dataset and used a Self-Trans approach, which integrated self-supervised learning with transfer learning, which learned robust and unbiased feature representations, in order to reduce the risk of over-fitting. Their model achieved an F1 score of 0.85 and an *Area Under Curve* (AUC) of 0.94.

Polsinelli et al. [30] proposed a light CNN design based on the SqueezeNet model. Their model achieved an accuracy of 0.83, a sensitivity of 0.85, a specificity of 0.81, a precision of 0.8173, and an F1 score of 0.8333. The average classification time of their model was also relatively lower as compared to other complex CNN models.

Loey et al. [31] used classic data augmentation techniques along with CGAN to increase the size of their dataset of CT scan images. They used five different CNN based models, namely AlexNet, VGGNet16, VGGNet19, GoogleNet, and ResNet50. They found that ResNet 50 was the best model to detect COVID-19 from CT scan images. It achieved an accuracy of 82.91%. Lokwani et al. [32] built a 2D segmentation model based on the U-Net architecture, whose output was the original CT scan with the region of infection identified. Their model achieved a sensitivity of 0.96428 and a specificity of 0.8839. They also developed a method to convert slice level predictions to scan level predictions, which helped them reduce the number of false positives.

Shaban et al. [33] proposed a new hybrid feature selection methodology, which selects the most informative features from those extracted from CT scan images. This methodology combines evidence from both filter and wrapper feature selection methods. They also proposed an enhanced K Nearest Neighbour Classifier, which overcomes the traditional KNN algorithm's trapping problem using

advanced heuristics in choosing the K nearest neighbors of the sample to be tested.

Azemin et al. [34] used a deep learning model based on the ResNet 101 architecture. Their model was first pre-trained on a dataset of a million images and then retrained to detect abnormalities in chest x-rays images. Their model achieved an AUC of 0.82, a sensitivity of 0.773, specificity of 0.718, and an accuracy of 0.719. Ouchicha et al. [35] in their paper proposed a model called CVDNet based on the residual network architecture. They constructed their model using two similar levels with different kernel sizes to capture the input chest x-rays images' local and global features.

Taresh et al. [36] evaluated the performance of different models on their ability to predict COVID-19 positive cases from chest x-rays images correctly. They found that the VGG 16 model had the best performance in overall scores and based-class scores. Yadav et al. [37] evaluated two pre-trained CNN models, namely, VGG16 and InceptionV3, using data augmentation techniques. The InceptionV3 model achieved the highest classification accuracy of 0.9935 for binary classifications, whereas the VGG16 model achieved the highest accuracy of 0.9884 for multiclass classification. Rahimzadeh et al. [38] proposed a novel method for increasing the classification accuracy of Convolutional Neural Networks. They used the ResNet50V2 network and a modified feature selection pyramid network. Their model achieved an accuracy of 0.9849, and their model was able to identify 234 out of 245 patients correctly.

Wang et al. [39] proposed a new joint learning framework to perform accurate COVID-19 detection by learning with heterogeneous datasets. They used a modified version of the COVID-Net model to improve accuracy and learning efficiency. On top of this model, they further conducted a separate feature normalization in latent space. Their model was able to outperform the COVID-Net model. This provides the evidence that chest CT scans and chest x-ray images can detect COVID-19. The existing models use CNNs as they were successful at many computer vision and biomedical imaging tasks.

Models proposed in [29–39] were tested on a single dataset which require more experimentation to arrive at a conclusion. Contrarily, the results obtained using our proposed model were more reliable since we have used five different datasets for experimentation. The models proposed in the above papers have been evaluated on either chest CT scans or chest x-rays images, whereas our model has been evaluated on both chest CT scans and chest x-rays images and could achieve a good performance in both types of inputs. Although, the model proposed in [33] uses a modified version of the KNN algorithm, and has a faster execution time as compared to other deep learning models when trained using a small dataset, it does not scale well

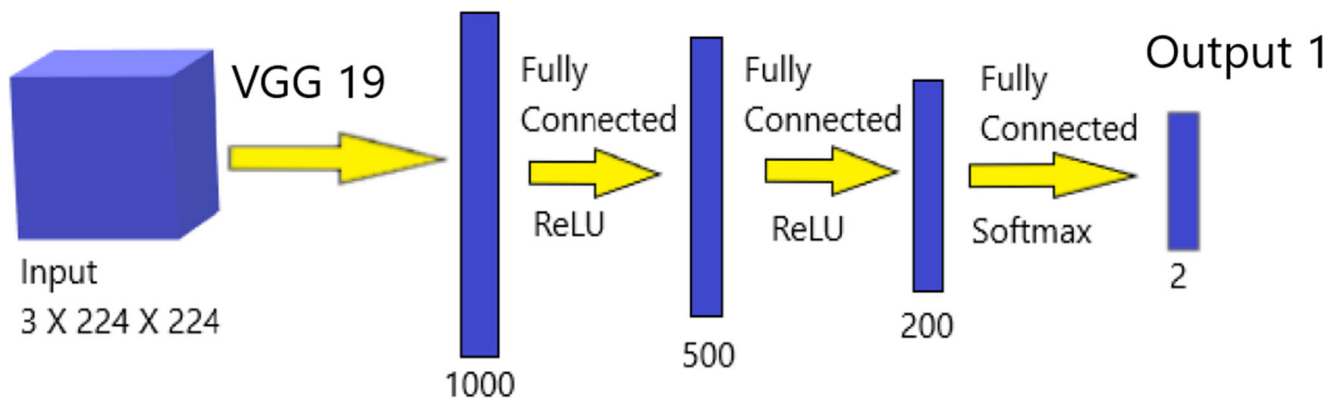


Fig. 1 Part 1 Model Architecture

to larger datasets, i.e., its execution time will be higher as compared to other deep learning models when trained on large datasets.

4 Proposed model

This section discusses the proposed model. The model consists of three parts. The first part uses a pre-trained VGG 19 model and three fully connected layers, as shown in Fig. 1. The VGG 19 model maps the input volume of size $(3 \times 224 \times 224)$ to a column vector, consisting of 1000 rows. The first fully connected layer converts this column vector into a column vector of 500 rows. The second fully connected layer further reduces this column vector into a column vector, which has 200 rows. The last fully-connected layer reduces this column vector into a column vector with as many rows as the number of classes (which is 2). The first two fully connected layers use a ReLU activation function, while the last fully connected layer uses a softmax activation function. A dropout layer, with a dropout probability of 0.5, is applied between each of the fully connected layers, to prevent the model from over-fitting to the training data.

The second part uses a pre-trained DenseNet 169 model and one fully connected layer, as shown in Fig. 2. The DenseNet 169 model maps the input volume of size $3 \times 224 \times 224$ to a column vector, consisting of 1000 rows, just like the VGG model. The fully connected layer maps this

column vector to a column vector with 2 rows (equal to the number of classes). It uses a softmax activation function. This part also uses a dropout layer with a probability of 0.5.

The third part uses a pre-trained DenseNet 169 model and three fully connected layers, as shown in Fig. 3. The third part is identical to the first part except that it uses the DenseNet 169 model instead of the VGG 19 model.

