Choquet Integral and Coalition Game-Based Ensemble of Deep Learning Models for COVID-19 Screening From Chest X-Ray Images

Pratik Bhowal, Subhankar Sen, Jin Hee Yoon, *Member, IEEE*, Zong Woo Geem, *Senior Member, IEEE*, and Ram Sarkar

***Abstract*—Under the present circumstances, when we are still under the threat of different strains of coronavirus, and since the most widely used method for COVID-19 detection, RT-PCR is a tedious and time-consuming manual procedure with poor precision, the application of Artificial Intelligence (AI) and Computer-Aided Diagnosis (CAD) is inevitable. Though, some vaccines have now been authorized worldwide, it will take huge time to reach everyone, especially in developing countries. In this work, we have analyzed Chest X-ray (CXR) images for the detection of the coronavirus. The primary agenda of this proposed research study is to leverage the classification performance of the deep learning models using ensemble learning. Many papers have proposed different ensemble learning techniques in this field, some methods using aggregation functions like Weighted Arithmetic Mean (WAM) among others. However, none of these methods take into consideration the decisions that subsets of the classifiers take. In this paper, we have applied Choquet integral for ensemble and propose a novel method for the evaluation of fuzzy measures using coalition game theory, information theory, and Lambda fuzzy approximation. Three different sets of fuzzy measures are calculated using three different weighting schemes along with information theory and coalition game theory. Using these three sets of fuzzy measures, three Choquet integrals are calculated and their decisions are finally combined. Besides, we have created a database by combining several image repositories developed recently. Impressive results on the newly developed dataset and the challenging COVIDx dataset support the efficacy and robustness of the proposed method. Our experimental results outperform many recently proposed methods.**

***Index Terms*—COVID-19, Chest X-Ray Images, Choquet Integral, Coalition Game, Deep Learning, Information Theory, Lambda Fuzzy.**

I. Introduction

CORONAVIRUS disease 2019 (COVID-19) is a contagious viral disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which was declared a worldwide pandemic by the World Health Organization (WHO) on 11 March 2020, following its accelerated worldwide dissemination and an unprecedented rise in the number of patients affected. The global initiative to produce an efficient and reliable COVID-19 vaccine is paying dividends. Nonetheless, a handful of vaccines have now been authorized worldwide while many more remain in production and it will take time to reach everyone, especially in under-developed countries. At present, reverse transcriptase-polymerase chain reaction (RT-PCR), a procedure carried out on swab samples taken from the respiratory tract, is the primary method used to diagnose COVID-19 disease. The RT-PCR assessments, however, are time-consuming and a repetitive manual procedure that has often contributed to a great deal of subjectivity.

Computer-Aided Diagnosis (CAD) technologies, combined with deep learning models are used to enhance the efficiency of the diagnosis and identification of COVID-19 infections from radiological images such as chest X-ray (CXR), computed tomography (CT), or lung ultrasound (LUS), and to minimize the manual intervention and error [48]–[53]. Deep learning approaches using Convolutional Neural Network (CNN) are regarded as one of the most robust and effective frameworks in diagnostic imaging assessments, particularly in image classification and segmentation problems.

Ensemble learning, which aggregates the decisions of multiple learning algorithms, further helps in increasing the accuracy of machine learning or deep learning models. Aggregation functions or operators like Arithmetic Mean, Weighted Arithmetic Mean (WAM), Geometric Mean, Median, etc. are some widely used approaches in ensemble learning. However, there are many problems with such aggregation functions. Some of them are - 1)

Manuscript received February 26, 2021; revised July 23, 2021; accepted September 1, 2021. Date of publication September 9, 2021; date of current version December 6, 2021. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) under Grant 2020R1A2C1A01011131. *(Corresponding authors: Jin Hee Yoon; Zong Woo Geem.)*

Pratik Bhowal is with the Department of Instrumentation and Electronics Engineering, Jadavpur University, Kolkata 700054, India (e-mail: pratikbhowal1999@gmail.com).

