Choquet fuzzy integral-based classifier ensemble technique for COVID-19 detection[☆]

Subhrajit Dey [a], Rajdeep Bhattacharya [b], Samir Malakar [c], Seyedali Mirjalili [d,e,f,*], Ram Sarkar [b]

[a] *Department of Electrical Engineering, Jadavpur University, Kolkata, 700032, India*
[b] *Department of Computer Science and Engineering, Jadavpur University, Kolkata, 700032, India*
[c] *Department of Computer Science, Asutosh College, Kolkata, 700026, India*
[d] *Centre for Artificial Intelligence Research and Optimisation, Torrens University Australia, Fortitude Valley, Brisbane, QLD, 4006, Australia*
[e] *Yonsei Frontier Lab, Yonsei University, Seoul, South Korea*
[f] *King Abdulaziz University, Jeddah, Saudi Arabia*

ARTICLE INFO

ABSTRACT

The COVID-19 outbreak has resulted in a global pandemic and led to more than a million deaths to date. COVID-19 early detection is essential for its mitigation by controlling its spread from infected patients in communities through quarantine. Although vaccination has started, it will take time to reach everyone, especially in developing nations, and computer scientists are striving to come up with competent methods using image analysis. In this work, a classifier ensemble technique is proposed, utilizing Choquet fuzzy integral, wherein convolutional neural network (CNN) based models are used as base classifiers. It classifies chest X-ray images from patients with common Pneumonia, confirmed COVID-19, and healthy lungs. Since there are few samples of COVID-19 cases for training on a standard CNN model from scratch, we use the transfer learning scheme to train the base classifiers, which are InceptionV3, DenseNet121, and VGG19. We utilize the pre-trained CNN models to extract features and classify the chest X-ray images using two dense layers and one softmax layer. After that, we combine the prediction scores of the data from individual models using Choquet fuzzy integral to get the final predicted labels, which is more accurate than the prediction by the individual models. To determine the fuzzy-membership values of each classifier for the application of Choquet fuzzy integral, we use the validation accuracy of each classifier. The proposed method is evaluated on chest X-ray images in publicly available repositories (IEEE and Kaggle datasets). It provides 99.00%, 99.00%, 99.00%, and 99.02% average recall, precision, F-score, and accuracy, respectively. We have also evaluated the performance of the proposed model on an inter-dataset experimental setup, where chest X-ray images from another dataset (CMSC-678-ML-Project GitHub dataset) are fed to our trained model and we have achieved 99.05% test accuracy on this dataset. The results are better than commonly used classifier ensemble methods as well as many state-of-the-art methods.

1. Introduction

COVID-19 or SARS-CoV-2 is a severe acute respiratory syndrome (SARS), which has resulted in a global pandemic and led to more than a million deaths.[1] The common symptoms of this virus are related to respiratory problems, fever, loss of smell and taste, cough, and breathing problems. However, it can also lead to fatal outcomes, where the infection can cause Pneumonia, severe respiratory problems, kidney failure, and even death.[2]

The COVID-19's primary problem is that, unlike other viruses, it has a very long incubation period (the time taken between the exposure to the virus and symptoms emerging), due to which a victim starts

---

spreading the infection unknowingly. The COVID-19 incubation period usually ranges from 1 to 12.5 days and may even extend to 14 days. It has also been noticed that some patients are carrying the COVID-19 without any symptoms, and these types of patients are known as asymptomatic patients.[3] Although vaccination has started in many countries, the problem is that the roll-out is slow due to storage and manufacturing issues. Also, because of new COVID-19 variants, it is still ideal to detect the virus as early as possible to reduce the risk of community transmission. Hence, an efficient system for the said virus detection is a necessity in the current situation.

Many research attempts in the fields of machine learning and image processing have been carried out to predict whether a person is carrying the virus by analysing the X-ray or computerized tomography (CT) chest scan images [1,2]. However, in this work, we have only used chest X-ray images because the whole purpose of developing a model for COVID-19 detection is to make the testing process inexpensive and, also, because getting chest X-ray images is cheaper than obtaining CT scan images. Moreover, CT scan images are computationally expensive, as they come in 3D format. Most importantly, the current health guidelines recommend the use of chest X-Ray over CT scan images.

In most of these research attempts [3–8], convolutional neural networks (CNNs) are preferred for classification tasks as they have been proven to produce appreciable performances while dealing with image classification problems, as can be seen in case of LeNet-5 [9] and AlexNet [10]. There are several improved CNN models in the literature, such as ResNet50 [11], InceptionNet [12], DenseNet [13], etc. These models are capable of capturing better perceptual features than hand-engineered ones, which helps with better image classification.

In the case of our work, the dataset is smaller, so it is important to utilize transfer learning to achieve competent results for our dataset classification. This is essential for an accurate classification of X-ray images for detecting COVID-19 at the early stage. As discussed by Das et al. [14], rather than using individual CNN models for chest X-ray image classification, we can assemble their outputs to get better results. It has been proven that ensembling of deep learning models can improve classification performance to a great extent. From the works by Liu et al. [15], and Kwak and Pedrycz [16], we note that, rather than using ensemble learning that uses normal weighted average, we can use Choquet fuzzy integral method to combine the outputs of different CNN models. According to Banerjee et al. [17], we can achieve classification accuracy better than the classification accuracy of individual models by using Choquet fuzzy integral to combine CNN models. Also, from previous research on COVID-19 detection, we can see that various models give accurate results in different situations, which is why such solutions cannot be generalized, posing a significant drawback. To solve this problem, we have introduced Choquet fuzzy integral to achieve COVID-19 classification accuracy higher than that of individual CNN models. Our dataset contains chest X-ray images for classifying patients into COVID-19 affected, Pneumonia affected, and Normal cases. For transfer learning, we have used three pre-trained models: DenseNet121 [13], InceptionV3 [12], and VGG19 [18]. Subsequently, we combine the three models using Choquet fuzzy integral method, as done by Banerjee et al. [17], that overrides the individual methods' performances. We use the validation accuracies of individual classifiers as the fuzzy-membership values for each classifier.

The rest of this paper is organized in the following way:

- In Section 2 (Related Work), we discuss the current state of research in the field of COVID-19 detection and how it motivated us to arrive at our solution.
- In Section 3 (Proposed Method), we propose our model and describe its key elements.

- In Section 4, we describe the dataset, the hyperparameters that we have used, and the results of our experimentation with the proposed network.
- In Section 5, we compare the results produced by our model on the dataset with other state-of-the-art models.
- Finally, in Section 6, the paper is concluded.

2. Related Work

The massive outbreak of COVID-19 has drawn the attention of many researchers in areas such as image processing and artificial intelligence. As a result, we can see many research attempts suggesting COVID-19 detection through machine learning. Majid et al. [3] trained a CNN-based model from scratch to extract deep features and then used machine learning algorithms like K-Nearest Neighbors algorithm (KNN) and Support-Vector Machine (SVM) for classification of COVID-19 chest X-ray images. Here, as the authors trained the model from scratch, they had to rely on the small COVID-19 chest X-ray image dataset that was available to them. However, in the given situation, they could have used pre-trained models which are already trained on large datasets like the ImageNet to achieve better classification accuracy.

Abbas et al. [4] used their previously developed CNN model, called decompose, transfer and compose (DeTrac) method, along with transfer learning for classification of chest X-ray images of COVID-19. In this case, DeTrac method makes the model computationally heavier. Moreover, it was trained and tested on a smaller dataset, which is why training and testing the method on a large dataset combined with the computationally heavy model of DeTrac becomes a problem for the model. A deep learning network, called COVNet [5], was used to extract features from volumetric chest CT scan images for COVID-19 detection. The problem in this work is that the training and testing data came from the same hospital and we do not know how well the model can be generalized on the data from other sources. Also, the authors use CT scan dataset for COVID-19 detection, which is a problem, as the CT scan images are computationally expensive and also more expensive than X-ray images.