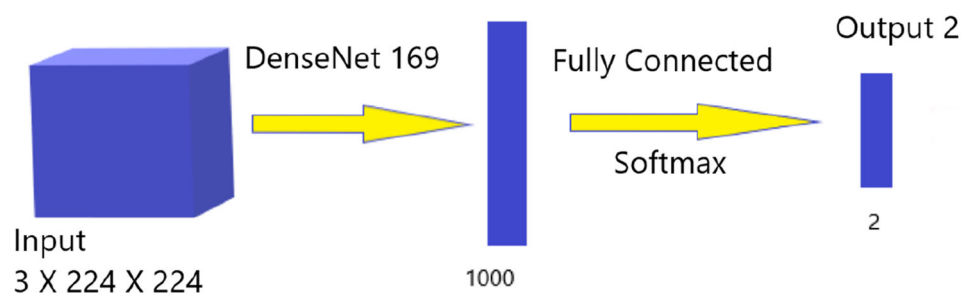
Finally, the outputs of each of the three parts are put through a single neuron to get the predicted class, as shown in Fig. 4. This single neuron uses a softmax activation function. This single neuron forms the stacking model, which assigns weights to the outputs of each of the three parts, and based on these weights and the outputs of the three parts it predicts the output class i.e. COVID-19 positive or COVID-19 negative.

The proposed model uses transfer learning so that the model can train faster. The weights of the pre-trained models are fine-tuned to the task at hand, which is detecting COVID-19. The three models are combined using stacking to predict the output class. In this model, the meta-model is a single neuron, which correctly predicts the output class based on the outputs of the three models discussed above.

5 Experimental data and methodology

This section discusses the datasets, experimental methodology and evaluation metrics.

Fig. 2 Part 2 Model Architecture



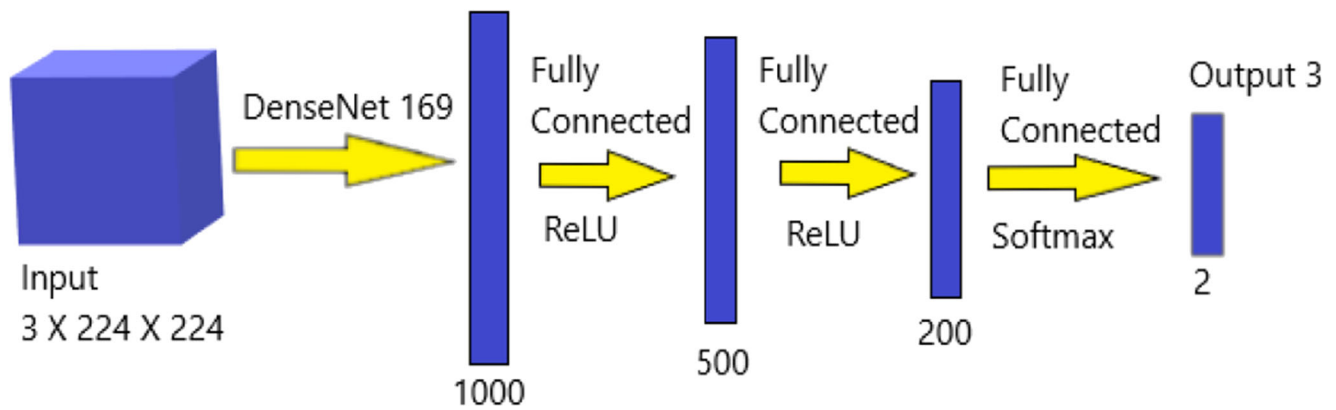


Fig. 3 Part 3 Model Architecture

5.1 Datasets

In this section, the datasets used for the evaluation of the proposed model are discussed. Five different datasets were obtained from different countries. Two of these datasets contain chest x-ray images, while the remaining datasets contain chest CT scans.

Each dataset has been split into test set, validation set and the training set. The test set was ensured to contain at least 200 images or at the most 400 images to obtain better assessment of the model's generality. The size of the validation set is based on the size of the test set, i.e., the bigger the test set, the more prominent will be the validation set and vice versa. The remaining images constituted the training set.

The test and validation sets were ensured to have the same proportion of positive and negative images. The validation set was ensured to be similar to the test set because the hyper-parameters were tuned according to the validation set. The training set's composition is immaterial as long as it had enough positive and negative images for the model to both classes' features.

1. **COVID-CT Dataset** [44]: This dataset contains 349 COVID-19 CT images from 216 patients and 397

non-COVID-19 CT images. The positive images and negative images were collected from preprints related to COVID-19. Sample COVID-19 negative images are shown in Fig. 5 and sample COVID-19 positive images are shown in Fig. 6. Test and validation sets were collected from the hospitals.

- Source: <https://github.com/UCSD-AI4H/COVID-CT>
- Type of images: CT scans
- Dataset size: 746 images
- No. of COVID-19 positive images: 349
- No. of COVID-19 negative images: 397
- Train set size: 425 images
- Validation set size: 118 images
- Test set size: 203 images

To prevent the model from over-fitting to the training data, the size of the training set is increased to 1275 images using data augmentation techniques like random rotation, horizontal flip, and color jittering.

2. **Covid-19 Image Data Collection** [45]: This dataset was collected from public sources as well as from the hospitals and physicians.

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

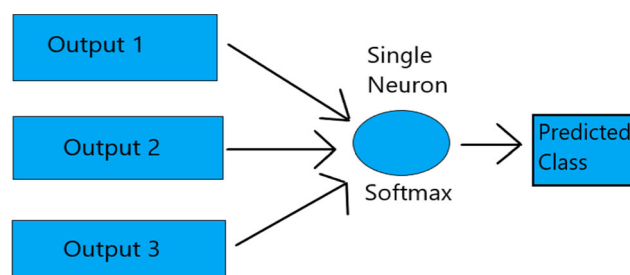
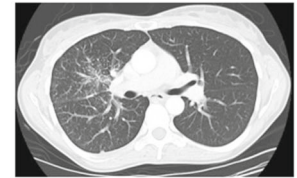


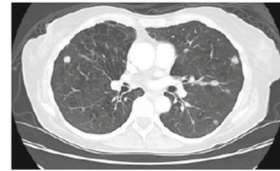
Fig. 4 Combined Model Architecture

Fig. 5 COVID-19 Negative CT scan images

(a)

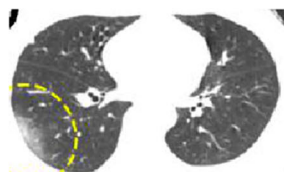


(b)

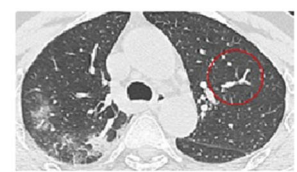


(c)

- Type of images: chest x-rays
 - Dataset size: 579 images
 - No. of COVID-19 positive images: 342
 - No. of COVID-19 negative images: 237
 - Train set size: 309 images
 - Validation set size: 70 images
 - Test set size: 200 images
3. **COVID-CTset** [46]: This dataset contains the full original CT scans of 377 persons. There are 15589 and 48260 CT scan images belonging to 95 COVID-19 and 282 normal persons, respectively. This dataset is from the Negin medical center, Sari, Iran.
- Source: <https://github.com/mr7495/COVID-CTset>
 - Type of images: CT scans
 - Dataset size: 12058 images
 - No. of COVID-19 positive images: 2282
 - No. of COVID-19 negative images: 9776
 - Train set size: 11400 images
 - Validation set size: 258 images
- Test set size: 400 images
4. **COVID-19 Radiography Database** [47]: This dataset consists of 1200 COVID-19 positive images, 1341 normal images and 1345 viral pneumonia images.
- Source: <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>
 - Type of images: chest x-rays
 - Dataset size: 2541 images
 - No. of COVID-19 positive images: 1200
 - No. of COVID-19 negative images: 2686
 - Train set size: 3086 images
 - Validation set size: 400 images
 - Test set size: 400 images
5. **SARS-CoV-2 CT scan dataset** [48]: This dataset contains 1252 CT scans that are positive for SARS-CoV-2 infection (COVID-19) and 1230 CT scans for patients non-infected by SARS-CoV-2, 2482 CT scans

Fig. 6 COVID-19 Positive CT scan images

(a)



(b)



(c)

in total. The data is from real patients in hospitals from Sao Paulo, Brazil.

