



# ET-NET: an ensemble of transfer learning models for prediction of COVID-19 infection through chest CT-scan images

Rohit Kundu<sup>1</sup> · Pawan Kumar Singh<sup>2</sup> · Massimiliano Ferrara<sup>3,4</sup> · Ali Ahmadian<sup>5,6,7</sup> · Ram Sarkar<sup>8</sup>

Received: 10 November 2020 / Revised: 7 July 2021 / Accepted: 21 July 2021 /  
Published online: 31 August 2021

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

## Abstract

The COVID-19 virus has caused a worldwide pandemic, affecting numerous individuals and accounting for more than a million deaths. The countries of the world had to declare complete lockdown when the coronavirus led to community spread. Although the real-time Polymerase Chain Reaction (RT-PCR) test is the gold-standard test for COVID-19 screening, it is not satisfactorily accurate and sensitive. On the other hand, Computer Tomography (CT) scan images are much more sensitive and can be suitable for COVID-19 detection. To this end, in this paper, we develop a fully automated method for fast COVID-19 screening by using chest CT-scan images employing Deep Learning techniques. For this supervised image classification problem, a bootstrap aggregating or Bagging ensemble of three transfer learning models, namely, Inception v3, ResNet34 and DenseNet201, has been used to boost the performance of the individual models. The proposed framework, called ET-NET, has been evaluated on a publicly available dataset, achieving  $97.81 \pm 0.53\%$  accuracy,  $97.77 \pm 0.58\%$  precision,  $97.81 \pm 0.52\%$  sensitivity and  $97.77 \pm 0.57\%$  specificity on 5-fold cross-validation outperforming the state-of-the-art method on the same dataset by 1.56%. The relevant codes for the proposed approach are accessible in: [https://github.com/Rohit-Kundu/ET-NET\\_Covid-Detection](https://github.com/Rohit-Kundu/ET-NET_Covid-Detection)

**Keywords** COVID-19 screening · Computer-aided detection · Deep learning · Coronavirus · Transfer learning · Bagging ensemble classifier · CT-scan image

## 1 Introduction

The COVID-19 pandemic caused by the novel coronavirus or SARS-CoV-2 originated in Wuhan, China and has affected more than 100 million people worldwide with more than 2 million deaths during January-October 2020. Although the mortality rate has dropped, the

---

✉ Ali Ahmadian  
ali.ahmadian@ukm.edu.my

Extended author information available on the last page of the article

pandemic is not over yet. The tests performed for the detection of COVID-19 are the rapid IgM-IgG combined antibody test [32] and the real-time Polymerase Chain Reaction (RT-PCR) [48]. The RT-PCR test has several limitations: (1) A long time is required for obtaining the test results; (2) It is a costly test requiring experts to perform the test and analyze the results (3) They have a high false-negative rate (sensitivity of 71%) [55]. Although the rapid antigen test can produce results within 15 minutes by detection of the IgG and IgM antibodies simultaneously in the human blood, it might take several days for the human body to form the antibodies and thus there is a risk of spread of the virus before being detected. This leads to a very high false-negative rate. Hence, as an alternative, an automated diagnosis tool is required that is sensitive as well as specific to the COVID-19 disease which can lead to fast predictions.

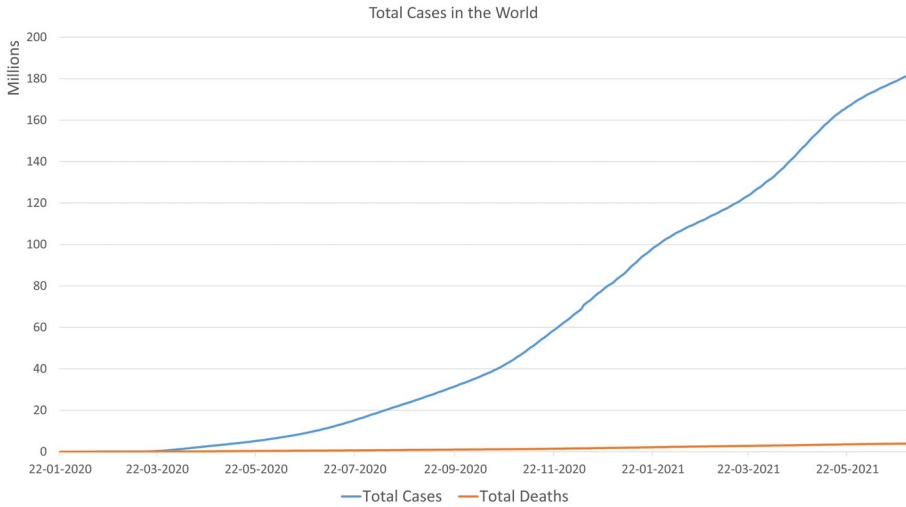
Currently, in the worldwide scenario, there are 108.3 million total COVID-19 cases and 2.38 million total deaths. The plots for the total cases and daily cases are shown in Fig. 1a, b, respectively. In India, there are more than 10.8 million cases in total and 155,000 deaths as shown in Fig. 2a and the daily cases and mortality rates are shown in Fig. 2b (All data for the graphs have been collected from the publicly available data by Roser et al. [35]). Due to the acute shortage of RT-PCR test kits, especially in developing countries like India, population-wise screening is not possible, which has led to uncontrolled community spread of the virus. Also, the RT-PCR test is a tedious and time-consuming process, so an appropriate and viable option can be the use of chest CT-scan images for COVID-19 screening.

Computed Tomography or CT-scan is a relatively common test [34] and can be performed more amply. It is much more sensitive (98%) than RT-PCR (71%) as established by Fang et al. [18]. Figure 3 shows two CT images, one of which is of a COVID-19 infected patient and the other, a tested negative patient. The most common finding from the chest CTs is “ground-glass opacities” scattered throughout the lungs. They represent tiny air sacs or alveoli, getting filled with fluid, and turning a shade of grey in the CT-scan turning into a white consolidation in more severe cases, as marked by the red circle in Fig. 3a. The disease severity is proportionate to the lung findings, meaning sicker individuals have more of such opacities in one of both lobes of the lungs in chest CT-scans.

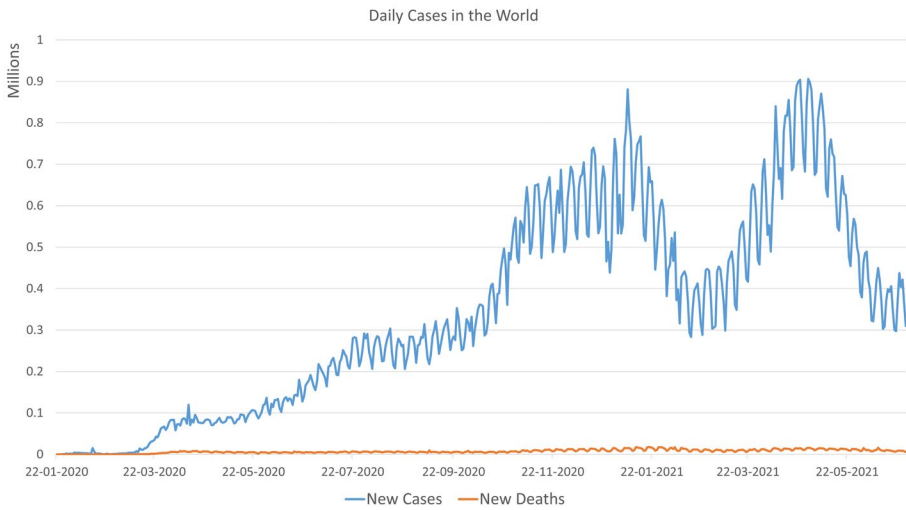
Also, to aid the clinicians in COVID-19 screening, automation based methods need to be developed that is both reliable and fast. Hence, researchers around the world have tried developing Computer-Aided Diagnosis tools for the detection of COVID-19 from chest X-rays or chest CT images. Chest CT images reveal more details than chest X-rays, and hence ET-NET considers chest CT images for the prediction of COVID-19 positive patients. Deep learning [30] is a powerful machine learning tool that uses structured or unstructured data for classification using a complex decision-making process.