In a similar work [6], the authors employed a swarm-based feature selection method, known as Marine Predators algorithm, to select the most important features extracted using a pre-trained Inception model. The problem with this model, as noted by the authors themselves, is that it is built with python, whereas the (FO-MPA) algorithm where feature selection is performed is built with MATLAB. Hence, the whole model is dependent on two different environments that might lead to file sharing and saving issues. Also, the feature extraction from images could have been improved by using multiple pre-trained CNN models rather than just one. Apostolopoulos et al. [19] made use of transfer learning with CNN models for automatic COVID-19 detection. The problem with this research is that the authors used individual pre-trained models for COVID-19 detection, but we know that some models can classify images better than other models in certain situations and so ensembling them would have boosted their performance by a huge margin.

In another work, Turkoglu [20] used a pre-trained AlexNet model for feature extraction and concatenation, followed by feature selection. These features are then used for COVID-19 detection using SVM classifier. Oh et al. [21] first used an extended DenseNet103 model for semantic segmentation of chest X-ray images, and then ResNet 18 model was used for COVID-19 detection from these chest X-ray images. In the work by Hoon et al. [22], a simple 2D deep learning model was developed with the help of a transfer learning approach as the backbone of the network, and the method was named Fast track COVID-19 classification network (FCONet). Another CNN model, named CovXNet, was proposed by Mahmud et al. [23]. This network uses depth-wise convolution with varied dilation rates for efficient extraction of diversified features from COVID-19 chest X-ray images. A new CNN framework with cascaded deep-learning classifiers was proposed by Karar et al. [24]. It enhances the performance of a CAD (Computer Aided Diagnosis) system for highly

---

[3] https://swachhindia.ndtv.com/coronavirus-explainer-what-we-know-about-asymptomatic-covid-19-infections-so-far-49628/.

suspected COVID-19 and Pneumonia diseases in X-ray images. Panwar et al. [25] proposed a deep-transfer learning algorithm with the aim of accelerating the COVID-19 detection while using chest X-ray and CT scan images. Rajaraman et al. [26] used an iteratively pruned deep-learning ensembles for detecting pulmonary manifestation of COVID-19, with the help of chest X-rays. They use a customized CNN model and a selection of ImageNet pre-trained models for training and evaluation at the patient level on publicly available chest X-Ray image collections. In another work [27], we see the use of an Auxiliary Classifier Generative Adversarial Network (ACGAN) based model, referred to as CovidGAN, for the generation of synthetic chest X-ray images, similar to the case in Ref. [28].

From the above-mentioned research attempts, it can be concluded that most of the researchers use transfer learning for detection of COVID-19 cases from chest X-ray or CT scan images. The reason behind such a frequent use of transfer learning is that the dataset size of COVID-19 patients' chest X-ray or CT scan images is usually not large enough to train a CNN model from the scratch whilst simultaneously obtaining satisfactory performance. As a result, researchers use pre-trained CNN models trained on ImageNet as feature extractors, and then, in many cases, the extracted features are analyzed prior to their use in COVID-19 detection. As an alternative approach, for achieving satisfactory classification performance, researchers either generate synthetic data (e.g., Ref. [27]) or introduce completely unique and novel deep-learning architectures (e.g., Refs. [22,23]).

Certainly, both of these methods are popularly used schemes in image processing and computer vision problems, with data scarcity as a challenge. However, in both of these scenarios, the computational cost is much higher than required when transfer learning is used. Additionally, the synthetic data are more biased towards the source data (i.e., data that are mimicked in synthetic data) and, hence, do not contain outliers, which play a crucial role in improving the current model. Also, mimicking the real data to produce synthetic data is not an easy task and a lack of approved synthetic data generation makes the process more critical. When generating the new CNN model, we might fail to prove its robustness and reusability in solving similar problems. That is why, after considering all these facts, we have opted for using transfer learning.

We use complex and deep CNN models, which are trained on ImageNet dataset as base models for image classification of chest X-ray and CT scan images. In order to utilize the strengths of all the models, we create an ensemble model, which performs better than all individual ones. As previously mentioned, we use Choquet fuzzy integral method to combine the results of individual CNN models and produce an ensemble CNN model that can effectively classify chest X-ray images. The individual models are spanning across a variety of architectures and individual strengths so that the final model is well-versed with all types of inputs and can efficiently classify the input image.

In some deep learning related research [14,26,29], which deals with COVID-19 detection, it can be seen that ensembling methods are used for combining classification decisions with multiple classifiers to improve the final classification results. In most cases, combination happens by majority voting or weighted average, but the problem with them is that these techniques are not dynamic. However, in our case, we use Choquet fuzzy integral as we know that it tailors predictions based on confidence scores of each classifier, which varies with each data point and leads to more dynamism even after the fuzzy measures have been set. It is a unique technique of combining classifiers, which we can see in other applications, as mentioned in the works [15–17]. Thus, we use this method as it has been rarely used in the case of COVID-19 detection.

3. Proposed work

In this section, we propose a model which uses the transfer-learning technique to perform the classification of chest X-ray images into Normal subjects, COVID-19-affected subjects, and Pneumonia-affected subjects. This is achieved by the pre-trained models of VGG19,

DenseNet121, and Inception-v3 as feature extractors of our dataset, after which the features are passed through classifiers for classification. The classifiers' outputs are then combined with the help of Choquet fuzzy integral, having fuzzy measures of each classifier with an experimentally set ratio. The model is shown in Fig. 1.

3.1. Feature extraction

It has been seen that CNNs are used in many studies related to COVID-19 detection from chest X-ray or CT scan images. This is due to the incredibly flexible nature of CNN models in image recognition tasks and object detection tasks, compared to other machine learning algorithms. It all started with the initial success of CNN models like LeNet-5 [9] and AlexNet [10] in several image classification problems. This motivated researchers to design more sophisticated networks that are either deeper, like ResNet50 [11] and DenseNet121 [13], or wider, like Inception-v3 [12]. For our ensemble model, we choose a variety of CNN models so that all characteristics of CNNs are represented in our output. To achieve high image classification accuracy, however, deep learning models need a large amount of data and huge processing power. As, in our case, the dataset for COVID-19 classification is small in terms of the number of samples, we need to use transfer learning so that the models are already capable of identifying lower-level features and need to learn only higher-level features from the given training images.

In the case of transfer learning, we utilize the learned weights of some pre-trained CNN models. A pre-trained model is one that has already been trained on large datasets, like ImageNet having 20,000 categories. The learned weights from these pre-trained CNN models are used by freezing some layers of the model while training on our small dataset, which is unique but similar to the dataset on which the models are trained. If we do not use transfer learning and just train the model from scratch on our small dataset, the model cannot learn to obtain more comprehensive features of COVID-19 images required for better classification. Thus, we see that, by using transfer learning, we can eliminate the need for large datasets and large processing power. A brief description of each model is given below, with the help of Fig. 2.