- Source: <https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset>
- Type of images: CT scans
- Dataset size: 2482 images
- No. of COVID-19 positive images: 1252
- No. of COVID-19 negative images: 1230
- Train set size: 1800 images
- Validation set size: 282 images
- Test set size: 400 images

5.2 Data augmentation techniques

Large datasets are needed to train a model based on Deep Learning. When the available datasets are smaller in size, their size can be increased using the data augmentation techniques. In the present study, as COVID-CT Dataset [44] is a comparatively smaller dataset, the following data augmentation techniques were used to increase the size of this dataset.

- **Random Resized Crop** refers to cropping the given image to a random size and aspect ratio.
- **Random Rotation** refers to the rotation of the given image randomly by an angle in the given range.
- **Random Horizontal Flip** refers to flipping the given image horizontally randomly with a given probability.
- **Colour Jittering** refers to changing the brightness, contrast, and saturation of the given image randomly.

5.3 Training details

The following are the training details and parameters which were maintained constant throughout the experiment.

- Deep Learning Framework = PyTorch
- Number of epochs = 100
- Optimizer = Adam
- Learning rate = 1e-3
- Loss function = Cross Entropy Loss
- Batch size = 16
- Random Resized Crop size = 224
- Random Resized Crop Scale = (0.5, 1.0)
- Random Rotation angle range = [-5 degrees, 5 degrees]
- Random Horizontal Flip probability = 0.5

5.4 Hyper-parameter tuning

The values of the hyper-parameters namely Random Resized Crop size, Random Resized Crop Scale Random Rotation angle range and Random Horizontal Flip probability have been obtained based on the performance of the validation set.

Grid search method has been employed for hyper-parameter tuning. The values considered for each of the hyper-parameters are as follows: for Random Resized Crop size: 128, 200, and 224; for Random Resized Crop Scale: (0.5, 1.0), (1.0, 0.5) and (0.5, 0.5); for Random Rotation angle: [-3 degrees, 3 degrees], [-5 degrees, 5 degrees], and [-10 degrees, 10 degrees] ranges; and for Random Horizontal Flip probability: 0.3, 0.5, and 0.7.

Table 1 Comparison among the proposed model and other baseline models on COVID-CT Dataset [44]

Model	Precision	Recall	Accuracy	F1 Score
<i>VGG19</i> ₀	0.7321	0.8367	0.7734	0.7810
<i>VGG19</i> ₁	0.7434	0.8571	0.7882	0.7962
<i>VGG19</i> ₂	0.7345	0.8469	0.7783	0.7867
<i>ResNet101</i> ₀	0.6404	0.7449	0.6748	0.6887
<i>ResNet101</i> ₁	0.7629	0.7551	0.7685	0.7590
<i>ResNet101</i> ₂	0.6170	0.8877	0.6798	0.7280
<i>DenseNet169</i> ₀	0.7234	0.9082	0.7882	0.8054
<i>DenseNet169</i> ₁	0.7841	0.7041	0.7635	0.7419
<i>DenseNet169</i> ₂	0.7611	0.8776	0.8079	0.8152
<i>WideResNet50</i> ₂₀	0.7353	0.7653	0.7537	0.75
<i>WideResNet50</i> ₂₁	0.7064	0.7857	0.7389	0.7440
<i>WideResNet50</i> ₂₂	0.7586	0.6735	0.7389	0.7135
<i>Combination1</i>	0.7699	0.8878	0.8177	0.8246
<i>Combination2</i>	0.7679	0.8776	0.8128	0.8190
<i>Proposed Model</i>	0.7913	0.9286	0.8473	0.8545

Bold values denote the highest values obtained for a given parameter (Column)

Table 2 Comparison among the proposed model and other baseline models on Covid-19 Image Data Collection [45]

Model	Precision	Recall	Accuracy	F1 Score
VGG19 ₀	0.92	0.92	0.92	0.92
VGG19 ₁	0.8942	0.93	0.91	0.9118
VGG19 ₂	0.9048	0.95	0.925	0.9268
ResNet101 ₀	0.9878	0.81	0.9	0.8901
ResNet101 ₁	0.8824	0.9	0.89	0.8911
ResNet101 ₂	0.7983	0.95	0.855	0.8676
DenseNet169 ₀	0.8969	0.87	0.885	0.8832
DenseNet169 ₁	0.8273	0.91	0.86	0.8667
DenseNet169 ₂	0.9109	0.92	0.915	0.9154
WideResNet50 2 ₀	0.89	0.89	0.89	0.89
WideResNet50 2 ₁	0.9024	0.74	0.83	0.8132
WideResNet50 2 ₂	0.9881	0.83	0.91	0.9022
Combination1	0.902	0.92	0.91	0.9109
Combination2	0.8468	0.94	0.885	0.891
Proposed Model	0.93	0.93	0.93	0.93

Bold values denote the highest values obtained for a given parameter (Column)

5.5 Evaluation metrics

Four evaluation metrics are used to measure the performance of the proposed model. They are as follows:

- **Precision** : Precision is the fraction of positive predictions that belong to the positive class.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **Recall** : Recall is the fraction of positive examples in the dataset that are predicted positive.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- **F1 Score** : F1 Score is the harmonic mean of precision and recall.

$$\text{F1 Score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

- **Accuracy** : Accuracy is the fraction of the total predictions that are correct.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{False Positives} + \text{False Negatives} + \text{True Negatives}}$$

Table 3 Comparison among the proposed model and other baseline models on COVID-CTset [46]

Model	Precision	Recall	Accuracy	F1 score
VGG19 ₀	0.9949	0.98	0.9875	0.9874
VGG19 ₁	0.9896	0.955	0.9725	0.9720
VGG19 ₂	0.9897	0.96	0.975	0.9746
ResNet101 ₀	0.9948	0.95	0.9725	0.9719
ResNet101 ₁	1	0.935	0.9675	0.9664
ResNet101 ₂	0.975	0.975	0.975	0.975
DenseNet169 ₀	0.9946	0.915	0.955	0.9531
DenseNet169 ₁	0.9896	0.95	0.97	0.9694
DenseNet169 ₂	0.9897	0.965	0.9775	0.9772
WideResNet50 2 ₀	0.9843	0.94	0.9625	0.9616
WideResNet50 2 ₁	0.9947	0.93	0.9625	0.9612
WideResNet50 2 ₂	0.9948	0.955	0.975	0.9745
Combination1	0.9898	0.975	0.9825	0.9824
Combination2	1	0.975	0.9875	0.9873
Proposed Model	1	0.98	0.99	0.9899