The current image classification problem [13] is a supervised learning task. Supervised learning [4, 22] refers to the learning procedure, where an algorithm is trained on a labelled dataset, meaning that the true classes of the samples are already provided for the model to tune its parameters based on the training accuracy. Transfer learning is a technique where a deep learning model used for one task is utilized for another separate task. This method is particularly effective when the task at hand has less amount of data available for training the model, and the parameters trained from the previous task are loaded and trained with the new data for fine-tuning.

Ensemble learning allows the fusion of the salient features of multiple base learners, leading to more accurate predictions than the individual models. Such a learning scheme is robust since the variance in the prediction errors is reduced upon ensembling. An ensemble model aims to capture the complementary information from the base models and makes superior predictions. In the present study, the Bagging technique is used as a method to



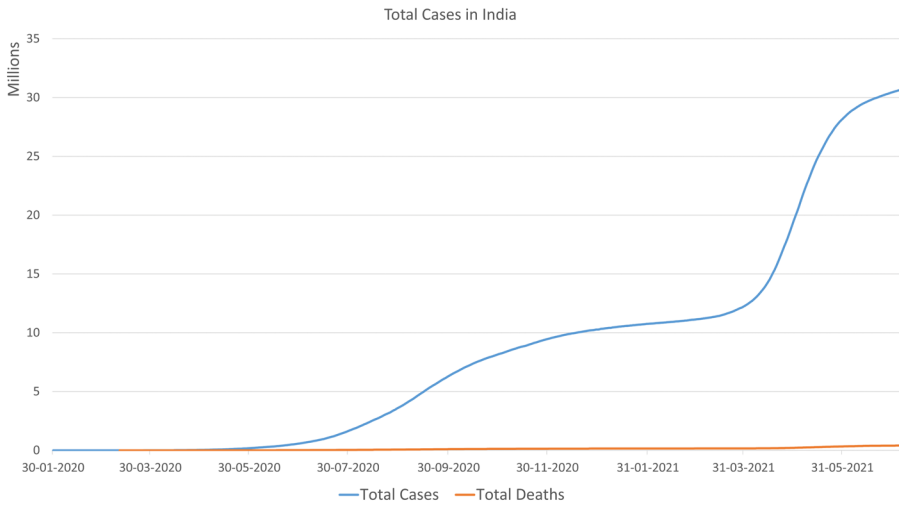
(a) Total Cases and Deaths



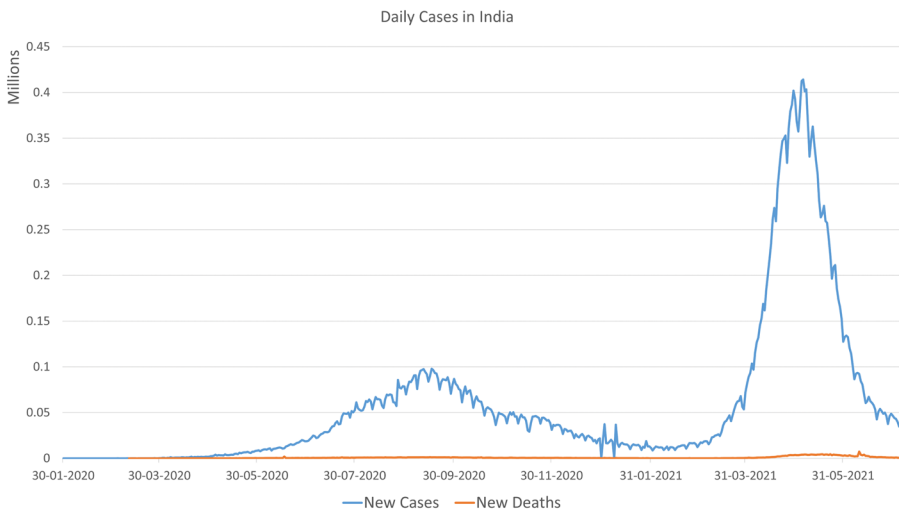
(b) Daily Cases and Deaths

**Fig. 1** Current worldwide statistics of COVID-19 (a) Total cases and deaths in the world (b) Daily new cases and deaths in the world [35]

fuse the important aspects of all the transfer learning models considered here to form the ensemble. The Bagging technique is preferred over the Boosting algorithm in the present work, since the dataset available has only a small amount of data, and might lead to excessive overfitting using the Boosting technique. The bagging technique, on the other hand, reduces overfitting and hence is beneficial for the current problem. Thus, the ET-NET or Ensemble Transfer learning Network is proposed in this paper.



**(a)** Total cases and deaths due to COVID-19 in India



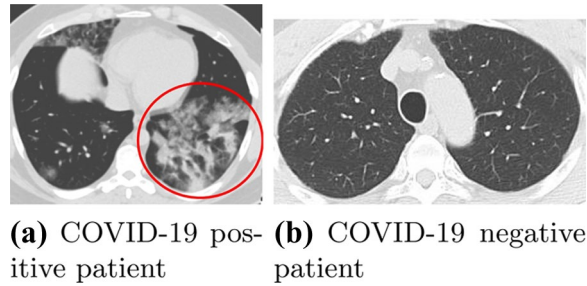
**(b)** Daily new cases and deaths due to COVID-19 in India

**Fig. 2** Current statistics of COVID-19 in India: **(a)** Total cases and deaths **(b)** Daily new cases and deaths [35] (the graph has been formed from the data using Google Sheets)

## 2 Literature survey

A vast amount of research is being conducted to help stop the COVID-19 pandemic [5]. However, the existing methods are time consuming and expensive, while also being less accurate. Yang et al. [54] showed that chest CT-scans can serve as an important make-up for the diagnosis of COVID-19. They used respiratory samples including nasal and throat

**Fig. 3** Illustration of chest CT image findings of two patients having: **(a)** COVID-19 positive and **(b)** COVID-19 negative. The COVID-19 infection's characteristic "Ground Glass Opacity" has been marked with a red circle in the COVID-19 infected chest CT image



swabs, and bronchoalveolar lavage fluid (BALF) to draw comparisons. The accuracy in the detection of COVID-19 was only 88.9% for severe cases and 82.2% for mild cases using sputum samples. Nasal swabs and throat swabs gave even lower accuracies (73.3% for nasal swabs and 60.0% for throat swabs). Table 1 shows some of the recent methods proposed in the literature for the automated diagnosis of COVID-19 from either CT-scan or Chest X-ray images.

Several automated frameworks for the screening of COVID-19 infected patients have been proposed since the outbreak of the pandemic, a majority of which have used chest X-ray images [11, 50]. According to clinicians and doctors, CT-scans are more reliable and sensitive than radio-graph (X-ray) images, and hence a better input for screening.

Deep learning has been widely used as a computer-aided detection tool for COVID-19 screening like in [31, 57]. Gozes et al. [23] utilized deep learning by fusing two subsystems, one being a 2D slice model and the other a 3D volumetric model for CT image classification. Li et al. [31] developed COVNet for extracting visual features from volumetric chest CT images. The COVNet they developed extracted both 2D local and 3D global features using a ResNet50 backbone and fused the features using a max-pooling layer and employing a final fully connected layer for generating the probability scores.