- **Inception-v3:** In InceptionNet, the architecture is wider rather than deeper as seen in Fig. 2. The architecture has multiple filters with varied filter sizes at the same level. The intuition for InceptionNet is that using smart factorization methods, convolutions can be made efficient. For example, 5×5 convolution can be factorized into two 3×3 convolutions and a 5×5 convolution is 2.78 times more complex than a 3×3 convolution. Hence, this leads to a boost in performance. Therefore, we use Inception-v3 in our work, which consists of RMS-Prop optimizers, a factorized 7×7 convolution, BatchNorms in the auxiliary classifiers, and Label Smoothing, which is a regularization technique that prevents the network from becoming too confident in one class, hence preventing it from overfitting.
- **DenseNet121:** In the DenseNet architecture, each layer is connected to every other layer, so, in Fig. 2, we can say that, if the output of all convolutional blocks or maxpool blocks is connected to the input of other blocks, then it is possible to implement the DenseNet architecture. For example, if we have n number of layers then the total number of connections is given by $n(n+1)/2$. In the case of each layer, feature maps of all the previous layers are used as input for the current layer, and its own feature maps are used as inputs for next layers. So, we see that the DenseNet architecture facilitates both the down-sampling and the feature concatenations by dividing the network into multiple densely connected networks, while the feature map size remains the same.
- **VGG19:** VGG19 network as shown in Fig. 2 accepts 224×224 images in RGB format. The input image is passed through a number of convolutional layers where each layer has a receptive field size of 3×3 and stride equal to 1. After that, maxpooling is performed
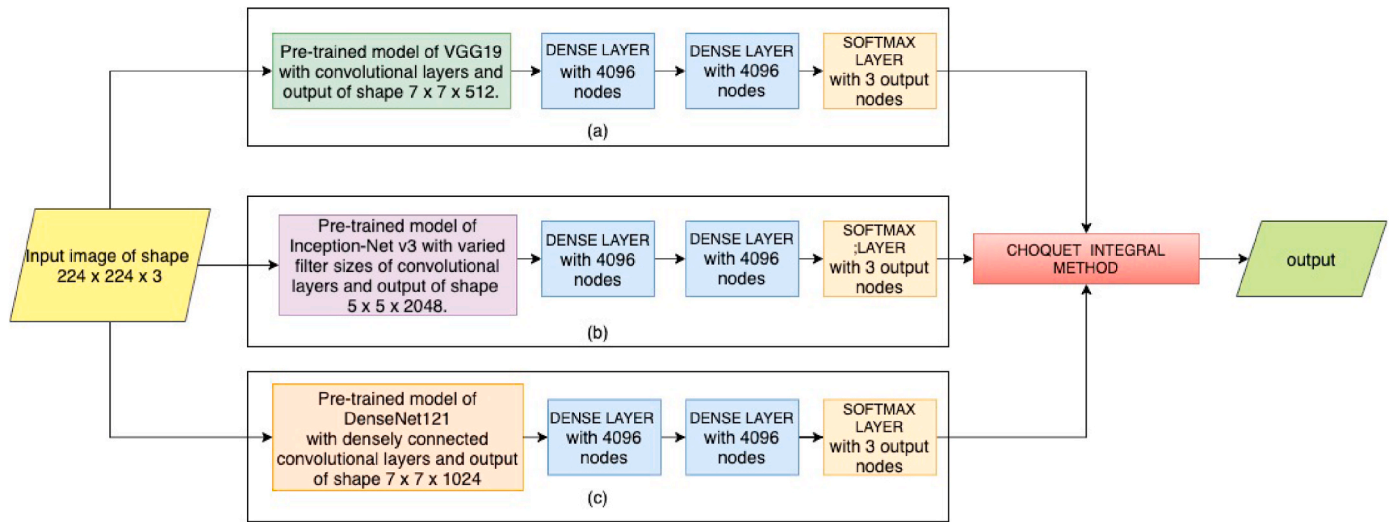
Fig. 1. Our proposed Choquet fuzzy integral based COVID-19 classification model. In the figure, (a) denotes the CNN model, which uses the pre-trained VGG19 model, (b) denotes the CNN model with pre-trained model of Inception-v3, and (c) denotes the model using pre-trained DenseNet121 model.
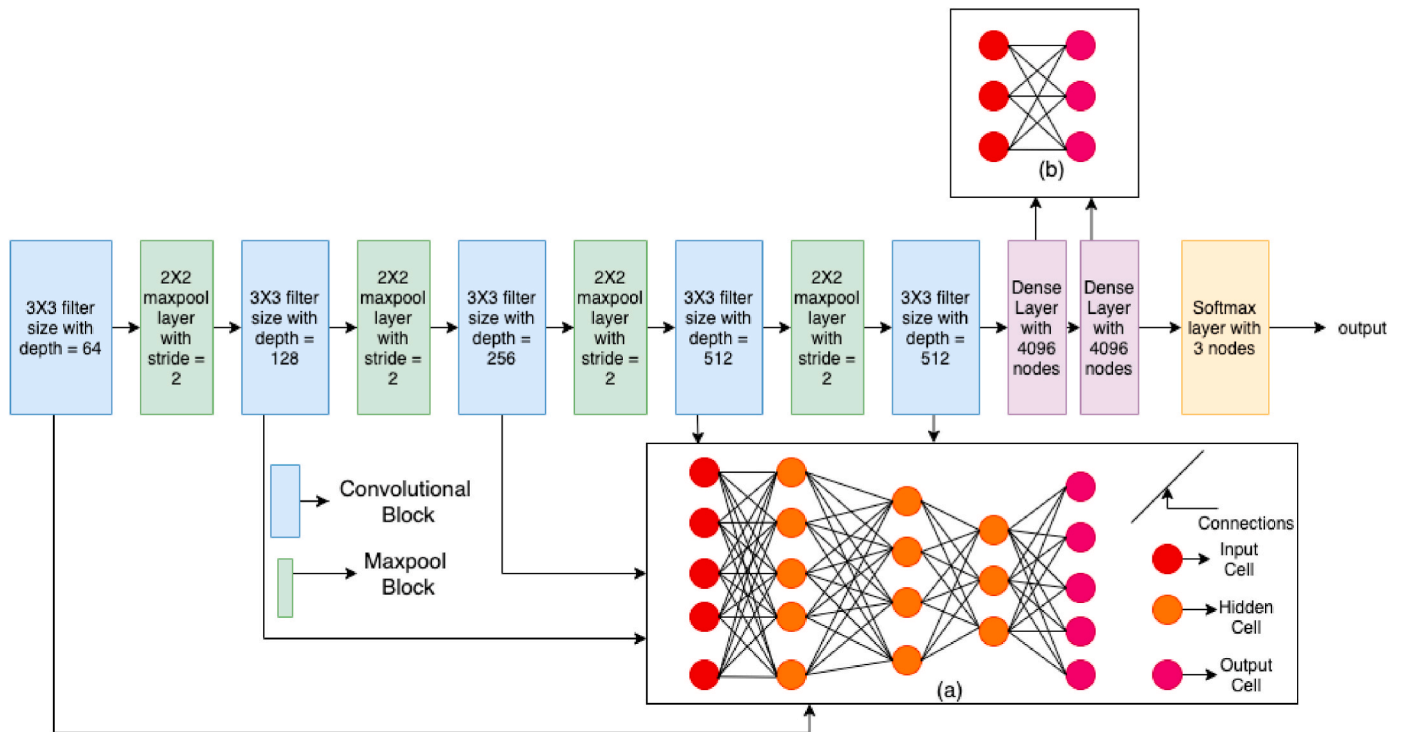

Fig. 2. In this figure we represent a VGG19 model and the illustrations of deep convolutional layer are shown in (a) and fully connected layer in (b).

over a window of 2 × 2 and with stride equal to 2. However, a max-pool layer does not come after every convolutional layer. The hidden layers have Rectified Linear Unit (ReLU) as their activation function. The size of the feature layer that we extract is of dimension 7 × 7 × 512. After this, we have flatten it for classification purposes.

3.2. Choquet fuzzy integral

Aggregation in this work refers to the process of collecting performance scores of each classifier into a single global score. The function that is used to combine the scores into a single global score is known as aggregation operator.

Normally as the aggregation operator, we have weighted average or quasi arithmetic means but in our case it is the Choquet fuzzy integral operator. The Choquet fuzzy integral method has been used previously in many pattern recognition problems [15–17]. The advantage of using it is that it harnesses the degree of uncertainty that is present in the decision scores that we get as additional information during the fusion of classifiers which is absent in normal ensembling methods. So the end result is the generalization of aggregation operators on a set of confidence scores which are known as fuzzy measures. These fuzzy measures are weighted values given to each classifier.

To use Choquet fuzzy integral, we need to identify the fuzzy measure values of each classifier. The fuzzy measure values determine the strength of each classifier and also the strength of all possible classifier

combinations. In the case of Choquet fuzzy integral, there are boundary conditions, which are as follows:

- If, in any combination, all the classifiers are present, that combination has the maximum strength.
- If, in any combination, all the classifiers are absent, that combination has no strength at all.

The fuzzy measure expresses three types of interactions that can take place between the classifiers under consideration. Let us assume that X and Y are two subsets of classifiers and that they are mutually exclusive (i.e., $X \cap Y = \Phi$). We also assume that $f(S)$ indicates the strength of a set of classifiers S in the underlying classification problem. All three types of interactions are as follows:

- $f\{X \cup Y\} = f\{X\} + f\{Y\}$: here the strength of the $X \cup Y$ is equal to the sum of the strengths assigned to X and Y. In this case, it can be said that X and Y are sharing additive effects or are independent of each other.
- $f\{X \cup Y\} \leq f\{X\} + f\{Y\}$: here the strength of $X \cup Y$ is less than or equal to the sum of the strengths assigned to X and Y. So in this case we can say that the X and Y are sharing sub-additive effects or are redundant to each other.
- $f\{X \cup Y\} \geq f\{X\} + f\{Y\}$: here the strength of $X \cup Y$ is greater than or equal to the sum of the strengths assigned to the X and Y. Here, it can be said that X and Y are expressing super-additive nature or have a synergistic effect.