Bold values denote the highest values obtained for a given parameter (Column)

Table 4 Comparison among the proposed model and other baseline models on COVID-19 Radiography Database [47]

Model	Precision	Recall	Accuracy	F1 score
<i>VGG19</i> ₀	1	0.95	0.975	0.9744
<i>VGG19</i> ₁	1	0.955	0.9775	0.9770
<i>VGG19</i> ₂	0.9949	0.98	0.9875	0.9874
<i>ResNet101</i> ₀	0.9949	0.975	0.985	0.9848
<i>ResNet101</i> ₁	1	0.97	0.985	0.9848
<i>ResNet101</i> ₂	1	0.91	0.955	0.9529
<i>DenseNet169</i> ₀	0.9948	0.96	0.9775	0.9771
<i>DenseNet169</i> ₁	0.985	0.985	0.985	0.985
<i>DenseNet169</i> ₂	1	0.96	0.98	0.9796
<i>WideResNet50</i> 2 ₀	0.99	0.985	0.9875	0.9875
<i>WideResNet50</i> 2 ₁	0.9851	0.99	0.9875	0.9875
<i>WideResNet50</i> 2 ₂	1	0.92	0.99	0.9899
<i>Combination1</i>	1	0.98	0.99	0.9899
<i>Combination2</i>	1	0.99	0.995	0.9950
<i>Proposed Model</i>	0.995	1	0.9975	0.9975

Bold values denote the highest values obtained for a given parameter (Column)

6 Experimental results

The proposed model was evaluated on the five datasets and the results are presented in this section.

6.1 Performance analysis of the proposed model

The proposed model is evaluated based on five different datasets of chest CT scans and chest x-rays i.e. COVID-CT-Dataset [44], Covid-19 Image Data Collection [45], COVID-CTset [46], COVID-19 Radiography Database [47] and SARS-CoV-2 CT scan dataset [48].

Two ensembles were designed: i) Ensemble 1, also named as Combination 1, was designed using *DenseNet169*₀, *DenseNet169*₁, *DenseNet169*₂, ii) Ensemble 2, also named as Combination 2, was designed using *DenseNet169*₀, *DenseNet169*₂, *WideResNet50* 2₂. The performance of the proposed model is compared with both the ensembles and also with the base line classifiers obtained using the models namely, *VGG19*, *ResNet101*, *DenseNet169*, *WideResNet50* 2.

The following notations were adopted in this paper: i) *Model*₀ represents the model with a softmax layer ii) *Model*₁ represents the model with a fully connected

Table 5 Comparison among the proposed model and other baseline models on SARS-CoV-2 CT scan dataset [48]

Model	Precision	Recall	Accuracy	F1 score
<i>VGG19</i> ₀	0.8389	0.885	0.8575	0.8613
<i>VGG19</i> ₁	0.8691	0.83	0.8525	0.8491
<i>VGG19</i> ₂	0.8960	0.775	0.8425	0.8311
<i>ResNet101</i> ₀	0.9341	0.78	0.8625	0.8501
<i>ResNet101</i> ₁	0.8848	0.73	0.8175	0.8
<i>ResNet101</i> ₂	0.8942	0.845	0.8725	0.8689
<i>DenseNet169</i> ₀	0.9270	0.825	0.88	0.8730
<i>DenseNet169</i> ₁	0.9045	0.71	0.8175	0.7955
<i>DenseNet169</i> ₂	0.9096	0.855	0.885	0.8814
<i>WideResNet50</i> 2 ₀	0.9223	0.89	0.9075	0.9059
<i>WideResNet50</i> 2 ₁	0.9179	0.895	0.9075	0.9063
<i>WideResNet50</i> 2 ₂	0.8653	0.835	0.8525	0.8499
<i>Combination1</i>	0.9302	0.8	0.87	0.8602
<i>Combination2</i>	0.9171	0.83	0.8775	0.8714
<i>Proposed Model</i>	0.9137	0.9	0.9075	0.9068

Bold values denote the highest values obtained for a given parameter (Column)

Table 6 Performance of the proposed model on COVID-CT Dataset [44] under varied thresholds

Threshold	Precision	Recall	Accuracy	F1 score
0.1	0.7164	0.9796	0.8030	0.8276
0.2	0.7308	0.9694	0.8128	0.8333
0.3	0.7561	0.9490	0.8276	0.8417
0.4	0.7686	0.9490	0.8374	0.8493
0.5	0.7913	0.9286	0.8473	0.8545
0.6	0.8074	0.8980	0.8473	0.8502
0.7	0.8367	0.8367	0.8424	0.8367
0.8	0.8571	0.7959	0.8374	0.8254
0.9	0.9125	0.7449	0.8424	0.8202

Bold values denote the highest values obtained for a given parameter (Column)

layer and a softmax layer, and iii) $Model_2$ represents the model with two fully connected layers and a softmax layer. The model can be any of the following deep learning models namely: *VGG19*, *ResNet101*, *DenseNet169*, *WideResNet50 2*.

The first dataset considered for the experiment is the COVID-CT Dataset [44]. It is comprised of 349 COVID-19 positive images and 397 COVID-19 negative images. This dataset was split into a training set consisting of 425 images, a validation set consisting of 118 images, and a test set consisting of 203 images. The test set contained 98 COVID-19 positive images and 105 COVID-19 negative images. Of the 98 COVID-19 positive images, the proposed model correctly classified 91 images, i.e., the number of true positives is 91, and misclassified 7 images, i.e., the number of false negatives is 7. Of the 105 COVID-19 negative images, the proposed model correctly identified 81 images, i.e., the number of true negatives is 81 and misclassified 24 images, i.e., the number of false positives is 24. Therefore the proposed model achieved an accuracy of 0.8473 and an F1 score of 0.8545. Table 1 summarizes the experiment results for the proposed model, ensembles, and various deep learning models.

The second dataset considered for the experiment is the Covid-19 Image Data Collection [45]. This dataset was split into a training set of 309 images, a validation set of 70 images, and a test set of 200 images. Of the 200 images in the test set, 100 are COVID-19 positive, and 100 are COVID-19 negative. Of the 100 COVID-19 positive images, the proposed model correctly identified 93 images, i.e., the number of true positives is 93, and incorrectly classified 7 images, i.e., the number of false negatives is 7. Of the 100 COVID-19 negative images, the proposed model correctly identified 93 images, i.e., the number of true negatives is 93, and incorrectly classified 7 images, i.e., the number of false positives is 7. Therefore, our model achieved an accuracy of 0.93 and an F1 score of 0.93. The results obtained by proposed model, ensembles for Covid-19 Image Data Collection [45] dataset are summarized in Table 2.