Zhang et al. [56] proposed a novel deep learning model for utilizing 3D chest CT volumes for the classification of infected patients and localization of swelling regions in the CT-scans. They used a pretrained U-net for segmentation of the 3D CT-scans and fed the 3D segmented chest areas into a deep neural network for forecasting the infection probability. The computation time for the detection of test images in their method is only 1.93 seconds per image. Abdel et al. [1] proposed a semi-supervised meta learning-based lung segmentation model for COVID-19 detection. Karbhari et al. [29] proposed a Generative Adversarial Network (GAN) framework to address the challenge of data scarcity for COVID-19 detection and used the generated data for training a classification model. Das et al. [14] proposed a bi-level classification model that uses pre-trained VGG-19 for feature extraction and then a shallow classifier for the final predictions. Sen et al. [42] and Chattopadhyay et al. [11] proposed deep features extraction and classification framework using meta-heuristics to reduce the feature set dimensionality. Garain et al. [21] developed a Spiking Neural Network-based model for the detection of COVID-19 from CT-scan images.

Most of the previous methods as shown in Eq. (1) use a single model for the predictions, however, we propose an ensemble scheme for the detection of COVID-19. Using the complementary information provided by the different base classifiers, based on the confidence scores, enhances the overall performance and robustness by reducing the variation in prediction errors. The ensemble method is a kind of fusion mechanism that uses the outputs or features from more than one model to compute the final prediction

**Table 1** Some recent methods for COVID-19 detection

Work Ref.	Approach	Type of Images
Wang et al. [50]	Tailored Covid-Net CNN Model	X-ray
Chattopadhyay et al. [11]	Golden Ratio Optimizer-based feature selection on deep features	X-ray
Zhang et al. [57]	Residual CNN model	X-ray
Karbhari et al. [29]	GAN for synthetic data generation. Classification using deep feature extraction and Harmony Search-based feature selection	X-ray
Sen et al. [42]	Deep feature extraction using VGG-19 and classification using shallow classifiers like KNN, XGB, etc.	X-ray
Zhang et al. [56]	Lung Segmentation using U-Net and classification using a novel 3D-CNN	3D CT-scan
Li et al [31]	Novel COVNet CNN Model	CT-scan
Gozes et al. [23]	Fusion of 2D slice model and 3D volumetric model	CT-scan
Abdel et al. [1]	Few shot semi-supervised lung segmentation using deep learning	CT-scan
Garain et al. [21]	Spiking Neural Network	CT-scan
Angelov and Soares [6]	Feature extraction using non-pretrained GoogLeNet and classification using MLP classifier	CT-scan
Panwar et al. [37]	Modified VGG-19 transfer learning model	CT-scan
Jaiswal et al. [28]	Transfer learning using DenseNet201	CT-scan
Wang et al. [51]	Novel joint learning framework and contrastive learning	CT-scan
Silva et al. [43]	Majority Voting between predictions on different slices of the same patient	CT-scan

of the input [36, 39, 40]. It aims to enhance the performance of the framework beyond the reach of the individual models. Ensemble learning works better than the individual models, because of the diversification of the information considered. When more than one model's opinion is accounted for, less noisy predictions are produced. Hence, such a technique has been employed in the present work. A large variety of ensemble techniques [17, 25, 36, 52] have been proposed in literature, two of the most popular techniques being Bagging [10, 38] and Boosting [9].

## 2.1 Motivation and contributions

In light of the current pandemic situation, the medical practitioners and healthcare professionals are working tirelessly, fighting the disease. However, the current gold standard method for COVID-19 screening, the RT-PCR test, is slow and tedious, and hence inadequate for population-wise screening resulting in an uncontrolled number of infected individuals. Several researchers, therefore, are trying to develop systems for faster and more efficient screening of the infected patients, which is the primary motivation behind the current paper. (**Vaccine?**) Ensemble learning allows the fusion of salient properties of the base classifiers, thus achieving an overall enhanced performance. Such models are robust since computing the ensemble model decreases the spread (or dispersion) of the predictions of the base models. That is, the variance in the prediction errors are diminished and complementary information is captured. Figure 4 shows a diagram depicting the overall workflow of the proposed ET-NET model.

The contributions of this paper are as follows:

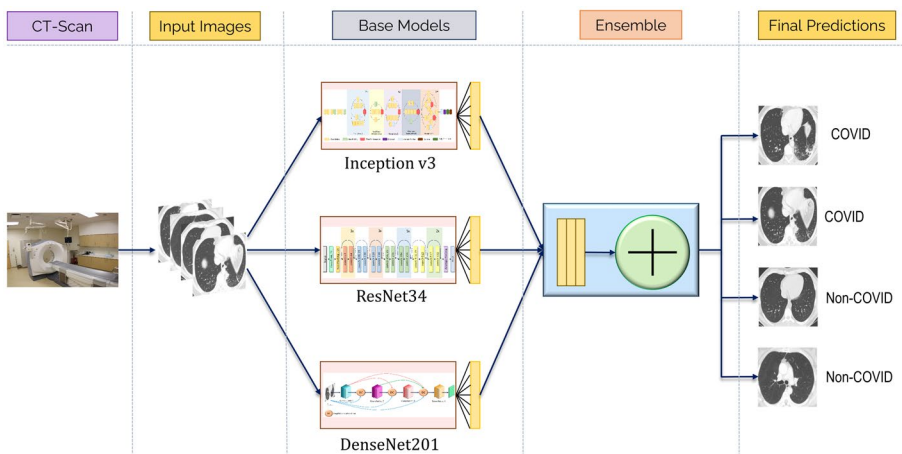
1. An ensemble-based COVID detection approach has been used that boosts the performance of the individual CNN classifiers: Inception v3 [47], ResNet34 [24] and DenseNet201 [27]. For this, a bagging ensemble technique has been used that uses the average of the decision scores generated by each model for each class of the dataset.
2. The proposed model, called ET-NET, has been evaluated on a publicly available dataset [45] using 5-fold cross-validation, outperforming the previous state-of-the-art method by 1.56%.
3. Most of the previous works considered chest X-ray images which are less sensitive than lung CT images used in this work. To account for the less availability of publicly available data, Transfer Learning has been used to generate the decision scores. The ensembling technique helps capture complementary information, thus outperforming individual models.

CT-scan images have been used, generally requiring no prior segmentation, for the classification of the chest CT-scans into two categories: COVID or Non-COVID.

The rest of the paper has been organized as follows: Sect. 3: Proposed Method, explains in detail the working of ET-NET in the current study; Sect. 4: Results and Discussion, highlights the results obtained by the ET-NET on a publicly available dataset, compares it to existing models and discusses the efficacy of ET-NET and Sect. 5: Conclusions, concludes the findings and contributions of this paper, and discusses the possibilities of future works on the proposed model.

### 3 Proposed method

Convolutional Neural Networks (CNNs) are preferred for image classification problems since an image is a 2D matrix of pixel intensities, and it might help to look at an image in parts, for example, a 300x300 image can be seen 3x3 parts at a time, for, say, object



**Fig. 4** Overall workflow of the proposed ET-NET ensemble classifier model for COVID-19 detection from chest CT-scan images

detection, etc., which is achieved by the convolution operation. The pooling operation [53] helps in dimensionality reduction. CNNs are shift-invariant [49, 58] and have less number of parameters in comparison to deep fully connected neural networks and hence are computationally more efficient even while accommodating a very deep network [19, 20].