After we get the complete set of fuzzy measure scores and the available performance scores of each classifier, we can apply the Choquet fuzzy integral method to compute the aggregated global score of the classifiers. Let $g()$ calculates fuzzy measure score of a set of classifiers $D = \{d_1, d_2, d_3, ..., d_n\}$ and $A = \{a_1, a_2, a_3, ..., a_n\}$ are the performance scores of the individual classifiers in D. Let us assume that L_i represents a subset such that $L_i = \{d_1, d_2, d_3, ..., d_i\}$ of D where $1 \leq i \leq n$. It means that, if $i = 1$ then $L_1 = \{d_1\}$ and if $i = 3$ then $L_3 = \{d_1, d_2, d_3\}$. We also assume that $a_1 \geq a_2 \geq a_3 ... ... \geq a_n$, so then the aggregated score can be calculated using Choquet fuzzy integral with the help of the following equation,

$$Choquet_g(a_1, a_2, a_3, ..., a_n) = a_n^* g(L_n) + (a_{n-1} - a_n)^* g(L_{n-1}) + \cdots$$
$$+ (a_1 - a_2)^* g(L_1) \quad (1)$$

For the calculation of fuzzy membership values of combinations of classifiers, we first have to calculate the value of λ given by the following equation,

$$1 + \lambda = \Pi_{i=1}^n(g(\{d_i\})\lambda + 1) \quad (2)$$

Once we get the roots of the characteristic equation, we can determine the value of λ. Using this, we can calculate the fuzzy membership value of any classifier combination by repetitive use of the following equation.

$$g(\{d_l, d_o\}) = g(\{d_l\}) + g(\{d_o\}) + \lambda g(\{d_l\})g(\{d_o\}) \quad (3)$$

where $1 \leq l, o \leq n$.

When dealing with classifier combinations, we have p classifiers which are used for classifying q classes. Let, x_{ij} represent the confidence score of jth class of ith classifier. Now, for each j, Choquet integral is used to obtain fuzzy confidence score. In this case, for each $j = 1, 2, ..., q$, $b_i = x_{ij}$ where $i = 1, 2, ..., p$, the fuzzy membership values for each classifier (i.e., $g(b_i)$) are experimentally set. However, in this work, we set the fuzzy measure $g(\{b_i\}) = \frac{w_i}{(w_1 + w_2 + w_3 .... w_p)}$, where w_i represents the validation accuracy of the ith classifier. The rest of the fuzzy measure values for the classifier combination are calculated using Eq. (3), after determining the value of λ from Eq. (2) The ensembling of classifiers using Choquet integral has been shown in Fig. 3.

The advantage of Choquet fuzzy integral is that even after setting the values of fuzzy measures, it can alter the weightage of each classifier based on the decision scores provided by other classifiers. This makes the system dynamic, which is different from other ensembling methods that use majority voting system or weighted average, where setting some parametric values makes the system completely static.

The complexity of Choquet fuzzy integral is given by $O(K \times (n \times log(n)))$, where K denotes the number of classes and n denotes the number of classifiers.

The main steps of the proposed method are shown in Algorithm 1.

Algorithm 1.

Algorithm 1: Key steps of the Proposed method

1 Images for training I
2   Images for testing S Normal or COVID-19 or Pneumonia cases.
  `/* Model training phase and fuzzy measure value calculation`   `*/`
3 Resize training images I to $224 \times 224 \times 3$.
4 Extract feature maps from I using pre-trained models of VGG19, InceptionV3 and DenseNet121 (as described in Section 3.1). Let the feature maps generated by VGG19, InceptionV3 and DenseNet121 be $FS1$, $FS2$ and $FS3$ respectively.
5 Build classifiers $C1$, $C2$ and $C3$ using two FC layers of 4096 nodes each on the feature maps and a softmax layer of 3×1 on the top of $FS1$, $FS2$ and $FS3$ respectively.
6 Split the images I into training and validation sets in the ratio of 7:3
7 Train the classifiers $C1$, $C2$ and $C3$ using the training set data.
8 Use the validation set to generate validation accuracies w_1, w_2 and w_3 using $FS1$, $FS2$ and $FS3$ features respectively.
9 Calculate the fuzzy measure values using w_1, w_2 and w_3. The fuzzy measure values F for the ith classifier is $w_i/(w_1 + w_2 + w_3)$.
  `/* Model testing phase.`   `*/`
10 Resize testing images S to $224 \times 224 \times 3$.
11 Use the trained classifiers $C1$, $C2$ and $C3$ to generate prediction scores P of size 3×1 based on the testing samples.
12 Apply Choquet integral to the prediction scores P using fuzzy measures F (see Eq 1) to calculate the final prediction scores y.
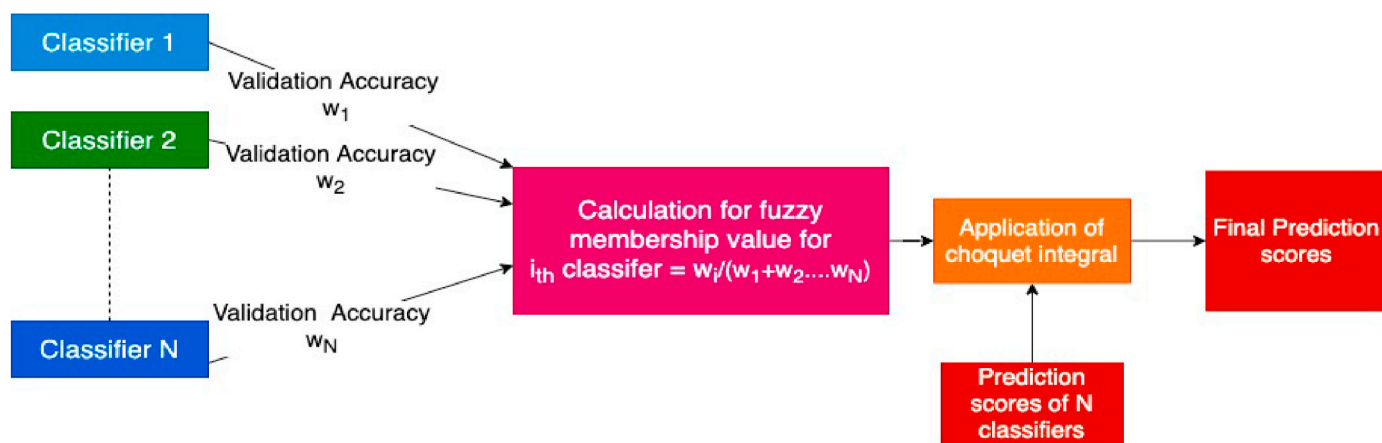13 Use prediction scores y to determine whether a given sample is a Normal or COVID-19 or a Pneumonia case.

Fig. 3. Classifier ensembling using Choquet fuzzy integral method. Here the fuzzy membership values are calculated as $g(\{c_i\}) = \frac{w_i}{(w_1 + w_2 + w_3 \ldots w_N)}$ where $w_1, w_2 \ldots w_N$ are validation accuracies of classifiers $c_1, c_2 \ldots c_N$ respectively.

Table 1
Dataset distribution for training and evaluation of the proposed model.

Dataset Name	#Train samples			#Test samples		
	Covid-19	Normal	Pneumonia	Covid	Normal	Pneumonia
Kaggle Pneumonia dataset + IEEE covid dataset	739	1072	3100	185	269	775

4. Experimental results and discussion

4.1. Dataset description

Dataset is an integral part of any research, as it helps evaluate newly designed model. To train and evaluate our work, we have combined the chest X-ray images of COVID-19 patients, Pneumonia, and Normal subjects from two public datasets: covid-chest xray-dataset[4] and chest-xray-Pneumonia.[5] The first dataset was made to create a repository of chest X-ray and CT scan images of COVID-19 patients. Here, we have only considered the chest X-ray images. The second dataset contains the chest X-ray images of patients suffering from Pneumonia and the chest X-ray images of Normal patients taken from Guangzhou Medical Center. All low quality and unreadable images are removed in the initial screening. The distribution of train and test sets prepared for our experiments is shown in Table 1. Some sample images of all three categories of chest X-ray images are shown in Fig. 4. In order to get the validation accuracies of the individual classifiers we use 30% of the training samples as validation samples and we train our model on the remaining 70% samples of the training set, apart from this customized dataset, we have also used a new dataset (CMSC-678-ML-Project GitHub[6]) to check the performance of our learned model on this new and unseen dataset.