Subhankar Sen is with the Department of Computer Science and Engineering, Manipal University, Jaipur 303007, India (e-mail: subhankarsen2001@gmail.com).

Jin Hee Yoon is with the School of Mathematics and Statistics, at Sejong University, Seoul 05006, South Korea (e-mail: jin9135@sejong.ac.kr).

Zong Woo Geem is with the College of IT Convergence, Gachon University, Seongnam 13120, South Korea (e-mail: zwgeem@gmail.com).

Ram Sarkar is with the Department of Computer Science and Engineering, Jadavpur University, Kolkata 700054, India (e-mail: ram.sarkar@jadavpuruniversity.in).

Digital Object Identifier 10.1109/JBHI.2021.3111415

Aggregation functions like WAM do not consider the correlation among the inputs (in our case classifiers). 2) Outliers have a huge influence on the average/mean values used here. This results in a considerable decrease in the accuracy of the model even if only one classifier is weak.

To deal with these problems, more complex aggregation functions can be found in the literature, one of them being fuzzy integrals. Fuzzy integrals like Choquet and Sugeno, unlike other ensemble methods take into consideration the importance of the subsets of classifiers. Therefore, Fuzzy integrals are considered as a more generalized aggregation function as they assign weights to not only the individual classifiers but even the subset of classifiers.

To understand why considering subsets of classifiers is important, we apply the following idea. A classifier may not perform well individually, however it can do well when combined with other classifiers. The vice versa is also true, i.e. a classifier may perform well individually, but when its decision is combined with the decision of other classifiers, performance of the overall framework may degrade. Another possibility is that two classifiers have no influence on each other when their decisions are taken together. The effect of classifiers deciding on a group is not captured by the conventionally used aggregation methods like majority voting and weighted averaging. To deal with this problem, the concept of *Fuzzy integrals* has been introduced in the literature.

The calculation of fuzzy measures is an NP-hard problem [25]. Additive fuzzy measures have been proposed to deal with this problem. However, additive fuzzy measures are too simplified to emulate practical situations. The alternative which has been proposed is neither oversimplified nor too complex to calculate and it is *lambda fuzzy approximation*. In this paper, we use this method which is supported by the *coalition game theory* to calculate the fuzzy measures. Further, we use *Choquet integral* for the aggregation of the three CNN models.

A. Motivation and Contributions

Many breakthroughs and promising outcomes in the area of medical image processing and healthcare have been accomplished with the use of transfer learning, and ensemble techniques [1]–[5]. Due to their compact genomic organization, high mutation rate, and poor evolutionary conservation [6], the modern SARS-CoV2 has created new challenges to the scientists, and radiologists and medical professionals are not completely familiar with all the complexities of COVID-19 lung infection. To this end, the primary motivation of our research study is to develop an automated system for efficient computerized detection of COVID-19 from the CXR image with the aid of deep learning models and fuzzy ensemble learning.

The main contributions of our paper are enlisted below:

1) We apply deep feature extraction techniques on the input CXR images using fine-tuned deep convolutional neural network (DCNN) architectures, pre-trained on the ImageNet dataset for obtaining salient image descriptors.
2) We propose a novel method of calculating the fuzzy measures of individual classifiers using *coalition game theory* and *Information Theory* before using the *lambda fuzzy approximation* for calculating the fuzzy measures of the set of classifiers. We also use three different *weighting schemes* for calculating the *Shapley values*.
3) We make results intuitive and interpretable with the aid of *Explainable AI* to enhance the transparency of the black-boxed AI systems. This includes heat map-based visualization of the network activation maps by employing the RISE algorithm [7] to provide explainability about the class prediction. This will enable radiologists to localize the region(s) in the CXR for examining pulmonary damage severity.
4) We have created a *new CXR image dataset* known as Novel COVID-19 Chestxray Repository by combining three publicly available CXR image repositories for future use and reference by researchers in the field of COVID-19 identification and diagnosis.