---

#### Algorithm 1 Pseudo-Algorithm of ET-NET

---

##### Inputs:

$D \rightarrow$  Dataset

Optimizer  $\rightarrow$  SGD

$\alpha \rightarrow$  Learning Rate = 0.001

$\beta \rightarrow$  Momentum = 0.99

BS  $\rightarrow$  Batch Size = 4

$\epsilon \rightarrow$  Number of epochs = 100

Models  $\rightarrow$  [Inception v3, ResNet34, DenseNet201]

$\omega \rightarrow$  Pretrained Weights

##### Output:

Predicted Class of image

##### for *model* in Models do

    Initialize the training and validation sets for *model*

    Load  $\omega$  for *model*

##### while *epochs* $\leq \epsilon$ do

    Perform forward propagation and compute binary cross-entropy loss on train set

    Perform backward propagation and update SGD optimizer

##### end while

    Save prediction probability scores for *model* on validation set

##### end for

##### for image $\in$ test set do

    final probability score = average probability scores from the three models

    prediction = class with highest probability

##### end for

---

In the proposed work, three models namely, Inception v3 [47], ResNet34 [24] and DenseNet201 [27] pretrained on ImageNet [15] have been used, which are then fine-tuned using the chest CT-scan dataset. The number of layers and parameters of each deep transfer learning model have been shown in Table 2.

### 3.1 Inception v3

The characteristic features of the Inception v3 model developed by Szegedy et al. [47] in 2016, are the three types of inception block, which have parallel convolutions. Such modules account for more efficient computation in the deep architecture, while also addressing the overfitting problem. The architecture of the Inception v3 CNN has been illustrated in Fig. 5a.



**Table 2** Number of layers and parameters in each network

CNN	No. of layers	Filter size used in various layers	No. of parameters
<b>Inception v3</b> [47]	48	(1x1), (3x3), (5x5), (1x7), (7x1)	21.8M
<b>ResNet34</b> [24]	35	(3,3), (7,7)	21.3M
<b>DenseNet201</b> [27]	201	(1x1), (3x3), (7x7)	20.2M

### 3.2 ResNet34

The salient features of Residual Networks or ResNets developed by He et al. [24] in 2016 are that they have skip connections that directly concatenate the current layer with features from a previous layer, resulting in preservation of features from past layers, which might be important. ResNet34 is one such network that is 34 layers deep (and one fully connected classification layer), the architecture of which is shown in Fig. 5b.

### 3.3 DenseNet201

In DenseNets by Huang et al. [27] in 2017, each layer is a concatenation of feature maps of the current layer and all preceding layers. As a result, these networks are compact (that is, less number of channels), and hence in terms of computation and memory requirement, it is efficient, while also having rich features representation for the input images. The architecture of the DenseNet201 is shown in Fig. 5c.

### 3.4 Loss function

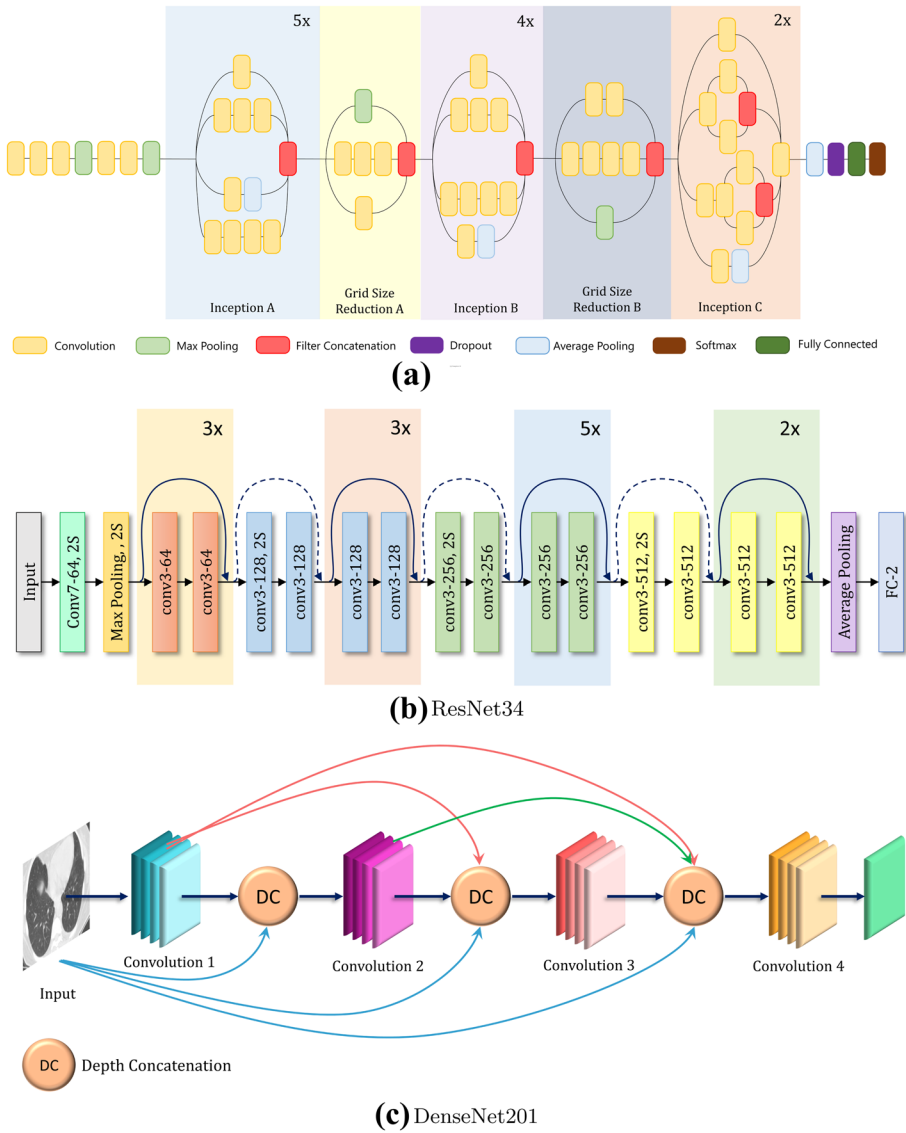
A loss function is a measure of the performance of a deep learning model. The main objective of a deep learning model is to minimize the error between the predicted and the original labels, which is calculated during backward propagation [12] in a neural network.

In the current study, the cross-entropy loss function is used, which evaluates the performance of the classifier which outputs a matrix of probabilities (each probability value between 0 and 1). Since the present classification problem deals with only two classes, the loss function is called Binary Cross-Entropy Loss function. The cross-entropy loss function is chosen since it performs well for binary classification problems which have a large decision boundary [33]. This loss function also helps curb the vanishing gradient descent problem since the use of logarithm nullifies any exponential behaviour which occurs due to the sigmoid (or softmax) activation function. The logarithm avoids saturation of the gradients at extreme values which is beneficial since large gradients are essential for making significant progress through the iterations.

Suppose for an input  $x$ , the true label is  $y$  and the predicted label from the classifier is  $\hat{y}$ , which is given by Eq. 1, where  $w$  is the weight matrix associated with the neural network and  $b$  is the bias matrix associated with it.  $f$  is the non-linear activation function associated with the layers in the neural network. For the present work, the activation function Rectified Linear Unit or ReLU [3] has been used.

$$\hat{y} = f(w^T \cdot x + b) \quad (1)$$

The ReLU activation function is given as in Eq. 2.