4.2. Hyperparameters used

We train all three CNN models on the chest X-ray dataset with the same set of hyperparameters. During the input process, we resize the chest X-ray images to a size of 224×224. For training, we use 60 epochs and a learning rate of 2×10^{-4}, both of which are small enough to avoid overfitting of the CNN models in use. For compilation, we consider RMSProp as our optimizer and after extraction of features by the pre-trained models, we have used two Dense layers with 4096 neurons each, as part of the classifier with ReLU as the activation

function. The last layer is a Softmax layer with three output nodes. Since the output layer has three classes, we have used categorical-cross entropy as the loss function for our model training. In Fig. 5 the accuracy vs epoch, and loss vs epoch curves are depicted for the three base classifiers during the training phase.

4.3. Results

In the present scope of work, we have used three pre-trained models: VGG19, DenseNet121 and Inception-v3 and trained them on the chest X-ray dataset whose results are given in Tables 2–4, respectively. From the results recorded in these tables, it is evident that VGG19 gives the best average individual test classification accuracy among all CNN models in use, which is 98.20%. However, in terms of average precision score, the VGG16 and MobileNetV2 models perform best with a score of 98.00%. In terms of the average recall, both the VGG16 and MobileNetV2, with the highest value of 97.33% and, in the case of average F1-score, both VGG16 and MobileNetV2 give the best F1-score of 97.67%.

After getting the outputs from the mentioned CNN models, we combine the outputs with Choquet fuzzy integral based ensemble technique, and the new classification scores are shown in Table 5. Inspecting the results in this table, it is evident that, with the help of Choquet fuzzy integral method based classifier ensemble, we are able to improve the average test accuracy to 99.02%, which is better than the 98.20% obtained using VGG19 model. However, through the method of ensembling classifiers, our model increases the previous best average precision score by 1.00% and the previous best average recall and F1-score values by 1.67% and 1.33%, respectively.

4.4. Error case analysis

After performing experiments with our proposed ensemble model and comparing them to other models, we have observed that our method outperforms many known CNN models and achieves very high performance accuracy in COVID-19 detection. However, with its classification accuracy of 99.02%, there are cases where it fails to predict the correct class of the chest X-ray image, which is dangerous due to the disease's fatal nature. Some examples of chest X-ray images where it fails to properly classify the images are shown in Fig. 6.
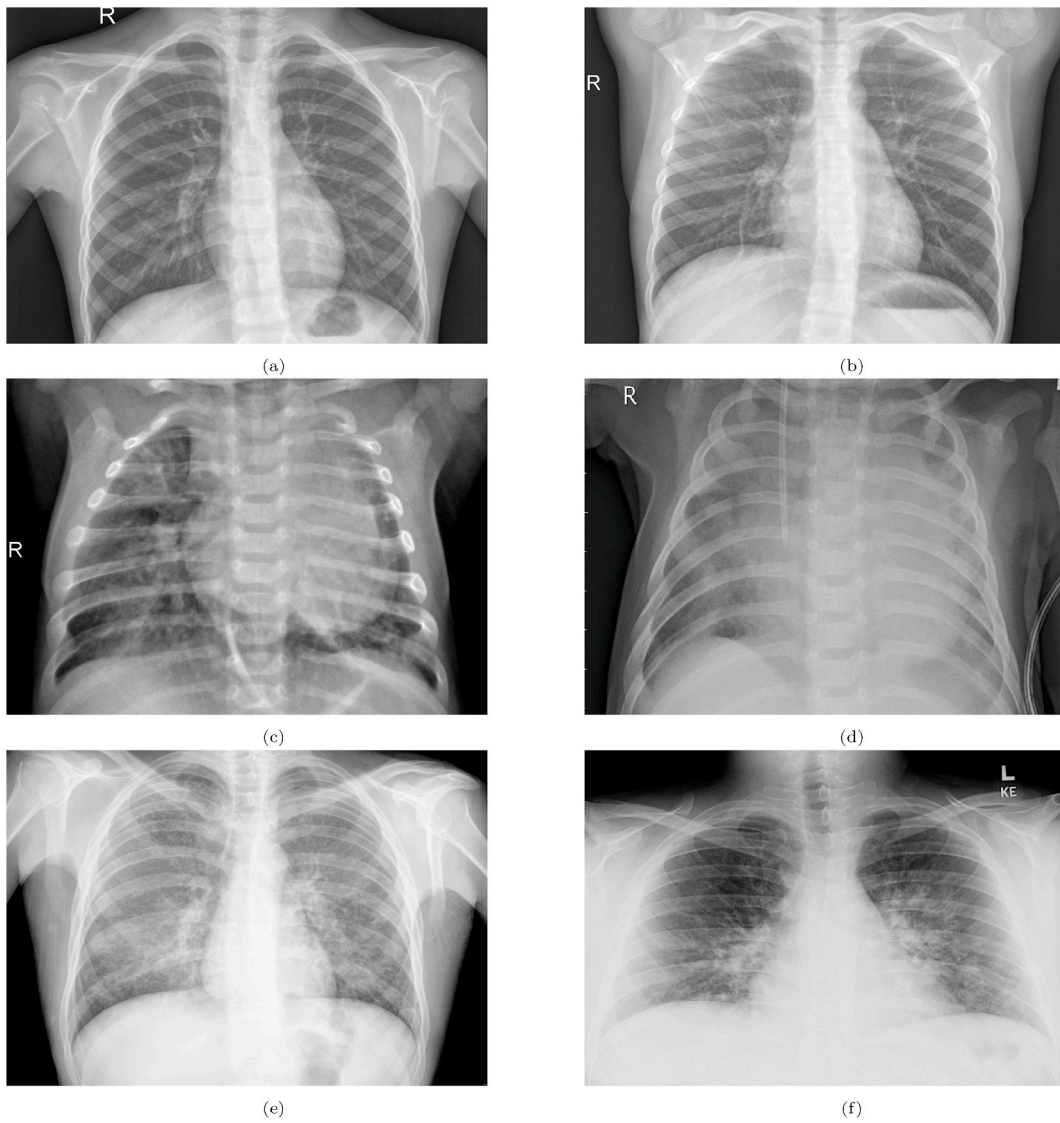

Fig. 4. Sample chest X-ray images of normal people (a–b), Pneumonia patients (c–d), and COVID-19 cases (e–f).
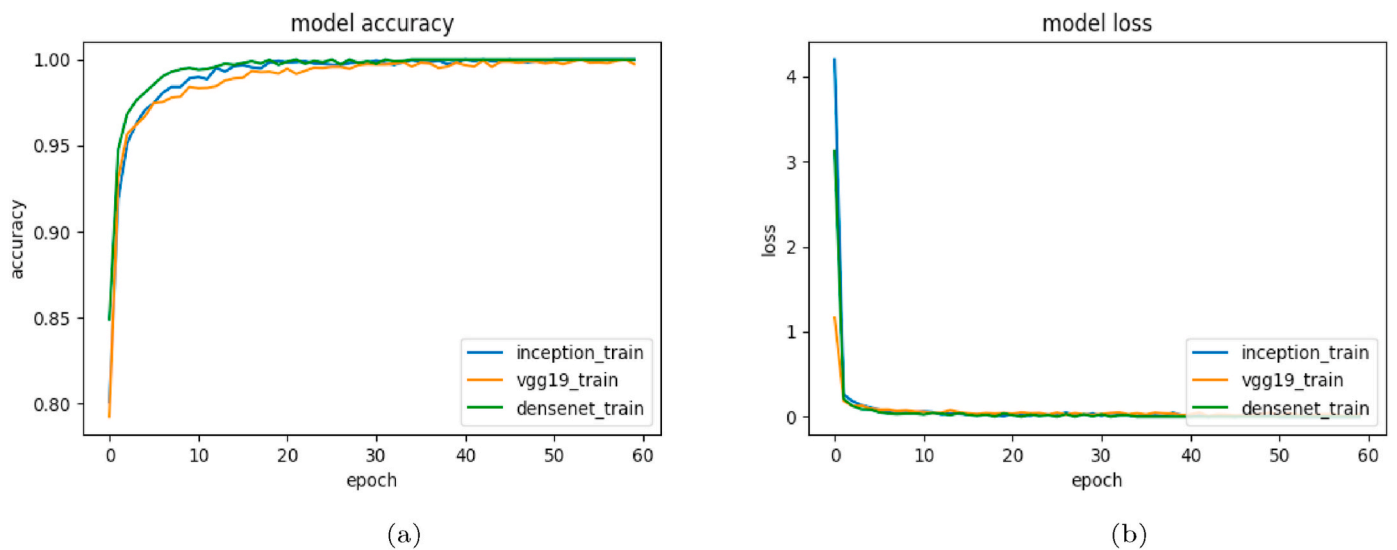

Fig. 5. Illustration of results: a) represents the accuracy of VGG19, DenseNet121 and Inception-v3 models on the training dataset while training, with respect to the number of epochs, whereas (b) represents the loss with respect to the number of epochs while training.