The third dataset under consideration is the COVID-CTset [46]. It is a large dataset that consisted of 63849 CT scan images. For this study, a smaller version of this dataset comprising of 12,058 images was used [46]. This dataset was split into a training set consisting of 11400 images, a validation set consisting of 258 images, and a test

Table 7 Performance of the proposed model on Covid-19 Image Data Collection [45] under varied thresholds

Threshold	Precision	Recall	Accuracy	F1 score
0.1	0.8151	0.97	0.875	0.8858
0.2	0.8661	0.97	0.91	0.9151
0.3	0.8739	0.97	0.915	0.9194
0.4	0.9118	0.93	0.92	0.9208
0.5	0.93	0.93	0.93	0.93
0.6	0.9394	0.93	0.935	0.9347
0.7	0.9333	0.84	0.89	0.8842
0.8	0.9432	0.83	0.89	0.8830
0.9	0.9733	0.855	0.765	0.8343

Bold values denote the highest values obtained for a given parameter (Column)

Table 8 Performance of the proposed model on COVID-CTset [46] under varied thresholds

Threshold	Precision	Recall	Accuracy	F1 score
0.1	0.985	0.985	0.985	0.985
0.2	0.9949	0.985	0.99	0.99
0.3	1	0.985	0.9925	0.9924
0.4	1	0.98	0.99	0.9899
0.5	1	0.98	0.99	0.9899
0.6	1	0.97	0.985	0.9848
0.7	1	0.965	0.9825	0.9822
0.8	1	0.965	0.9825	0.9822
0.9	1	0.945	0.9725	0.9717

Bold values denote the highest values obtained for a given parameter (Column)

set consisting of 400 images. In the test set, the number of COVID-19 positive and COVID-19 negative images are equal. Out Of the 200 COVID-19 positive images, the proposed model correctly classified 196 images, i.e., the number of true positives is 196, and incorrectly classified 4 images, i.e., the number of false negatives is 4. Of the 200 COVID-19 negative images, our model correctly classified all the images, i.e., the number of true negatives is 200, and there are no false positives. Hence, the proposed model achieved an accuracy of 0.99 and an F1 score of 0.9899. Table 3 lists the results obtained using COVID-CTset [46] for different models.

The fourth dataset is the COVID-19 Radiography Database [47]. It is comprised of 1200 COVID-19 positive images and 2686 COVID-19 negative images (Normal and Viral Pneumonia). This dataset is split into a training set consisting of 3086 images, a validation set consisting of 400 images, and a test set consisting of 200 COVID-19 positive images and 200 COVID-19 negative images. Out of the 200 COVID-19 positive images, our model correctly classified all 200 images, i.e., the number of true positives is 200, the number of false negatives is 0. Of the 200 COVID-19 negative images, our model correctly classified all 199

images, i.e., the number of true negatives is 199, and the number of false positives is 1. Hence, the proposed model achieved an accuracy of 0.9975 and an F1 score of 0.9975. The results obtained for different models using Covid-19 Image Data Collection [45] dataset are summarized in Table 4.

The fifth dataset used for evaluating the proposed model is the SARS-CoV-2 CT scan dataset [48]. It consists of 2482 CT scans, out of which a training set consists of 1800 images, a validation set consists of 282 images, and a test set consists of 400 images. Of the 400 images in the test set, 200 are COVID-19 positive, and 200 are COVID-19 negative. Of the 200 COVID-19 positive images, the proposed model correctly classified 180 images, i.e., the number of true positives is 180, and misclassified 20 images, i.e., the number of false negatives is 20. Of the 200 COVID-19 negative images, the proposed model correctly classified 183 images, i.e., the number of true negatives is 183, and misclassified 17 images, i.e., the number of false positives is 17. Therefore, the proposed model achieved an accuracy of 0.9075 and an F1 score of 0.9068. Table 5 shows the comparison of the performance of the proposed model against other models in terms of precision, recall, accuracy, and F1 score on the SARS-CoV-2 CT scan dataset [48].

Table 9 Performance of the proposed model on COVID-19 Radiography Database [47] under varied thresholds

Threshold	Precision	Recall	Accuracy	F1 score
0.1	0.995	1	0.9975	0.9975
0.2	0.995	1	0.9975	0.9975
0.3	0.995	1	0.9975	0.9975
0.4	0.995	1	0.9975	0.9975
0.5	0.995	1	0.9975	0.9975
0.6	0.995	1	0.9975	0.9975
0.7	0.995	0.995	0.995	0.995
0.8	0.995	0.995	0.995	0.995
0.9	1	0.995	0.9975	0.9975

Bold values denote the highest values obtained for a given parameter (Column)

Table 10 Performance of the proposed model on SARS-CoV-2 CT scan dataset [48] under varied thresholds

Threshold	Precision	Recall	Accuracy	F1 score
0.1	0.8527	0.955	0.895	0.9009
0.2	0.8942	0.93	0.91	0.9118
0.3	0.915	0.915	0.915	0.915
0.4	0.9146	0.91	0.9125	0.9123
0.5	0.9137	0.9	0.9075	0.9068
0.6	0.9293	0.855	0.895	0.8906
0.7	0.9486	0.83	0.8925	0.8853
0.8	0.9586	0.81	0.8875	0.8780
0.9	0.9682	0.76	0.8675	0.8515

Bold values denote the highest values obtained for a given parameter (Column)

From Tables 1 to 5, it is evident that the proposed model performs better than the other models in terms of accuracy and F1 score on all the datasets. Since the proposed model uses stacking to combine three different models, other models compensate for the misclassification made by one model, hence increasing the accuracy and F1 score of the proposed model. Therefore, the proposed model can perform better than the individual models.

6.2 Evaluation of the proposed model under varied thresholds

Once the proposed model's performance was evaluated at a constant threshold, the threshold was varied and the model was evaluated at different thresholds. The threshold is increased from an initial value of 0.1 by a step size of 0.1 until 0.9. For each threshold value, the values obtained for the evaluation metrics were recorded. The Tables 6-10 show the performance of the proposed model on different datasets by varying the threshold above which an image is predicted positive.

For the COVID-CT Dataset [44], with the increase in threshold, the precision increased, and recall decreased; this is because as the threshold increased, the number of false negatives increased and the number of false positives decreased. The F1 score is maximum when the threshold is 0.5, and accuracy is maximum when the threshold is 0.5 and 0.6. Table 6 shows the values obtained from the experiment for different thresholds. With threshold on the x-axis and a scale of 0 to 1 on the y-axis, the evaluation metrics (Precision, Recall, Accuracy, and F1 score) for COVID-CT Dataset [44] are shown in Fig. 7.

For the Covid-19 Image Data Collection [45] dataset, the precision increased with an increase in threshold. The recall decreased with an increase in threshold. The accuracy and F1 score are maximum at a threshold of 0.6.

Table 7 shows evaluation metrics obtained by varying threshold values for the proposed model on Covid-19 Image Data Collection [45] dataset. With threshold on the x-axis and a scale of 0 to 1 on the y-axis, the evaluation metrics (Precision, Recall, Accuracy, and F1 score) for Covid-19 Image Data Collection [45] are shown in Fig. 8.

For the COVID-CTset [46] also, the precision increased and the recall decreased with the increase in the threshold. The accuracy and F1 score are maximum at a threshold of 0.3. Table 8 lists Precision, Recall, Accuracy, and F1 score obtained for the proposed model at different thresholds. With threshold on x-axis and a scale of 0 to 1 on y-axis, evaluation metrics (Precision, Recall, Accuracy and F1 score) for COVID-CTset [46] are shown in the Fig. 9.