The remaining paper is organized as follows. Section II offers a summary of the literature relating to the diagnosis of pulmonary disease relevant to COVID-19 disease. A detailed overview of our proposed solution is given in Section III. Section IV provides the dataset description, experimental results obtained by employing the proposed model, and further discussions including error analysis and comparison with state-of-the-art. Finally, the paper is concluded in Section V by outlining some future research directions.

II. Literature Survey

1) Recent Deep Learning Approaches: In the recent past, machine learning and deep learning-based approaches have been employed for COVID-19 identification and diagnosis from CXR images, CT scans, and LUS images. In this section, we have discussed the automated CAD systems used for COVID-19 detection and diagnosis.

Islam *et al.* [8] used CNN followed by long short-term memory(LSTM) network for automated detection of COVID-19 from CXR images using deep learning. In this proposed model, CNN was used as a feature extractor, and the LSTM network was used to classify COVID-19 based on the CNN-derived deep image descriptors. The framework proposed, however, had some drawbacks. The developed system could not distinguish other views of CXRs such as anterior-posterior (AP), lateral, etc., as it only focused on the posterior-anterior (PA) view of CXR images. Besides, COVID-19 images contained various symptoms of the disease that they had failed to efficiently classify.

A COVID-19 diagnostic method called COVIDetectioNet was proposed by Turkoglu *et al.* [9]. Deep features were extracted from the CXR images in this method using the pre-trained AlexNet architecture, preceded by the selection of the most salient features from the image descriptors obtained using a feature selection technique known as Relief algorithm followed by generation of decision scores using Support Vector Machine(SVM). However, estimating the optimal parameters for the Relief algorithm and the SVM classifier was seen as a limitation of the proposed workflow. A deep learning approach for COVID-19 detection was developed by Sahlol *et al.* [10] in

which Inception was employed for deep feature extraction from CXR images accompanied by filtering the resulting features using a swarm-based feature selection algorithm, known as the Marine Predators Algorithm, and fractional-order calculus for selection of the most relevant features.

In [13], Ucar *et al.* implemented a deep Bayesian optimization-based SqueezeNet model called COVIDiagnosis-Net. By fine-tuning the pre-trained SqueezeNet model, and performing hyperparameter optimization the robustness and efficiency of the proposed method were leveraged substantially. The SqueezeNet is preferred in embedded applications due to its functional layout and generalization efficiency, thereby opening the door for robust product deployment that can run mobile and attract the end-user. In [14], a deep neural network architecture was used by Roy *et al.* for decision scores prediction associated with a single image of LUS and recognizes the regions comprising pathological significant objects in a weakly supervised manner. A lightweight methodology based on uni-norms for the aggregation of frame-level prediction scores was also adopted and the score associated with the LUS video series was computed.

In [15], the hyper learning binary dragonfly algorithm (HLBDA) was introduced by Too *et al.* as a feature selection technique to hand-pick the vital features from the images. For the COVID-19 classification, the proposed model was used and the results showed the supremacy of HLBDA in leveraging the precision of the classification and reducing the number of features selected. A machine learning approach was used by Elaziz *et al.* [16] for a 2-class classification problem(COVID-19 and non-COVID) using CXR images. The features were extracted from the CXRs using Fractional Multichannel Exponent Moments (FrMEMs), followed by a modified Manta-Ray Foraging Optimization based on differential evolution used for selecting salient features. The feature selection technique enabled leveraging the classification efficiency and minimizing the resource consumption and computation complexity. However, the authors could not obtain optimal hyperparameters for training the KNN classifier and was hence a drawback of this study. Babukarthik *et al.* [18] propounded a deep learning approach namely Genetic Deep Learning CNN (GDCNN). The GDCNN method involved population initialization where the population was arbitrarily initialized and the population was continually progressed on a generation-by-generation basis to create improved architectures by utilizing redefined genetic operations.