Table 2

The performances of VGG19 for classification of Normal, COVID-19, and Pneumonia cases.

Image Type	Precision (in %)	Recall (in %)	F1-score (in %)	Accuracy (in %)
Normal	94.00	94.00	94.00	98.20
Pneumonia	98.00	98.00	98.00	
COVID-19	98.00	97.00	98.00	
Average	96.67	96.33	96.67	

Table 3

The performances of DenseNet121 for classification of Normal, COVID-19, and Pneumonia cases.

Image Type	Precision (in %)	Recall (in %)	F1-score (in %)	Accuracy (in %)
Normal	95.00	99.00	97.00	97.96
Pneumonia	99.00	98.00	98.00	
COVID-19	99.00	96.00	98.00	
Average	97.67	96.67	96.67	

Table 4

The performances of Inception-v3 for classification of Normal, COVID-19, and Pneumonia cases.

Image Type	Precision (in %)	Recall (in %)	F1-score (in %)	Accuracy (in %)
Normal	93.00	95.00	94.00	96.69
Pneumonia	98.00	98.00	98.00	
COVID-19	98.00	97.00	98.00	
Average	96.33	96.67	96.67	

Table 5

The performances of proposed Choquet fuzzy integral based classifier ensemble technique for classification of Normal, COVID-19, and Pneumonia cases using the pre-trained models of VGG19, Inception-v3, and DenseNet121.

Image Type	Precision (in %)	Recall (in %)	F1-score (in %)	Accuracy (in %)
Normal	97.00	99.00	98.00	99.02
Pneumonia	100.00	99.00	99.00	
COVID-19	100.00	99.00	100.00	
Average	99.00	99.00	99.00	

4.5. Performance on an inter-dataset experimental set-up

The authors of the sate-of-the-art methods in the literature, in general, have used the samples from same dataset to train and evaluate their model's performance. In order to make a system that works well on

variety of data samples, samples from these datasets are first merged and then randomly divided into train test parts. To the best of our knowledge, no study for COVID-19 screening from chest X-ray images has evaluated the performance of their model in inter-dataset set-up. In other words, it is not explored how these models behave on new samples

Table 6

Performances of proposed Choquet fuzzy integral method on the CMSC-678-ML-Project GitHub dataset.

| Image Type	Precision (in %)	Recall (in %)	F1-score (in %)	Accuracy (in %)
Normal	100.00	96.00	98.00	99.05
Pneumonia	98.00	100.00	99.00	
COVID-19	100.00	100.00	100.00	
Average	99.00	99.00	99.00	

Table 7

Depicts the performance comparison of present ensemble methods with some of the standard ensembling techniques.

Models	Average			
	Precision (in %)	Recall (in %)	F1-score (in %)	Accuracy (in %)
Average Voting	95.33	97.33	96.00	96.50
Weighted Average	96.67	98.33	97.33	97.47
Majority Voting	98.00	98.67	98.34	98.62
Choquet fuzzy integral based	99.00	99.00	99.00	99.02

Fig. 7. Confusion matrix of our proposed Choquet fuzzy integral method on the CMSC-678-ML-Project GitHub dataset.
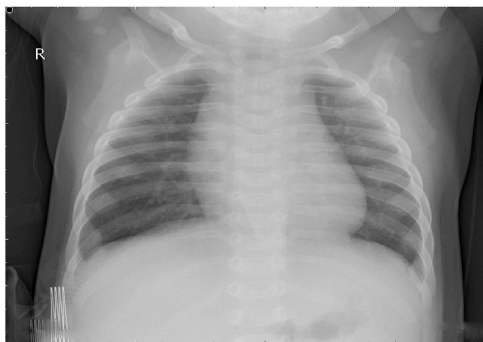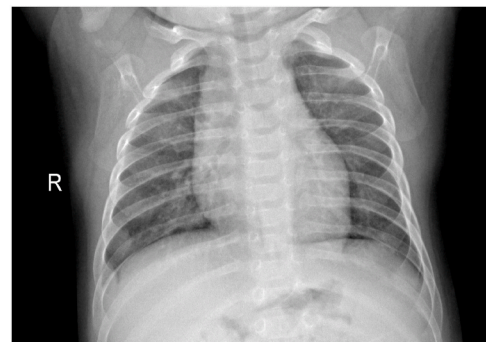
Fig. 6. Examples of erroneously classified images: (a) Normal Subject's x-ray image is classified as COVID-19 case and (b) Normal Subject's x-ray image is classified as Pneumonia case.

from different sources or capturing mediums. To this end, to evaluate the performance of proposed model for the mentioned scenario, we have fed samples from a new dataset into our model, as described in the previous experiments. The new dataset is CMSC-678-ML-Project GitHub, which is divided in two forms: (i) 3 class dataset containing chest x-ray images of COVID-19 patients, normal subjects, and Bacterial Pneumonia infected patients, and (ii) 4 class dataset chest X-ray images of Normal subjects, COVID-19 infected patients and Pneumonia patients either having Viral Pneumonia or Bacterial Pneumonia. Since the present model does not consider chest X-ray images of Pneumonia patients separately, we prefer to use the second dataset to test the performance of our model. During testing, we have considered chest X-ray images of Viral and Bacterial Pneumonia as common Pneumonia patients' chest x-ray images. All the samples of this dataset are fed to our previously trained system and it achieves an accuracy of 99.05%, as shown in Table 6, which is very close to our model's performance (see Table 7) and comparable with the state-of-the-art results, as mentioned in Ref. [30], where training and testing was performed on the samples of only this dataset. The confusion matrix is shown in Fig. 7, which is showing that our model fails to properly classify only 3 Normal chest X-ray images.

4.6. Comparative performance analysis

It has already been indicated that we have used Choquet fuzzy integral based classifier ensemble method to improve the classification accuracy achieved by individual classifiers and, thereby, to improve the prediction of COVID-19 cases. Therefore, in this section, we have compared our method with other standard classifier ensemble

techniques and state-of-the-art deep learning methods for COVID-19 detection. In this context, we would like to mention that the datasets used in state-of-the-art methods are different from the present ones, either in terms of sample count or train test division. Hence, to perform uniform and robust comparison, we have evaluated all the methods used for comparison on the present dataset.

4.6.1. Comparison with standard ensemble techniques

To compare the performance of our Choquet fuzzy integral based ensemble method for COVID-19 cases identification on the present dataset, we have experimented with three standard ensembling methods: majority voting, weighted average, and average voting. For uniform performance comparison, we have tested these models on the present dataset. The comparative results are shown in Table 7. From the table records, it can be seen that our method generates 0.40%, 1.00%, 0.33%, and 0.66% better performance than the weighted average ensemble technique in terms of accuracy, precision, recall, and F1-score, respectively.

From Table 7, we note that the majority voting ensemble learning technique is the only other ensem bling technique whose results are close to those of the proposed Choquet fuzzy integral method. Also, from Fig. 8, we can see the confusion matrices of both the Choquet fuzzy integral method and the majority voting ensemble method. In the case of predicting the COVID-19 cases, both methods show the same results, but we also see that, in the case of majority voting, there are two cases where the method predicts the cases as those of COVID-19, although they belong to the Normal and Pneumonia classes. In the case of correct prediction of Normal cases, our method correctly predicts 267 cases, whereas the majority voting method correctly predicts 263 cases. In the

Fig. 8. Confusion matrices of (a) Majority Voting Ensemble and (b) Our proposed Choquet fuzzy integral method.