For the COVID-19 Radiography Database [47], the variation of the threshold has almost no effect on the values of precision, recall, accuracy, and F1 score. These values varied slightly as the threshold increased from 0.1 to 0.9. This slight variance is because the model is very confident

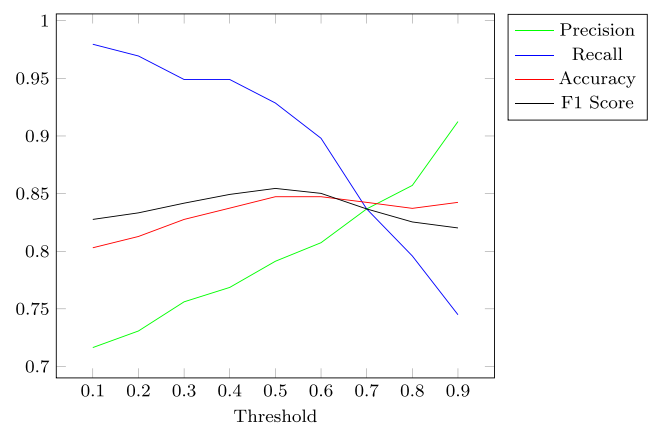


Fig. 7 Variation of Precision, Recall, Accuracy and F1 score with threshold on COVID-CT Dataset [44]

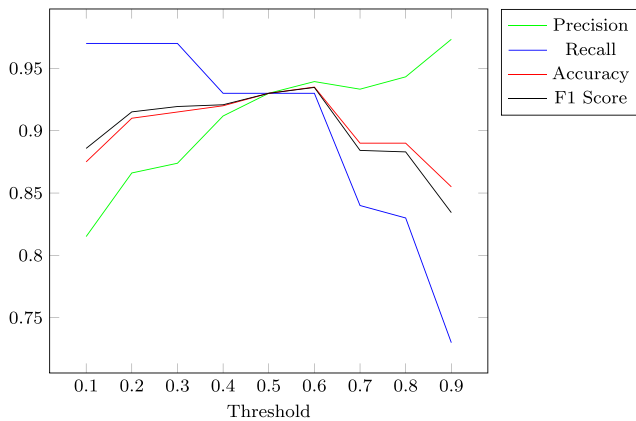


Fig. 8 Variation of Precision, Recall, Accuracy and F1 score with threshold on Covid-19 Image Data Collection [45]

of all its predictions, i.e., for most of the positive examples, the model assigned a probability greater than 0.9, and for most of the negative examples, it assigned a probability less than 0.1. Here, the probability refers to the probability that a particular example is positive. Table 9 lists Precision, Recall, Accuracy, and F1 score obtained for the proposed model at different thresholds. With threshold on the x-axis and a scale of 0 to 1 on the y-axis, evaluation metrics (Precision, Recall, Accuracy, and F1 score) for COVID-19 Radiography Database [47] are shown in Fig. 10.

For the SARS-CoV-2 CT scan dataset [48] the precision increased with an increase in the threshold. On the other hand, the recall decreased with an increase in the threshold. The relationship among precision, recall, and the threshold is similar to the first three datasets. The accuracy and F1 score are maximum at a threshold of 0.3. Table 10 lists Precision, Recall, Accuracy, and F1 score obtained for the proposed model at different thresholds. With threshold on the x-axis and a scale of 0 to 1 on the y-axis, evaluation metrics (Precision, Recall, Accuracy, and F1 score) for SARS-CoV-2 CT scan dataset [48] are shown in Fig. 11.

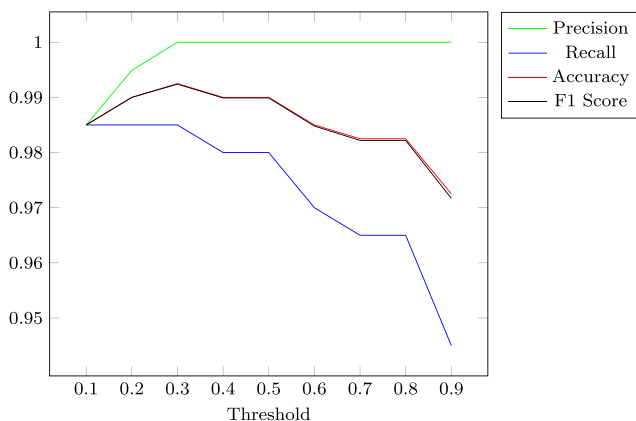


Fig. 9 Variation of Precision, Recall, Accuracy and F1 score with threshold on COVID-CTset [46]

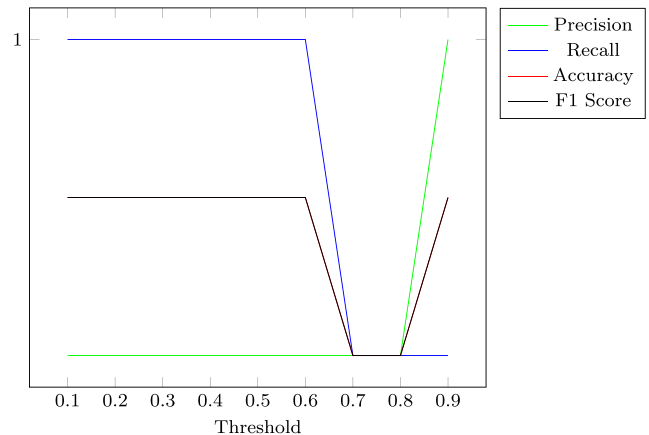


Fig. 10 Variation of Precision, Recall, Accuracy and F1 score with threshold on COVID-19 Radiography Database [47]

7 Comparison with existing models

The proposed model is able to outperform the existing models in [29–39], because the proposed model unlike the existing models does not use single model, but instead uses a stacked model, where three CNN based models are used as base models and a single neuron is used as the meta model. Moreover, the proposed model was trained on five different datasets and has seen more examples than the other models. Thus, the proposed model achieved better performance, as it could learn more relevant features from the datasets.

The proposed model comprised of three different models. These three models consisted of a pre-trained model and additional fully connected layers, and these additional fully connected layers helped the model to learn the features specific to a particular dataset. The presence of the additional fully connected layers is another reason for the proposed model's improved performance. Table 11 shows that the proposed model does better than the models proposed in previous research papers (Figs. 12–14).

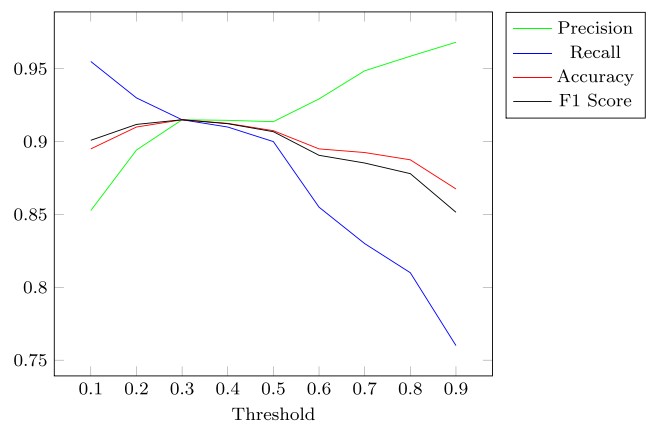


Fig. 11 Variation of Precision, Recall, Accuracy and F1 score with threshold on SARS-CoV-2 CT scan dataset [48]

Table 11 Comparison between the proposed model and models proposed in previous research papers.