Abbas *et al.* [45] proposed a deep CNN architecture based on class decomposition, coined as Decompose, Transfer, and Compose (DeTraC) model for enhancing the adaptation of pre-trained transfer learning models for detection of COVID-19 cases. The proposed class decomposition technique improved the low variance classifiers which facilitated investigating the decision boundaries. As a result, DeTraC addressed the issue of irregularities in the image dataset. Wang *et al.* [38] revamped the original COVID-Net by improving network architecture for COVID-19 detection from CT scan images. The authors also introduced contrastive cross-site learning and a learning rate scheduling strategy based on cosine annealing which enhanced the joint training protocol and facilitated domain invariance enhancement, remarkably reducing the inter-site data heterogeneity. Panetta *et al.* [44] presented a shape-dependent Fibonacci -p patterns-based feature extractor for distilling out the intricate textural features from CXR images.

1) Fuzzy Ensemble Techniques: The authors of [23] presented a Mobius-like transformation of discrete fuzzy measures and proposed an evaluation formula for the corresponding Choquet-like integral, including the Sugeno integral. Murillo *et al.* [42] proposed a revamped Heuristic Least Mean Squares (HLMS) implementation, a gradient-based algorithm for identification of fuzzy measures, that enhanced the convergence by improving the formula for iterative estimation of fuzzy measure coefficients and the monotonicity check. [39] devised a mathematical programming approach for the derivation of fuzzy measures based on the correlation coefficient. Kundu *et al.*. [47] proposed an ensemble strategy that generates fuzzy ranks of classifiers using Gompretz function. They adaptively fused the decision scores of the classifiers to make the final predictions of the test cases.

Beliakov *et al.* [40] introduced k-interactivity which reduced the number of variables and constraints which further facilitated the reduction of the complexity of learning fuzzy measures. The fuzzy measures were learned using a linear programming problem. In [41], Grabisch *et al.* introduced a gradient algorithm for identifying fuzzy measures from empirical data. The paper presented the use of a mix of standard optimization algorithms, such as Lemke's method, and utility theory for identifying the 2^n coefficients based on semantical considerations. Choquet integral of fuzzy number-valued functions based on $\sigma - \lambda$ rules was formulated by the authors of [26]. The paper further proposed genetic algorithm (GA) based optimization for computing fuzzy measures on fuzzy number-valued data. For learning monotone models, Tehrani *et al.* [27] demonstrated the use of Choquet integral as an aggregator operator in machine learning problems.

III. Proposed Work

A. Method Overview

For the COVID-19 detection from CXR images, in the present work, we have proposed a lambda fuzzy-based ensemble model of DCNN architectures. At first, the CXR images have been preprocessed. Then fine-tuned, well-established DCNN architectures, pre-trained over the ImageNet dataset [36] namely VGG16, Xception, and InceptionV3 are for feature extraction. The image descriptors obtained are then fed as input into a Multi-layer Perceptron (MLP) classifier with softmax output for a 3-class classification problem (COVID-19, Pneumonia, and Normal). The confidence scores obtained per image, across the three DCNN models used, are then combined using Choquet integral into a confidence matrix. For the evaluation of the Choquet integral and the Choquet integral itself, the fuzzy measures required are calculated as follows. We calculate the Shapley values, using coalition game theory and information theory, which become the fuzzy measures of the single classifier set. We introduce three different weighting schemes to calculate the