Table 8
The performance comparison of different direct classification models.

Method	Technique Used	Average			
		Precision (in %)	Recall (in %)	F1-score (in %)	Accuracy (in %)
Makris et al. [31]	Inception-v3	96.33	96.67	96.67	96.90
Hemdan et al. [33]	DenseNet121	97.67	96.67	96.67	97.96
Apostolopoulos at al [19].	VGG19	96.67	96.33	96.67	98.20
Makris et al. [31]	VGG16	98.00	97.33	97.67	98.04
Horry et al. [32]	ResNet50	85.67	90.67	86.67	88.77
Ardakani et al. [34]	ResNet 101	94.33	91.00	92.67	94.38
Hussain et al. [35]	CoroDet	96.34	96.00	96.00	96.66
Ismael et al. [36]	End-to-End CNN	95.67	94.67	95.00	96.09
Das et al. [37]	Bi-Level prediction model	97.87	98.14	98.00	98.45
Apostolopoulos at al [19].	MobileNetV2	98.00	97.33	97.67	97.80
Proposed Method	Choquet fuzzy integral based ensemble	99.00	99.00	99.00	99.02

case of Pneumonia prediction, our method outperforms the majority voting ensembling technique by 4 cases. In the case of majority voting, it incorrectly predicts 1 Pneumonia case as COVID-19 case and 9 Pneumonia cases as Normal cases, while, in our case, we only predict 6 Pneumonia cases as Normal cases.

4.6.2. Comparison with state-of-the-art methods

In this section, we describe the method level comparison with a number of state-of-the-art techniques proposed by the authors in Refs. [19,31–36], and [37]. To perform uniform, exhaustive and robust comparison, we have evaluated the performance of the state-of-the-art deep learning models on the present dataset, which means that we have implemented the method from scratch, trained the mentioned models on present training dataset and then evaluated them on the present test set. The comparative results are shown in Table 8. From the other state-of-the-art models, the highest overall accuracy is achieved by Bi-level prediction model proposed by Das et al. [37], the highest average precision value is achieved by VGG16 and MobileNetV2, which were proposed by Makris et al. [31] and Apostolopoulos at al [19]. We also see that the highest average recall and F1-score are achieved by the method proposed by Das et al. [37], where the author used a Bi-Level prediction model. When we compare the above results with those of our method, we see that our method outperforms all other ones, in terms of all performance metrics used for comparison.

From the results in Table 8, we can say that our method achieves the highest accuracy, as each classifier we use has a high accuracy. Further, Choquet fuzzy integral adds sensitivity to each classifier, because the final prediction given by the fuzzy integral fusion is influenced by the intermediate decisions based on the confidence scores of each classifier. Hence, this process gives fuzzy probabilities that have been proven to be more robust than normalized softmax probabilities. Our method also outperforms other models in terms of precision, recall and F1-score, by utilizing the fuzzy probability property to combine the precision, recall, and F1-scores of base classifiers.

Also, Fig. 9 shows that our Choquet fuzzy integral method outperforms other models when it comes to correct prediction of the true classes. In the case of COVID-19 cases prediction, our model correctly predicts 184 cases, whereas, in the case of other models, the second best result is given by both the DenseNet121 and MobileNetv2 models where each one correctly predicts 181 out of 185 cases. Also, we observe that our method correctly predicts 769 Pneumonia cases, whereas Bi-Level model predicts the second highest number of cases (768). MobileNetV2 performs the worst in Normal case detection, with 259 correct predictions and 10 incorrect predictions for Normal cases, predicting them as Pneumonia cases instead of Normal cases.

5. Conclusion

The severe outbreak of the COVID-19 has caused a pandemic. Many researchers have tried to find ways to assist the global response to COVID-19 and beyond. In this work, we have designed a method that can be used to predict COVID-19 cases by examining patients' chest X-ray images. We have proposed a Choquet fuzzy integral based classifier ensembling method, which can satisfactorily isolate chest X-ray images
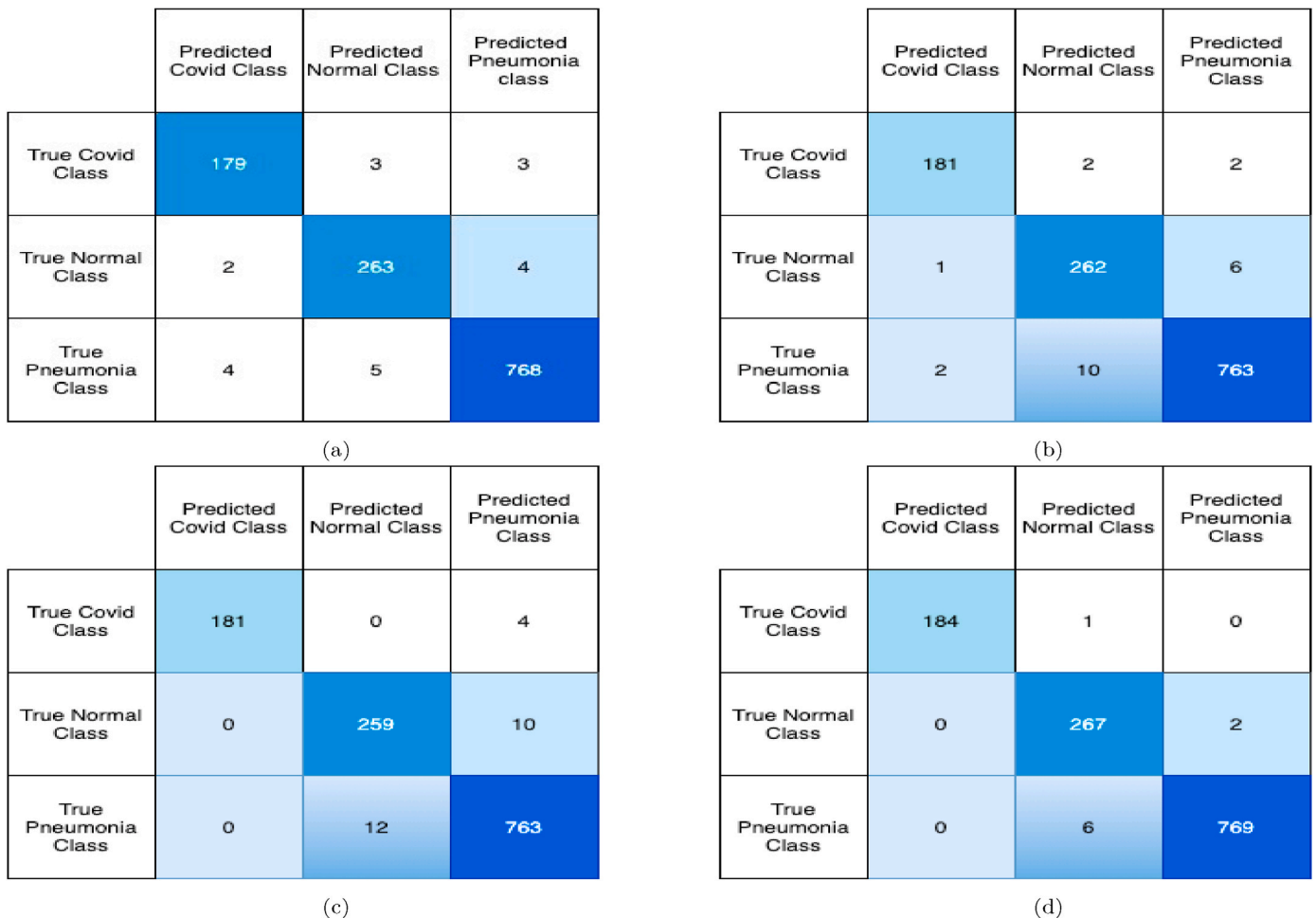

Fig. 9. Confusion matrices of some state-of-the-art models: (a) Bi-Level prediction model, (b) DenseNet121 model, (c) MobileNetV2, and (d) Our proposed Choquet fuzzy integral method.

of COVID-19 patients from the chest X-ray images of Normal subjects and Pneumonia patients. To this end, we have used three popular CNN models, namely Inception-v3, DenseNet121, and VGG-19 for feature extraction from the chest X-ray images, before using two densely connected neural networks to classify the images. Finally, classification results from individual CNN models are combined using a classifier ensemble technique to generate the final results. Since the datasets are too small to train the CNN models from scratch, we leverage the transfer learning approach to train the CNN models, and we have prepared a dataset taking images from publicly available chest X-ray image repositories. The experimental results demonstrated that the Choquet fuzzy integral based ensembling technique not only improves the performance of individual classifiers but also provides better results when compared to standard classifier combination methods like majority voting and weighted average. We also find that the present technique outperforms some of the state-of-the-art methods which we have considered for comparison using the present dataset. Further, we have tested the performance of our model on a new dataset, and its performance is comparable with the state-of-the-art results.