**Fig. 5** Architectures of the three CNN base classifiers: **(a)** Inception v3, **(b)** ResNet34, and **(c)** DenseNet201 used to form the proposed ensemble model called ET-NET

$$ReLU(x) = \max(0, x) \tag{2}$$

Then the loss function  $L$  is given by Eq. 3 where  $N$  denotes the number of classes in the problem.  $N = 2$  for the present study.

$$L(\hat{y}^{(i)}, y^{(i)}) = - \sum_{i=1}^N y^{(i)} \log \hat{y}^{(i)} \tag{3}$$

For  $m$  training samples, the cost function is given by Eq. 4

$$J(w, b) = -\frac{1}{m} \sum_{i=1}^m L(\hat{y}^{(i)}, y^{(i)}) \tag{4}$$

Using the cost function in Eq. 4, the weights and biases associated with the layers in the neural networks are updated.

### 3.5 Ensemble

The ensemble approach adopted for the current is the bootstrap aggregating or “Bagging” ensemble [8]. This machine learning-based ensemble technique was developed to make the machine learning classification algorithms more stable and accurate. Bagging ensemble techniques help to reduce overfitting problems, in contrast to Boosting ensemble technique [41] which increases the overfitting problem, because, in each stage of the Boosting algorithm, only the misclassified samples from the previous stage are used as training data.

In the current study, the Bagging ensemble technique uses the same training set for training the three pretrained models (Inception v3, ResNet34 and DenseNet201) independently and then predicts the class probabilities of the samples in the test set by the fine-tuned models to calculate the average probability score, thus giving equal weightage to all the three classifiers.

Suppose  $m$  models (classifiers) numbered as  $1, 2, \dots, m$  are used for a classification task of  $n$  classes, and the prediction probability scores are denoted by  $P$ . The prediction scores for a single image from model  $i$  can be expressed as a matrix as in Eq. 5.

$$P^{(i)} = \begin{bmatrix} p_1^{(i)} & p_2^{(i)} & \dots & p_n^{(i)} \end{bmatrix} \tag{5}$$

So the final prediction score  $P^{ensemble}$  using the average probability ensemble technique is given by Eq. 6.

$$\begin{aligned} P^{ensemble} &= \frac{\sum_{i=1}^m P^{(i)}}{m} \\ &= \left[ \frac{\sum_{i=1}^m p_1^{(i)}}{m} \quad \frac{\sum_{i=1}^m p_2^{(i)}}{m} \quad \dots \quad \frac{\sum_{i=1}^m p_n^{(i)}}{m} \right] \\ &= [p'_1 p'_2 \dots p'_n] \end{aligned} \tag{6}$$

Now, the class having the maximum probability out of the values  $p'_1, p'_2, \dots, p'_n$  is decided as the predicted class, which is then compared with the true labels to obtain the accuracy. In the current problem, there are 3 models and 2 categories to sort the images into, accounting for  $m = 3$  and  $n = 2$  in Eqs. 5 and 6.

## 4 Results and discussion

In this section, we will briefly describe the dataset used for the current study in Sect. 4.1, the evaluation metrics used for comparing and validating ET-NET in Sect. 4.2. The implementation of the developed methodology and the results thus obtained, are described in detail in Sect. 4.3, and the comparison with the existing literature and standard models are made in Sect. 4.5.

**Table 3** Class-wise distribution of images in the Kaggle dataset

Class	Category	Number of Images
1	COVID	1249
2	Non-COVID	1229

#### 4.1 Dataset description

For evaluating the performance of the proposed methodology, the dataset used is publicly available on Kaggle<sup>1</sup> developed by Soares et al. [45]. The dataset consists of a total of 2481 CT-scan images unevenly distributed into COVID and Non-COVID categories as shown in Table 3. For the proposed framework, 70% of the images (1736 scans) are used as training data and the rest 30% (745 scans) are used as testing data.

#### 4.2 Evaluation metrics

For evaluating the performance of ET-NET on the binary classification task at hand, parameters such as accuracy, precision, recall (or sensitivity), f1 score and specificity. For defining these terms, first the terms True Positive, True Negative, False Positive and False Negative needs to be defined.

In a binary classification problem, suppose the two classes are a positive class and a negative class. *True Positive (TP)* refers to a sample belonging to the positive class, being classified correctly. *False Positive (FP)* refers to a sample belonging to the negative class, but classified to be belonging to the positive class. Similarly, *True Negative (TN)* refers to a sample being classified correctly as belonging to the negative class. *False Negative (FN)* refers to a sample belonging to the positive class, but classified as being part of the negative class. Now the metrics can be defined as follows:

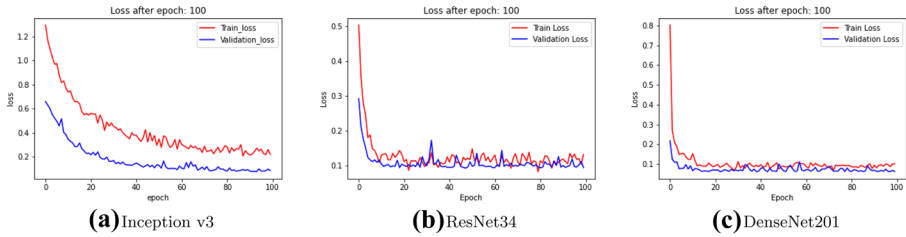
$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

**Table 4** Hyperparameters used for training each model

Hyperparameters	Values Used
Batch size	4
Number of Epochs	100
Optimizer	Stochastic Gradient Descent
Learning Rate	0.001
Momentum	0.99
Input Size	299x299x3 (Inception v3), 224x224x3 (Others)
Loss Criterion	Binary Cross-Entropy Loss

<sup>1</sup> <https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset>



**Fig. 6** Loss curves obtained using the base pre-trained classifiers after 100 epochs of re-training: (a) Inception v3 (b) ResNet34 and (c) DenseNet201

$$Recall \text{ (or Sensitivity)} = \frac{TP}{TP + FN} \tag{9}$$

$$F1Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{10}$$

$$Specificity = \frac{TN}{TN + FP} \tag{11}$$

### 4.3 Implementation

The CNN transfer learning models have been trained for 100 epochs, and the loss curves of the models have been shown in Fig. 6. The predictions of the models on the test set images have been saved. The hyperparameters used for training the three models are shown in Table 4.

The probability prediction matrices from the three classifiers have been averaged per sample to get the final prediction scores, and hence the predicted result for all the images are obtained.

The class-wise metrics obtained have been shown in Table 5 and the net result has been shown in Table 6. The confusion matrix for the test set has been shown in Fig. 7 and the Receiver Operating Characteristics (ROC) curves of the individual models and the proposed ET-NET are shown in Fig. 8.

**Table 5** Class-wise evaluation metrics generated by the base classifiers and the proposed ET-NET model on Fold-4 (best fold) of 5-fold cross-validation