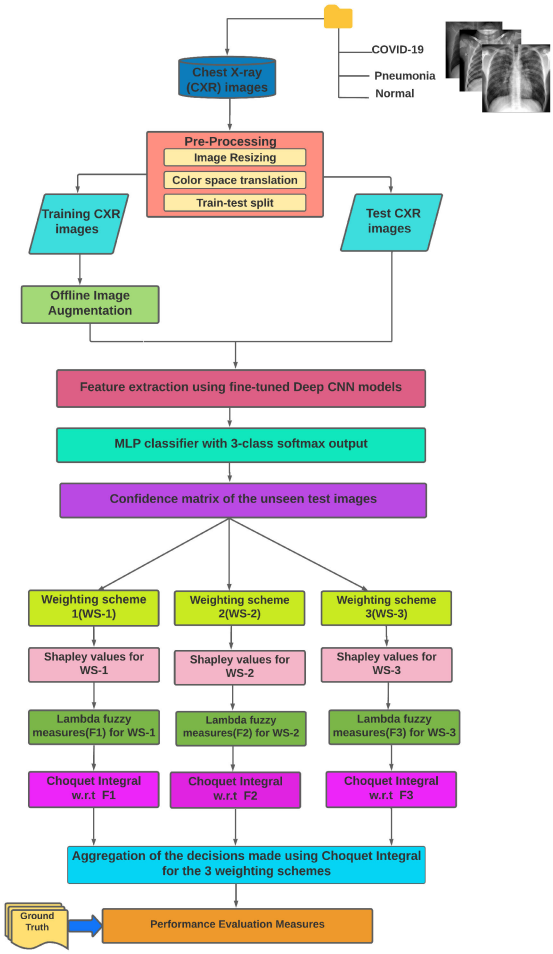

Fig. 1. Overall block diagram of the proposed ensemble model used for screening COVID-19 from CXR images.

Shapley values better. We then use lambda fuzzy to calculate the fuzzy measures of the other subsets of classifiers whose cardinality is greater than 1 and then use Choquet integral for aggregation. Three aggregations done with respect to the three weighting schemes are combined at the end. Fig. 1 demonstrates the flowchart of our proposed methodology. In Fig. 1 it can be observed that first the training images are augmented and features are extracted from both training and the test images. The images are then classified using the MLP classifier. As mentioned above, we use three different sets of fuzzy measures for the calculation of three different Choquet integrals. First, the weights are calculated for each set. In the next step, the Shapley values are calculated. Lambda fuzzy approximation is then applied to calculate the fuzzy measures for subsets with cardinality greater than one. In the next step, the three different Choquet integrals are calculated for each of the sets of fuzzy measures. Finally, as shown in Fig. 1, the results of the three Choquet integrals are aggregated using majority voting and the accuracies are calculated using different evaluation metrics. In addition to visualize the results, we yield saliency maps for our model's predictions using the RISE algorithm [7], an approach used to explain black-box models by visualizing the semantically significant regions of the input images based on the generated class prediction.

B. Data Preprocessing and Augmentation

The CXR images of varying resolutions are resized before feeding as input into the DCNN models. The images are downscaled to 512x512 pixels using bicubic interpolation, color space translation was done from RGB to grayscale. All input images are normalized using *proprocess_input* function of the Keras Applications module [37], which subtracts the mean RGB channels of the ImageNet dataset from each input CXR image.

Further, image augmentation techniques are applied to the preprocessed images. Various image augmentation techniques such as image rotation by 5°, 10°, 15°, and 20°, vertical flip, horizontal flip, shearing, horizontal and vertical image translation are performed.

C. Deep Feature Extraction and Model Training

The primary agenda of deep feature extraction is to distill out salient, discriminating information from the original raw images and represent that information in a lower dimensionality space. In this proposed work, we have employed the approach of deep feature extraction from CXR images using fine-tuned DCNN models. Using the GlobalAveragePooling operation, we then extract deep salient features from the convolutional layers directly before or after the max-pooling layer, from unique blocks of the pre-trained DCNN models. By employing the Keract library [17], the visualization of activation maps generated corresponding to layers in the DCNN architectures are studied, and consequently the suitable convolutional layers for deep visual feature extraction is determined. This method makes deep black-box neural networks highly interpretative and easy to debug. The features extracted from these specific layers of the pre-trained DCNN models, as represented in Table I, are then concatenated into a single feature vector known as a deep image descriptor.