In spite of the good results, there is still room for improvement. The proposed model has failed to classify 0.98% of test data, which is not desirable considering the critical nature of the COVID-19. The reason for such a failure might be relatively lower performance of individual CNN models. Therefore, designing or selecting better individual CNN models might help achieve better results. Also, in this study, we make use of only three base CNN models in the mentioned ensemble process. Therefore, increasing the number base classifier models could be another improvement of the present method. Last but not the least, the current method could be used for solving other image-classification-based problems to prove its robustness across fields.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

References

[1] S. Sen, S. Saha, S. Chatterjee, S. Mirjalili, R. Sarkar, A Bi-stage feature selection approach for Covid-19 prediction using chest Ct images, Appl. Intell. (2021) 1–16.

[2] A. Garain, A. Basu, F. Giampaolo, J.D. Velasquez, R. Sarkar, Detection of covid-19 from ct scan images: a spiking neural network-based approach, Neural Comput. Appl. (2021) 1–14.

[3] M. Nour, Z. Cömert, K. Polat, A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization, Appl. Soft Comput. 97 (A) (2020) 106580.

[4] A. Abbas, M.M. Abdelsamea, M.M. Gaber, Classification of COVID-19 in chest X-ray images using DeTrac deep convolutional neural network, Appl. Intell. 51 (2) (2021) 854–864.

[5] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, et al., Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest CT: evaluation of the diagnostic accuracy, Radiology 296 (2) (2020) E65–E71.

[6] A.T. Sahlol, D. Yousri, A.A. Ewees, M.A. Al-Qaness, R. Damasevicius, M. Abd Elaziz, COVID-19 image classification using deep features and fractional-order marine predators algorithm, Sci. Rep. 10 (1) (2020) 1–15.

[7] S. Chattopadhyay, A. Dey, P.K. Singh, Z.W. Geem, R. Sarkar, Covid-19 detection by optimizing deep residual features with improved clustering-based golden ratio optimizer, Diagnostics 11 (2) (2021) 315.

[8] S. Chattopadhyay, A. Dey, P.K. Singh, Z.W. Geem, R. Sarkar, Covid-19 detection by optimizing deep residual features with improved clustering-based golden ratio optimizer, Diagnostics 11 (2) (2021) 315.

[9] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.

[10] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.

[11] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[13] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[14] A.K. Das, S. Ghosh, S. Thunder, R. Dutta, S. Agarwal, A. Chakrabarti, Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network, Pattern Anal. Appl. (2021) 1–14, https://doi.org/10.1007/s10044-021-00970-4.

[15] X. Liu, L. Ma, J. Mathew, Machinery fault diagnosis based on fuzzy measure and fuzzy integral data fusion techniques, Mech. Syst. Signal Process. 23 (3) (2009) 690–700.

[16] K.-C. Kwak, W. Pedrycz, Face recognition: a study in information fusion using fuzzy integral, Pattern Recogn. Lett. 26 (6) (2005) 719–733.

[17] A. Banerjee, P.K. Singh, R. Sarkar, Fuzzy integral based CNN classifier fusion for 3D skeleton action recognition, IEEE Trans. Circ. Syst. Video Technol. (2020) 1–10, https://doi.org/10.1109/TCSVT.2020.3019293.

[18] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: International Conference on Learning Representations, ICLR, 2015.

[19] I.D. Apostolopoulos, T.A. Mpesiana, Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks, Phys. Eng. Sci. Med. 43 (2) (2020) 635–640.

[20] M. Turkoglu, COVID ectecioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble, Appl. Intell. 51 (3) (2021) 1213–1226.

[21] Y. Oh, S. Park, J.C. Ye, Deep learning COVID-19 features on CXR using limited training data sets, IEEE Trans. Med. Imag. 39 (8) (2020) 2688–2700.

[22] H. Ko, H. Chung, W.S. Kang, K.W. Kim, Y. Shin, S.J. Kang, J.H. Lee, Y.J. Kim, N. Y. Kim, H. Jung, et al., COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image: model development and validation, J. Med. Internet Res. 22 (6) (2020), e19569.

[23] T. Mahmud, M.A. Rahman, S.A. Fattah, CovXNet: a multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization, Comput. Biol. Med. 122 (2020) 103869.

[24] M.E. Karar, E.E.-D. Hemdan, M.A. Shouman, Cascaded deep learning classifiers for computer-aided diagnosis of COVID-19 and pneumonia diseases in X-ray scans, Complex Intell. Syst. 7 (1) (2021) 235–247.

[25] H. Panwar, P. Gupta, M.K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, V. Singh, 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, Solit. Fractals 140 (2020) 110190.

[26] S. Rajaraman, J. Siegelman, P.O. Alderson, L.S. Folio, L.R. Folio, S.K. Antani, Iteratively pruned deep learning ensembles for COVID-19 detection in chest X-rays, IEEE Access 8 (2020) 115041–115050.

[27] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, P.R. Pinheiro, CovidGAN: data augmentation using auxiliary classifier GAN for improved covid-19 detection, IEEE Access 8 (2020) 91916–91923.

[28] Y. Karbhari, A. Basu, Z. W. Geem, G. T. Han, R. Sarkar, Generation of synthetic chest X-ray images and detection of COVID-19: a deep learning based approach, Diagnostics 11 (5).

[29] T.B. Chandra, K. Verma, B.K. Singh, D. Jain, S.S. Netam, Coronavirus disease (covid-19) detection in chest x-ray images using majority voting based classifier ensemble, Expert Syst. Appl. 165 (2021) 113909.

[30] P. Saha, D. Mukherjee, P.K. Singh, A. Ahmadian, M. Ferrara, R. Sarkar, Graphcovidnet: a graph neural network based model for detecting covid-19 from ct scans and x-rays of chest, Sci. Rep. 11 (1) (2021) 1–16.

[31] A. Makris, I. Kontopoulos, K. Tserpes, COVID-19 detection from chest X-ray images using deep learning and convolutional neural networks, in: 11th Hellenic Conference on Artificial Intelligence, 2020, pp. 60–66.

[32] M.J. Horry, M. Paul, A. Ulhaq, B. Pradhan, M. Saha, N. Shukla, et al., X-ray Image Based Covid-19 Detection using Pre-trained Deep Learning Models. https://engrxiv.org/wx89s/download.

[33] E.E.-D. Hemdan, M.A. Shouman, M.E. Karar, COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images, 2020, pp. 1–14, arXiv preprint arXiv:2003.11055.

[34] A.A. Ardakani, A.R. Kanafi, U.R. Acharya, N. Khadem, A. Mohammadi, Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks, Comput. Biol. Med. 121 (2020) 103795.

[35] E. Hussain, M. Hasan, M.A. Rahman, I. Lee, T. Tamanna, M.Z. Parvez, Corodet: a deep learning based classification for covid-19 detection using chest x-ray images, Chaos, Solit. Fractals 142 (2021) 110495.

[36] A.M. Ismael, A. Şengür, Deep learning approaches for covid-19 detection based on chest x-ray images, Expert Syst. Appl. 164 (2021) 114054.

[37] S. Das, S.D. Roy, S. Malakar, J.D. Velásquez, R. Sarkar, Bi-level Prediction Model for Screening Covid-19 Patients Using Chest X-Ray Images, Big Data Research, 2021, p. 100233.