Model	Dataset 1 (F1 Score)	Dataset 2 (Accuracy)	Dataset 3 (Accuracy)	Dataset 4 (Accuracy)	Dataset 5 (Accuracy)
Model [31]	0.8114	N/A	N/A	N/A	N/A
Model [29]	0.85	N/A	N/A	N/A	N/A
Model [30]	0.8333	N/A	N/A	N/A	N/A
Model [33]	0.7150	N/A	N/A	N/A	N/A
Model [34]	N/A	0.719	N/A	N/A	N/A
Model [32]	0.794	N/A	N/A	N/A	N/A
Model [38]	N/A	N/A	0.9849	N/A	N/A
Model [35]	N/A	N/A	N/A	0.9720	N/A
Model [36]	N/A	N/A	N/A	0.9969	N/A
Model [37]	N/A	N/A	N/A	0.9935	N/A
Model [39]	0.7883	N/A	N/A	N/A	0.9083
Proposed Model	0.8545	0.93	0.99	0.9975	0.915

Bold values denote the highest values obtained for a given parameter (Column)

The threshold above which the positive prediction were made, varied with different datasets. The threshold also depends on what metric one wants to give more preference. If better precision is preferred, i.e., fewer false positives, the threshold should be higher, whereas a better recall is preferred, i.e., fewer false negatives, the threshold should be lower.

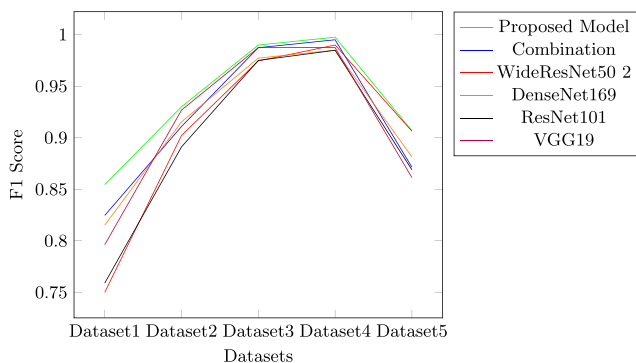
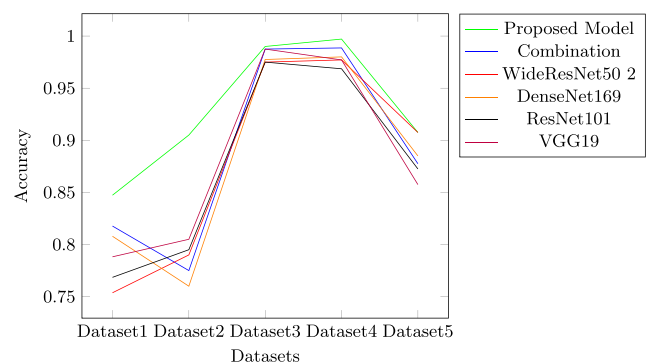
In the detection of COVID-19, it is vital not to make false negative predictions, as it can have significant consequences like the increased chance of an infected person spreading the disease to other people. Hence, recall should be given more preference along with accuracy. One method to increase the recall is to select a lower threshold. It is evident from Tables 6 to 10 and Figs. 7 to 11 that recall increased with the decrease in the threshold. Moreover, the recommended threshold varied from one dataset to another, and it lies between 0.3 and 0.5.

Another method to decrease the number of false positives is to modify the loss function to give a higher weight

to the positive examples, but again this decrease in the number of false negatives might increase the number of false positives.

8 Conclusion

In this paper, we proposed a stacked ensemble model to detect COVID 19, which has severely affected most parts of the world. Our model is a stacked ensemble of the VGG 19 and DenseNet 169 models. The proposed stacked ensemble model performed better than the base line models, other ensembles and existing models. Moreover, the proposed model achieved high accuracy and recall on all the five different datasets that consist of chest CT scans and chest x-rays Images. The recall of the model is high when the threshold is 0.5, and it increased further by decreasing the threshold. The lower the threshold, the higher is the recall, and the lower is the precision. However, there is still

**Fig. 12** F1 Score of Different models on different datasets**Fig. 13** Accuracy of Different models on different datasets

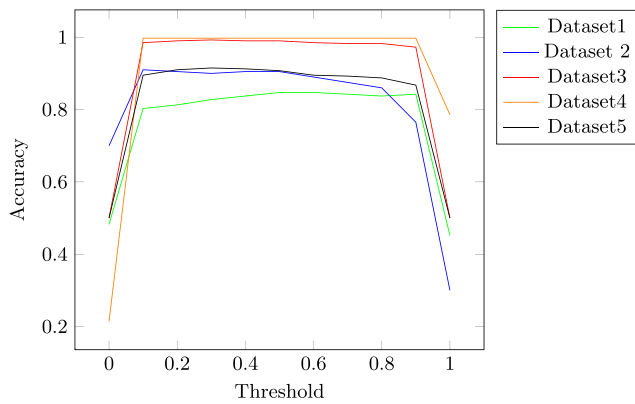


Fig. 14 Variation of Accuracy with threshold

room for improving the proposed model's performance by designing better and efficient pre-processing techniques.