For feature extraction, we have used three standard DCNN models, pre-trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [19], namely VGG16 [20], Xception [21] and InceptionV3 [22]. In Fig 2, VGG16 has been shown as an example for the extraction of discriminating features from the input CXR images. The global average pooling operation is applied to 5th(block2_conv2), 9th(block3_conv3), 13th(block4_conv3) and 17th(block5_conv3) convolutional layers comprising 128, 256, 512, and 512 channels respectively. These features are fused to form a single deep feature vector having a length of 1408. The other two deep CNN models, namely Xception and InceptionV3, are used in a similar manner for the extraction of rich image descriptors from CXR images. The modified DCNN models, with fully-connected layers added after the feature vector, are then fine-tuned and trained for 100 epochs for optimizing the weights of the model which thereby outputs a refined deep feature embedding The obtained feature vector after model training is finally fed into an MLP classifier, consisting of a Fully Connected (FC) layer of 512 neurons (with activation of Rectified Linear Unit or ReLU) followed by a

TABLE I
DETAILS OF REGIONS OF FEATURE EXTRACTION FROM THE GIVEN DCNN MODELS ALONG WITH FEATURE MAP DIMENSIONS OF THE CONVOLUTIONAL LAYER. THE THIRD COLUMN DEPICTS THE DIMENSION OF THE FEATURE VECTOR EXTRACTED FROM THE CXR IMAGES

Pretrained DCNN model	Model layer no.	Deep feature vector
VGG16	**5** (256x256x128), **9** (128x128x256), **13** (64x64x512), **17** (32x32x512)	**1408**
Xception	**27** (64x64x728), **37** (32x32x728), **126** (16x16x1536)	**2992**
InceptionV3	**11** (126x126x80), **18** (61x61x64), **28** (61x61x64), **51** (61x61x64), **74** (61x61x64), **101** (30x30x128), **120** (30x30x192), **152** (30x30x192), **184** (30x30x192), **216** (30x30x192), **249** (14x14x448), **263** (14x14x320), **294** (14x14x320)	**2320**

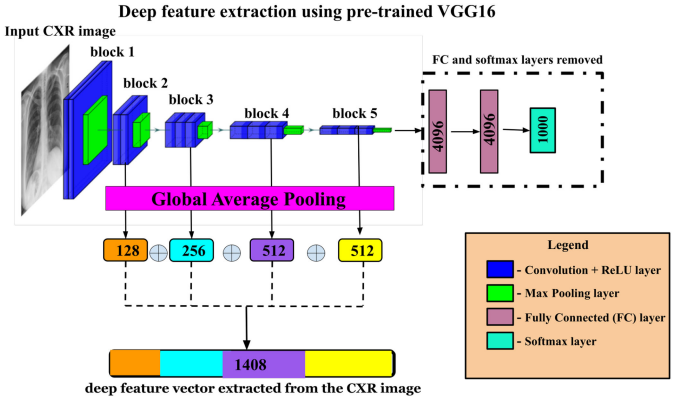

Fig. 2. Schematic diagram of VGG16 network architecture for deep feature extraction from the CXR images.

dropout rate of 0.5. Finally, there is the output layer of three neurons (with softmax activation) for the 3-class classification of the CXR images.

1) Loss Function: Inspired by the approach given by the authors of [38], we have used a combination of categorical cross-entropy loss and contrastive loss as our training objective $\mathcal{L}_{overall}$. The cross-entropy loss function $\mathcal{L}_{CE}$ is given by (1).

$$\mathcal{L}_{CE} = -\sum_{i=1} y_i \cdot \log \hat{y}_i \tag{1}$$

where $\hat{y}_i$ is the i-th predicted value in the network output, y_i is the corresponding target value. Categorical cross-entropy is used for measuring the classification performance of the DCNN models. Contrastive loss $\mathcal{L}_{con}$ [11] is a hardness-aware loss function that is used to map vectors that model the cosine similarities of a network output with an example of a positive class and a negative class. Contrastive learning also facilitates in regularizing the latent space and helps in addressing the issue of domain gap while dealing with heterogeneous CXR data sources, which resulted in robust joint training. A pair (z_i, z_j) is denoted as positive pair if they belong to the same class, otherwise they are denoted as negative. For a mini-batch consisting of K samples, the contrastive loss $l(z_i, z_j)$ over each positive pair (z_i, z_j) is given by

$$l(z_i, z_j) = -\log \frac{\exp\left(sim(z_i, z_j)/\tau\right)}{\sum_{k=1}^{K} f(z_i, z_k) \cdot \exp\left(sim(z_i, z_k)/\tau\right)} \tag{2}$$