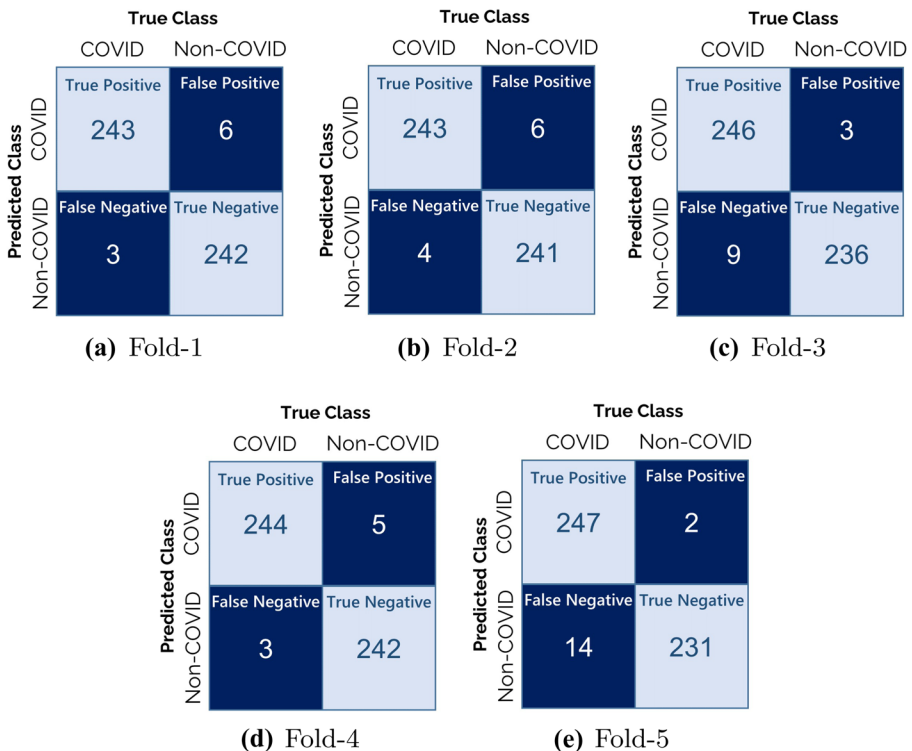
Class	Metrics	Inception v3	ResNet34	DenseNet201	Proposed ET-NET
COVID	Accuracy(%)	97.59	94.78	97.19	<b>98.79</b>
	Precision(%)	97.99	95.16	96.80	<b>98.38</b>
	Recall (or Sensitivity)(%)	97.59	94.78	97.19	<b>98.79</b>
	F1 Score(%)	97.98	94.97	96.99	<b>98.39</b>
	Specificity(%)	97.57	94.72	97.13	<b>98.78</b>
Non-COVID	Accuracy(%)	<b>98.37</b>	95.10	96.73	97.98
	Precision(%)	97.57	94.72	97.16	<b>98.78</b>
	Recall (or Sensitivity)(%)	<b>98.37</b>	95.10	96.73	97.98
	F1 Score(%)	97.97	94.91	96.93	<b>98.37</b>
	Specificity(%)	97.99	95.16	96.80	<b>98.38</b>

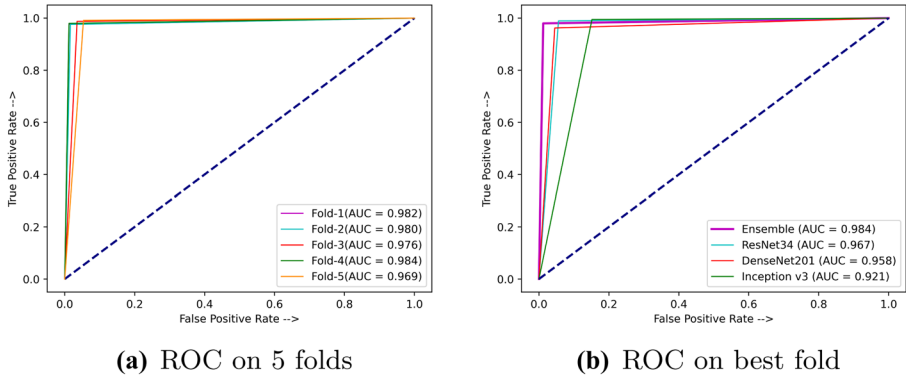
**Table 6** Evaluation metrics produced by the proposed ET-NET model on 5-fold cross-validation of the dataset

5-Fold cross-validation	Accuracy(%)	Precision(%)	Recall(%)	F1 Score(%)	AUC(%)
Fold-1	98.181	98.18	98.18	98.18	98.18
Fold-2	97.976	97.98	97.98	97.98	97.98
Fold-3	97.608	97.56	97.61	97.57	97.61
Fold-4	98.381	98.38	98.38	98.38	98.38
Fold-5	96.889	96.74	96.89	96.76	96.89
Average $\pm$ Std. Dev.	97.81 $\pm$ 0.53	97.77 $\pm$ 0.58	97.81 $\pm$ 0.52	97.77 $\pm$ 0.57	97.81 $\pm$ 0.53

#### 4.4 Error analysis

ET-NET performs very well for the current classification problem. Examples of correctly classified images from each class are shown in Fig. 9. In both the images, a part of the lungs has not been captured by the CT-scan properly, and as a result, it is an erroneous image. The contrast for Fig. 9a is also too high, while Fig. 9b is a hazy image. Even with all these limitations of the images in the dataset, ET-NET was able to classify them correctly, proving the model to be reliable even for imperfect imaging conditions. Hence, slightly noisy images do not affect the performance of ET-NET.

**Fig. 7** Confusion matrices of the predictions produced by the proposed ET-NET model on 5-Fold cross-validation of the dataset: (a) Fold-1 (b) Fold-2 (c) Fold-3 (d) Fold-4 and (e) Fold-5



**Fig. 8** Receiver Operating Characteristics (ROC) curves obtained on (a) 5 folds of cross-validation (b) the test set of Fold-4 (best result) and base CNN classifiers

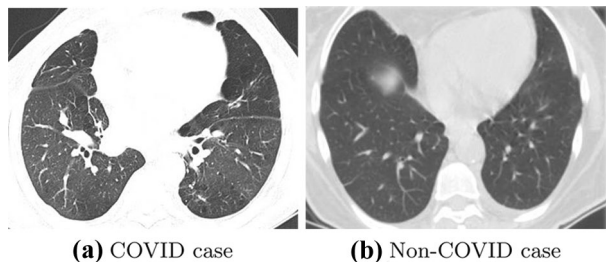
Figure 10 shows one misclassified image from each class of the dataset. Figure 10a belongs to class “COVID” of the dataset but was classified by ET-NET as “Non-COVID”. The prime reason for that is, the lung condition depicted in the CT-scan is one of a mild COVID condition, as a result, prominent ground-glass opacity has not yet developed in the lung alveoli. So, ET-NET was unable to detect the presence of COVID-19 infection from such a preliminary stage of infection. Figure 10b on the other hand, is a sample belonging to the “Non-COVID” class of the dataset, but ET-NET predicted it to be a “COVID” condition. One of the reasons for that is, the lung CT-scan quality is not appropriate, because visibly the lung shape has not been properly captured. The other reason might be the fact that the CT-scan is very hazy unlike the low level of noise present in Fig. 9b.

### 4.5 Comparison with existing models

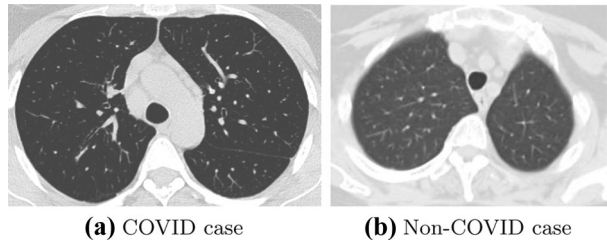
Several transfer learning models have been used for comparing the performance of the proposed approach, which has been shown in Table 7. Table 8 shows the comparison of ET-NET with some existing methods that use the same dataset.

Angelov and Soares [6] extracted features from non-pretrained GoogLeNet [46] and used a Multi-Layer Perceptron (MLP) for final classification. Panwar et al. [37] used the VGG19 [44] transfer learning model and added five more layers ahead and trained the network. Jaiswal et al. [28] used deep transfer learning technique with DenseNet201 [27] for feature extraction and classification.