References

1. Coronavirus Update (Live): 28,988,031 Cases and 925,320 Deaths from COVID-19 Virus Pandemic - Worldometer
2. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L (2020) Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. *Radiology*:200642
3. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, Ji W (2020) Sensitivity of chest ct for covid-19: comparison to rt-pcr. *Radiology*:200432
4. Kanne JP, Little BP, Chung JH, Elicker BM, Ketani LH (2020) Essentials for radiologists on covid-19: an updated radiology scientific expert panel. *Radiological Society of North America*
5. Kovács A, Palásti P, Veréb D, Bozsik B, Palkó A, Kincses ZT (2020) The sensitivity and specificity of chest ct in the diagnosis of covid-19. *Eur Radiol*:1–6
6. Bai HX, Hsieh B, Xiong Z, Halsey K, Choi JW, Tran TML, Pan I, Shi L-B, Wang D-C, Mei J et al (2020) Performance of radiologists in differentiating covid-19 from viral pneumonia on chest ct. *Radiology*:200823
7. Long C, Xu H, Shen Q, Zhang X, Fan B, Wang C, Zeng B, Li Z, Li X, Li H (2020) Diagnosis of the coronavirus disease (covid-19): rrt-pcr or ct? *Eur J Radiol*:108961
8. 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*:200343
9. Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, Cui J, Xu W, Yang Y, Fayad ZA et al (2020) Ct imaging features of 2019 novel coronavirus (2019-ncov). *Radiology* 295(1):202–207
10. Li X, Fang X, Bian Y, Lu J (2020) Comparison of chest ct findings between covid-19 pneumonia and other types of viral pneumonia: a two-center retrospective study. *Eur Radiol*:1–9
11. Ye Z, Zhang Y, Wang Y, Huang Z, Song B (2020) Chest ct manifestations of new coronavirus disease 2019 (covid-19): a pictorial review. *Eur Radiol*:1–9
12. Wang K, Kang S, Tian R, Zhang X, Wang Y (2020) Imaging manifestations and diagnostic value of chest ct of coronavirus disease 2019 (covid-19) in the xiaogan area. *Clinical Radiology*
13. Fan L, Li D, Xue H, Zhang L, Liu Z, Zhang B, Zhang L, Yang W, Xie B, Duan X et al (2020) Progress and prospect on imaging diagnosis of covid-19. *Chin J Acad Radiol*:1–10
14. Li Y, Xia L (2020) Coronavirus disease 2019 (covid-19): role of chest ct in diagnosis and management. *Am J Roentgenol* 214(6):1280–1286
15. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, Zheng D, Wang J, Hesketh RL, Yang L et al (2020) Time course of lung changes on chest ct during recovery from 2019 novel coronavirus (covid-19) pneumonia. *Radiology*
16. Liu H, Liu F, Li J, Zhang T, Wang D, Lan W (2020) Clinical and ct imaging features of the covid-19 pneumonia: Focus on pregnant women and children. *Journal of Infection*
17. Lei J, Li J, Li X, Qi X (2020) Ct imaging of the 2019 novel coronavirus (2019-ncov) pneumonia. *Radiology* 295(1):18–18
18. Himoto Y, Sakata A, Kirita M, Hiroi T, Kobayashi K-i, Kubo K, Kim H, Nishimoto A, Maeda C, Kawamura A et al (2020) Diagnostic performance of chest ct to differentiate covid-19 pneumonia in non-high-epidemic area in japan. *Japan J Radiol*:1
19. Ali M, Khan M, Tung NT et al (2018) Segmentation of dental x-ray images in medical imaging using neutrosophic orthogonal matrices. *Expert Syst Appl* 91:434–441
20. Razavian AS, Azizpour H, Sullivan J, Carlsson S (2014) Cnn features off-the-shelf: an astounding baseline for recognition
21. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*
22. Wu Z, Shen C, Van Den Hengel A (2019) Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recogn* 90:119–133
23. 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
24. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
25. Albawi S, Mohammed TA, Al-Zawi S (2017) Understanding of a convolutional neural network. In: *2017 International Conference on Engineering and Technology (ICET)*. IEEE, pp 1–6
26. Pan SJ, Yang Q (2009) A survey on transfer learning. *IEEE Trans Knowl Data Eng* 22(10):1345–1359
27. Sesmero MP, Ledezma AI, Sanchis A (2015) Generating ensembles of heterogeneous classifiers using stacked generalization. *Wiley Interdiscip Rev Data Min Knowl Discov* 5(1):21–34
28. Zagoruyko S, Komodakis N (2016) Wide residual networks. *arXiv:1605.07146*
29. He X, Yang X, Zhang S, Zhao J, Zhang Y, Xing E, Xie P (April 2020) Sample-Efficient Deep Learning for COVID-19 Diagnosis Based on CT Scans. *medRxiv*, pp 2020.04.13.20063941
30. Polsinelli M, Cinque L, Placidi G (April 2020) A Light CNN for detecting COVID-19 from CT scans of the chest. *arXiv:2004.12837* [cs, eess]
31. Loey M, Manogaran G, Khalifa NEM (May 2020) A Deep Transfer Learning Model with Classical Data Augmentation and CGAN to Detect COVID-19 from Chest CT Radiography Digital Images
32. Lokwani R, Gaikwad A, Kulkarni V, Pant A, Kharat A (June 2020) Automated Detection of COVID-19 from CT Scans Using Convolutional Neural Networks. *arXiv:2006.13212* [cs, eess]
33. Saleh AI, Shaban WM, Rabie AH, Abo-Elhoud MA (2020) A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. *Knowl-Based Syst* 205:106270
34. Azemin MZC, Hassan R, Tamrin MIM, Ali MAMd (2020) COVID-19 Deep Learning Prediction Model Using Publicly Available Radiologist-Adjudicated Chest X-Ray Images as Training Data: Preliminary Findings
35. CVDNet (2020) A novel deep learning architecture for detection of coronavirus (Covid-19) from chest x-ray images. *Chaos, Solitons Fractals* 140:110245

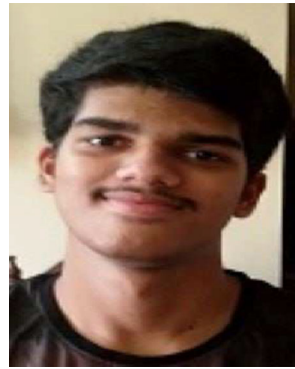
36. Taresh M, Zhu N, Ali TAA (August 2020) Transfer learning to detect COVID-19 automatically from X-ray images, using convolutional neural networks. medRxiv, pp 2020.08.25.20182170
37. Yadav SS, Bendre MR, Vikhe PS, Jadhav SM (August 2020) Analysis of deep machine learning algorithms in COVID-19 disease diagnosis. arXiv:2008.11639 [cs, eess]
38. Rahimzadeh M, Attar A, Sakhaei SM (September 2020) A Fully Automated Deep Learning-based Network For Detecting COVID-19 from a New And Large Lung CT Scan Dataset. medRxiv, pp 2020.06.08.20121541
39. Wang Z, Liu Q, Dou Q (2020) Contrastive Cross-site Learning with Redesigned Net for COVID-19 CT Classification. IEEE J Biomed Health Inf:1–1
40. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, Yu L, Chen Y, Su J, Lang G, Li Y, Zhao H, Xu K, Ruan L, Wu W (February 2020) Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia. arXiv:2002.09334 [physics]
41. Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P (June 2020) COVID-CT-Dataset: A CT Scan Dataset about COVID-19. arXiv:2003.13865 [cs, eess, stat]
42. Wang L, Wong A (May 2020) COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. arXiv:2003.09871 [cs, eess]
43. Farooq M, Hafeez A (March 2020) COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs. arXiv:2003.14395 [cs, eess]
44. Zhao J, Zhang Y, He X, Xie P (2020) Covid-ct-dataset: a ct scan dataset about covid-19. arXiv:2003.13865
45. Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M (2020) Covid-19 image data collection: Prospective predictions are the future. arXiv:2006.11988
46. Rahimzadeh M, Attar A, Sakhaei SM (2020) A fully automated deep learning-based network for detecting covid-19 from a new and large lung ct scan dataset. medRxiv
47. Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam KR, Khan MS, Iqbal A, Emadi NA, Reaz MBI, Islam MT (2020) Can ai help in screening viral and covid-19 pneumonia? IEEE Access 8:132665–132676
48. Soares E, Angelov P, Biaso S, Higa Froes M, Kanda Abe D (2020) Sars-cov-2 ct-scan dataset: A large dataset of real patients ct scans for sars-cov-2 identification. medRxiv

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



conferences. His research interests include Deep Learning, Big data analytics and medical image processing.

Ebenezer Jangam is currently working as an Assistant Professor in the Department of Information Technology at Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India. He is pursuing PhD from the Department of Computer Science and Engineering, Indian Institute of Technology (Indian School of Mines) Dhanbad, Jharkhand, India. He has more than 9 years of teaching experience and published twenty research papers in reputed journals and



Aaron Antonio Dias Barreto is an undergraduate student in the Department of Electronics Engineering at Indian Institute of Technology (Indian School of Mines) Dhanbad, pursuing Bachelor of Technology in Electronics and Communication Engineering. His research interests include machine learning and Artificial Intelligence.



in reputed journals and conferences. His research interests include Data and Knowledge-based System, Machine Learning, Evolutionary Algorithms, and Bioinformatics.

Chandra Sekhara Rao Annavarapu is currently working as an Assistant Professor in the Department of Computer Science and Engineering at Indian Institute of Technology (Indian School of Mines) Dhanbad, India. He received his PhD from Computer Science and Engineering department, Indian Institute of Technology (Indian School of Mines) Dhanbad, India. He has more than 15 years of teaching experience and published many research papers