where $f(z_i, z_k)$ outputs 0 for positive pair and 1 for negative pair. τ denotes the temperature parameter. The overall contrastive loss $\mathcal{L}_{con}$ is computed as the summation of $l(z_i, z_j)$, as given in 2, for all positive pairs in a given mini-batch for both (z_i, z_j) and (z_j, z_i). The overall training objective is given as:

$$\mathcal{L}_{overall} = \mathcal{L}_{CE} + \alpha \mathcal{L}_{con} \tag{3}$$

where α is a hyper-parameter.

2) Learning Rate Schedule: Cosine annealing [12] is used for the model training as a learning rate scheduler. The learning rate decay strategy, as given in 4), enhances the robustness of the model training and also facilitates in addressing the large variance present in the input space when dealing with heterogeneous CXR images.

$$\eta_t = \eta_{min} + \frac{1}{2}\left(1 + \cos\left(\frac{t}{T}\pi\right)\right)(\eta_0 - \eta_{min}) \tag{4}$$

where η_t is the updated learning rate after executing the scheduler module, η_{min} denotes the threshold of minimum learning rate, η_0 is the base learning rate (1e-3), t is the current epoch and T is the total number of training epochs.

D. Proposed Ensemble Method

As mentioned earlier, in the proposed method, we have used Choquet integral for aggregating the decision of the three deep learning models.

1) Summary of the Proposed Ensemble Method: As already mentioned, we have used Choquet integral to combine the decisions of the classifiers. However, for that, we need to calculate the fuzzy measures for all the subsets of classifiers. To calculate the fuzzy measures, we have used the Shapley value of the Coalition Game theory and Lambda fuzzy approximation. To calculate the Shapley value, marginal contribution has to be calculated. We have used mutual information and conditional mutual information to calculate the marginal contribution. It is to note that the direct calculation of mutual information is highly difficult, and so we have used heuristics. Finally, three different weighting schemes have been used, and a different set of fuzzy measures is obtained for each of them. These three sets of measures have been used to evaluate three different Choquet integrals, each of which used a different set of fuzzy measures. In the end, a majority voting scheme is used to combine the decisions of these three Choquet integrals.

2) Introducing Lambda Fuzzy Measures: However, for this, we need to calculate the fuzzy measures which is a very challenging task due to the monotonic conditions of the fuzzy measures and also the super-additivity and the sub-additivity schemes. Let X be a universe of discourse, then $\mu : 2^X \rightarrow [0, 1]$ is a fuzzy measure if it satisfies (i) (Boundary Condition) $\mu(\emptyset) = 0$ and $\mu(X) = 1$. (ii) (Monotonicity) For every $A, B \in 2^X$, if $A \subseteq B$, then $\mu(A) \leq \mu(B)$. (iii) (Continuity) For every sequence of X, if either $A_1 \subseteq A_2 \subseteq \ldots$ or $A_1 \supseteq A_2 \supseteq \ldots$, then

$\lim_{i\to\infty}\mu(A_i) = \mu(\lim_{i\to\infty} A_i)$. One easy way to approximate the fuzzy measures is to use the lambda fuzzy approximation. In this case, we just need to calculate the fuzzy measures of the individual classifiers, and the fuzzy measures for the subsets of classifiers would be calculated automatically using (5). In this equation λ is a parameter and in our case A_i and A_j represent classifiers.

$$\mu(A_i \cap A_j) = \mu(A_i) + \mu(A_j) + \lambda\mu(A_i)\mu(A_j) \quad given\ \mu(A_i \cup A_j)\ such\ that\ \lambda > -1 \tag{5}$$

λ is calculated using (6). In this equation, N is the number of classifiers and A_n represents the classifiers (in our case).