**Fig. 9** Examples of test cases where the proposed ET-NET model performs correct classification although the images were noisy: (a) COVID case and (b) Non-COVID case



**Fig. 10** Examples of test set images where the proposed ET-NET model fails to produce correct predictions: **(a)** COVID case and **(b)** Non-COVID case



## 4.6 Statistical test

The McNemar’s test [16] is performed for statistically analysing the performance of the proposed ET-NET ensemble model with the base CNN classifiers which have been used to form the ensemble, and other standard transfer learning classifiers. McNemar’s test is a non-parametric analysis of paired nominal data distribution. Table 9 displays the results obtained from McNemar’s test on the Kaggle dataset. The “ $p$  – value” signifies the probability that two models are similar, thus, a lower  $p$  – value is desired. To reject the null hypothesis that the two models are similar, the  $p$  – value needs to be smaller than 5% that is, if  $p$  – value  $<$  0.05, we can safely say that the two models under consideration are significantly different.

In Table 9, it can be noted that for every model with which the ET-NET is compared,  $p$  – value  $<$  0.05, thus rejecting the null hypothesis. So, it can be said that the proposed ensemble model captures complementary information from the constituent base classifiers, thus producing superior results while making the ensemble model markedly dissimilar from the base classifiers.

**Table 7** Comparison of ET-NET with some standard deep learning models

Model	Accuracy(%)	Precision(%)	Recall(%)	F1 Score(%)	AUC(%)
Inception v3 [47]	96.98	96.98	96.98	96.98	96.98
DenseNet201 [27]	96.96	96.97	96.96	96.96	96.96
VGG19 [44]	96.76	96.85	96.74	96.76	96.74
DenseNet161 [26]	96.36	96.36	96.36	96.36	96.36
ResNet34 [24]	94.94	94.94	94.94	94.94	94.94
ResNet152 [24]	94.74	94.74	94.73	94.74	94.73
<b>Proposed ET-NET</b>	<b>97.81</b>	<b>97.77</b>	<b>97.81</b>	<b>97.77</b>	<b>97.81</b>

**Table 8** Comparison of the proposed ET-NET with some existing models in literature on the Kaggle dataset

Model	Accuracy(%)	Precision(%)	Recall(%)	F1 Score(%)	AUC(%)
Jaiswal et al. [28]	96.25	96.29	96.29	96.29	-
Panwar et al. [37]	94.04	95.00	94.00	94.50	-
Wang et al. [51]	90.83	95.75	85.89	90.87	96.24
Angelov and Soares [6]	88.60	89.70	88.60	89.15	89.20
Silva et al. [43]	87.60	-	-	86.19	90.50
<b>ET-NET</b>	<b>97.81</b>	<b>97.77</b>	<b>97.81</b>	<b>97.77</b>	<b>97.81</b>



**Table 9** Results of the McNemar’s test performed between ET-NET and standard CNN models on the Kaggle dataset: Null hypothesis is rejected for all cases

McNemar’s Test	p-value
<b>Inception v3</b> [47]	0.045121
<b>ResNet34</b> [24]	0.001814
<b>DenseNet201</b> [27]	0.002199
<b>VGG-19</b> [44]	0.001456
<b>DenseNet161</b> [26]	0.042345
<b>ResNet152</b> [24]	0.0004803

## 5 Conclusions

The spread of COVID-19 has collapsed economies of the world and caused numerous deaths, and people are still suffering due to this pandemic situation. Although RT-PCR is used for the screening of COVID-19 patients, it is a tedious process with low sensitivity. ET-NET uses a more sensitive CT-scan based detection using Computer-Aided Diagnosis. Deep transfer learning and an average probability-based ensemble approach have been utilized for the binary classification task which obtained results superior to existing CT-scan based screening models achieving an accuracy of 97.73% which is impressive for the small dataset used. Also, the sensitivity and specificity of the proposed ET-NET is better than RT-PCR and hence can be used as a reliable and robust COVID-19 detection mechanism. The proposed ET-NET model is also domain-independent, and can be extended to problems in gait detection [2], action recognition [7], etc.

The primary limitation of this method is that the non-availability of the data may deter to prove the robustness and generalization ability of the method. Deep learning models essentially perform best with a very large database, but, the dataset used in this study, has only 2481 images, whereas the more efficient deep learning models need to be trained on larger datasets depending on the complexity of the problem for optimal performance. As a result, we had to use transfer learning models that were pretrained on ImageNet consisting of 14 million images and then fine-tuned using the chest CT-scan images from this study. Also, other pulmonary diseases like the Middle East respiratory syndrome (MERS) and Chronic obstructive pulmonary disease (COPD) are possible biases to the present work as compared to RT-PCR and IgG-IgM antibody tests. It might also be important to perform segmentation to improve the Non-Covid control group design, which we intend to address in the future.

We aim to perform more experiments once more extensive datasets of chest CT-scan become available and develop better models for classification. We shall try to use image enhancement techniques to address the limitations mentioned in Sect. 4.4. We may try more pretrained models to form the ensemble and try more sophisticated ensemble approaches like Dempster-Shafer theory, Choquet fuzzy integral or rank based fusions.

**Acknowledgements** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This research was financially supported for Ali Ahmadian by the Ministry of Higher Education, Malaysia with Fundamental Research Grant Scheme (FRGS) with the reference no FRGS/1/2018/STG06/UPM/02/6.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Abdel-Basset M, Chang V, Hawash H, Chakraborty RK, Ryan M (2021) Fss-2019-ncov: A deep learning architecture for semi-supervised few-shot segmentation of covid-19 infection. *Knowl-Based Syst* 212:106647
2. Achanta SDM, Karthikeyan T, Vinothkanna R (2019) A novel hidden markov model-based adaptive dynamic time warping (hmdtw) gait analysis for identifying physically challenged persons. *Soft Comput* 23(18):8359–8366
3. Agarap AF (2018) Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375
4. Al Hasan M, Chaoji V, Salem S, Zaki M (2006) Link prediction using supervised learning. In *SDM06: workshop on link analysis, counter-terrorism and security* 30:798–805
5. Ali A, Zhu Y, Zakarya M (2021) A data aggregation based approach to exploit dynamic spatio-temporal correlations for citywide crowd flows prediction in fog computing. *Multimed Tools Appl* pp 1–33
6. Angelov P, Soares E (2020) Explainable-by-design approach for covid-19 classification via ct-scan. medRxiv
7. Banerjee A, Singh PK, Sarkar R (2020) Fuzzy integral based cnn classifier fusion for 3d skeleton action recognition. *IEEE Trans Circuits Syst Video Technol*
8. Breiman L (1996) Stacked regressions. *Mach Learn* 24(1):49–64
9. Bühlmann P, Hothorn T et al (2007) Boosting algorithms: Regularization, prediction and model fitting. *Stat Sci* 22(4):477–505
10. Bühlmann PL (2003) Bagging, subbagging and bragging for improving some prediction algorithms. In *Research report/Seminar für Statistik, Eidgenössische Technische Hochschule (ETH)*, vol 113
11. Chattopadhyay S, Dey A, Singh PK, Geem ZW, Sarkar R (2021) Covid-19 detection by optimizing deep residual features with improved clustering-based golden ratio optimizer. *Diagnostics*
12. Chauvin Y, Rumelhart DE (1995) Backpropagation: theory, architectures, and applications. *Psychology Press*
13. Ciregan D, Meier U, Schmidhuber J (2012) Multi-column deep neural networks for image classification. In *2012 IEEE conference on computer vision and pattern recognition* pp 3642–3649
14. Das S, Roy SD, Malakar S, Velásquez JD, Sarkar R (2021) Bi-level prediction model for screening covid-19 patients using chest x-ray images. *Big Data Research* 25:100233
15. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conf Comput Vis Pattern Recognit* pp 248–255
16. Dietterich TG (1998) Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput* 10(7):1895–1923
17. Dietterich TG (2000) Ensemble methods in machine learning. In *International workshop on multiple classifier systems*. Springer pp 1–15
18. 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* p 200432
19. Fukushima K (1988) Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Netw* 1(2):119–130
20. Fukushima K, Miyake S (1982) Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*. Springer pp 267–285
21. Garain A, Basu A, Giampaolo F, Velasquez JD, Sarkar R (2021) Detection of covid-19 from ct scan images: A spiking neural network-based approach. *Neural Comput Appl* pp 1–14
22. Ghahramani Z, Jordan MI (1994) Supervised learning from incomplete data via an em approach. In *Adv Neural Inf Proces Syst* pp 120–127
23. Gozes O, Frid-Adar M, Greenspan H, Browning PD, Zhang H, Ji W, Bernheim A, Siegel E (2020) Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. arXiv preprint arXiv:2003.05037

24. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In Proc IEEE Conf Comput Vis Pattern Recognit pp 770–778
25. Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. *Stat Sci* pp 382–401
26. Huang G, Liu Z, Pleiss G, Van Der Maaten L, Weinberger K (2019) Convolutional networks with dense connectivity. *IEEE Trans Pattern Anal Mach Intell*
27. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In Proc IEEE Conf Comput Vis Pattern Recognit pp 4700–4708
28. Jaiswal A, Gianchandani N, Singh D, Kumar V, Kaur M (2020) Classification of the covid-19 infected patients using densenet201 based deep transfer learning. *J Biomol Struct Dyn* pp 1–8
29. Karbhari Y, Basu A, Geem Z-W, Han G-T, Sarkar R (2021) Generation of synthetic chest x-ray images and detection of covid-19: a deep learning based approach. *Diagnostics* 11(5):895
30. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
31. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q et al (2020) Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology*
32. Li Z, Yi Y, Luo X, Xiong N, Liu Y, Li S, Sun R, Wang Y, Hu B, Chen W et al (2020) Development and clinical application of a rapid igm-igg combined antibody test for sars-cov-2 infection diagnosis. *J Med Virol* 92(9):1518–1524
33. Mannor S, Peleg D, Rubinstein R (2005) The cross entropy method for classification. In Proceedings of the 22nd international conference on Machine learning pp 561–568
34. Masood A, Sheng B, Li P, Hou X, Wei X, Qin J, Feng D (2018) Computer-assisted decision support system in pulmonary cancer detection and stage classification on ct images. *J Biomed Inform* 79:117–128
35. Ortiz-Ospina E, Roser M, Ritchie H, Hasell J (2020) Coronavirus pandemic (covid-19). Our World in Data. <https://ourworldindata.org/coronavirus>
36. Opitz D, Maclin R (1999) Popular ensemble methods: An empirical study. *J Artif Intell Res* 11:169–198
37. Panwar H, Gupta P, Siddiqui MK, Morales-Menendez R, Bhardwaj P, Singh V (2020) A deep learning and grad-cam based color visualization approach for fast detection of covid-19 cases using chest x-ray and ct-scan images. *Chaos, Solitons Fractals* p 110190
38. Pham H, Olafsson S (2019) Bagged ensembles with tunable parameters. *Comput Intell* 35(1):184–203
39. Polikar R (2006) Ensemble based systems in decision making. *IEEE Circuits Syst Mag* 6(3):21–45
40. Rokach L (2010) Ensemble-based classifiers. *Artif Intell Rev* 33(1–2):1–39
41. Schapire RE (1990) The strength of weak learnability. *Mach Learn* 5(2):197–227
42. Sen S, Saha S, Chatterjee S, Mirjalili S, Sarkar R (2021) A bi-stage feature selection approach for covid-19 prediction using chest ct images. *Appl Intell* pp 1–16
43. Silva P, Luz E, Silva G, Moreira G, Silva R, Lucio D, Menotti D (2020) Covid-19 detection in ct images with deep learning: A voting-based scheme and cross-datasets analysis. *Informatics in Medicine Unlocked* 20
44. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*
45. Soares E, Angelov P, Biaso S, Froes MH, Abe DK (2020) Sars-cov-2 ct-scan dataset: A large dataset of real patients ct scans for sars-cov-2 identification. *medRxiv*
46. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In Proc IEEE Conf Comput Vis Pattern Recognit pp 1–9
47. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In Proc IEEE Conf Comput Vis Pattern Recognit pp 2818–2826
48. Tahamtan A, Ardebili A (2020) Real-time rt-pcr in covid-19 detection: issues affecting the results. *Expert Rev Mol Diagn* 20(5):453–454
49. Waibel A, Hanazawa T, Hinton G, Shikano K, Lang KJ (1989) Phoneme recognition using time-delay neural networks. *IEEE Trans Acoust Speech Signal Process* 37(3):328–339
50. Wang L, Wong A (2020) Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *arXiv preprint arXiv:2003.09871*
51. Wang Z, Liu Q, Dou Q (2020) Contrastive cross-site learning with redesigned net for covid-19 ct classification. *IEEE J Biomed Health Inform* 24(10):2806–2813
52. Wolpert DH (1992) Stacked generalization. *Neural Netw* 5(2):241–259
53. Yamaguchi K, Sakamoto K, Akabane T, Fujimoto Y (1990) A neural network for speaker-independent isolated word recognition. In First International Conference on Spoken Language Processing
54. Yang Y, Yang M, Shen C, Wang F, Yuan J, Li J, Zhang M, Wang Z, Xing L, Wei J et al (2020) Laboratory diagnosis and monitoring the viral shedding of 2019-ncov infections. *MedRxiv*
55. YAP JCH, ANG IYH, TAN SHX, Jacinta I, Pei C, LEWIS RF, Qian Y, YAP RKS, NG BXY, TAN HY (2020) Covid-19 science report: diagnostics. *NUS Libraries*

56. Zhang J, Chu Y, Zhao N (2020) Supervised framework for covid-19 classification and lesion localization from chest ct. *The Ethiopian Journal of Health Development (EJHD)* 34(4)
57. Zhang J, Xie Y, Li Y, Shen C, Xia Y (2020) Covid-19 screening on chest x-ray images using deep learning based anomaly detection. *arXiv preprint arXiv:2003.12338*
58. Zhang W, Itoh K, Tanida J, Ichioka Y (1990) Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Appl Opt* 29(32):4790–4797

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

## Authors and Affiliations

Rohit Kundu<sup>1</sup>  · Pawan Kumar Singh<sup>2</sup>  · Massimiliano Ferrara<sup>3,4</sup>  ·  
Ali Ahmadian<sup>5,6,7</sup>  · Ram Sarkar<sup>8</sup> 

<sup>1</sup> Department of Electrical Engineering, Jadavpur University, Kolkata 700032, India

<sup>2</sup> Department of Information Technology, Jadavpur University, Kolkata 700106, India

<sup>3</sup> Department of Law, Economics and Human Sciences & Decisions Lab, Mediterranean University of Reggio Calabria, Reggio Calabria 89125, Italy

<sup>4</sup> ICRIOS - The Invernizzi Centre for Research in Innovation, Organization, Strategy and Entrepreneurship, Bocconi University - Department of Management and Technology, Via Sarfatti 25, Milano 20136, MI, Italy

<sup>5</sup> Institute of IR 4.0, The National University of Malaysia, Bangi 43600 UKM, Selangor, Malaysia

<sup>6</sup> Department of Mathematics, Near East University, Nicosia TRNC, Mersin 10, Turkey

<sup>7</sup> Institute for Mathematical Research, Universiti Putra Malaysia, Seri Kembangan, Selangor 43400 UPM, Malaysia

<sup>8</sup> Department of Computer Science & Engineering, Jadavpur University, Kolkata 700032, India