$$\lambda + 1 = \prod_{n=1}^{N}(\lambda\mu(A_n) + 1) \tag{6}$$

3) Use of Coalition Game Theory for the Calculation of Fuzzy Measures of Individual Classifiers: For calculating the fuzzy measures of the individual elements or in this case classifiers, we take the help of coalition game theory and information theory. Shapley value, a concept used in the coalition game theory, is applied to calculate the payoff for each element (in this case classifier) when the grand coalition is formed. The formula for calculating the Shapley value is shown in (7). In (7), n is the total number of players and the summation is done over all the subsets S of the grand coalition N such that they do not contain the element i. In this context n is the total number of classifiers, S is all the possible subsets of classifiers of the grand coalition of classifiers N, which do not contain the classifier i. The function v is known as the Characteristic function and $v(S)$ indicates the worth of the subset or in other words, it indicates the total payoff that is available for division among the members (in this case classifiers) of S. The payoff, φ that is paid to i in the coalition is calculated using Shapley Value.

$$\varphi_i(v) = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(n-|S|-1)!}{n!}(v(S\cup\{i\}) - v(S)) \tag{7}$$

The expression $(v(S\cup\{i\}) - v(S))$ is known as the marginal contribution. Shapey value can also be interpreted as an expected marginal contribution. It is the difference between the value of the characteristic function of a subset containing the element i and the value of the characteristic function of the subset not containing i but containing all other elements present in the last subset. This difference indicates the contribution that i has on the subset. A higher value of marginal contribution indicates that adding i to the subset increases the value of v a lot, thereby increasing the profit of that subset.

4) Calculating Marginal Contribution Using Information Theory: We take the help of information theory to calculate the marginal contribution. Classifiers whose decisions are highly correlated with each other and the correlation of whose decision to the true label decreases when they are considered together are known as redundant classifiers. If two classifiers are completely unrelated and if their relevance to the true label does not increase when they are considered together they are known as independent classifiers. Two classifiers are said to be interdependent if their relevance to the true label increases when they are considered together instead of being considered individually. We use mutual information and conditional mutual information to understand how interdependent, redundant, or independent a classifier is with the set of classifiers already selected. Let L be the subset that has already been selected and let f_k be the classifier that we consider to be added to this subset. Further, $I(L; class)$ is the mutual information between the decisions made by the classifiers in the set L and the *true labels*, while $I(f_k; class|L)$ is the conditional mutual information between the predictions of the classifier f_k and the *true labels* given the decisions made by the classifiers in the set L. The following formulas hold.

$$I(f_k; class|L) < I(L; class)\ for\ redundancy \tag{8}$$

$$I(f_k; class|L) > I(L; class)\ for\ interdependency \tag{9}$$

$$I(f_k; class|L) = I(L; class)\ for\ independence \tag{10}$$

We define the difference between mutual information and conditional mutual information as the marginal contribution. We choose this difference as it gives us some idea about the contribution of f_k in the subset L. In other words, we capture the information about the true label that we gain after the inclusion of f_k to the already selected subset of classifiers L. Therefore, the equation becomes,

$$marginal\ contribution = I(f_k; class|L) - I(L; class) \tag{11}$$

5) Heuristics to Calculate Mutual Information: However, it is highly impractical and in some cases impossible to calculate marginal contribution using (11), because of the difficulties in calculating the formulas regarding mutual information and conditional mutual information. Hence, we choose to use a heuristic method for calculating the marginal contribution. The marginal contribution of f_k would be high when the mutual information between f_k and the true label is high, while the mutual information between f_k and L is low. Also, the mutual information between the elements of the set L must be low. When these conditions are met, the marginal contribution of f_k becomes larger. Let l be the cardinality of set L. We frame the equation as shown in 12, where $I(f_k, f_i)$ is the mutual information between the classifiers f_k and f_i, where, f_k is the classifier under consideration as mentioned above, while, f_i is a classifier of the set L.

$$marginal\ contribution = I(f_k; class) - \frac{1}{l}\sum_{f_i\in L} I(f_k, f_i) \tag{12}$$

6) The Three Weighting Schemes: The assumption on which the (12) is based, however, is not entirely correct. The assumption is that we assign maximum values to classifiers, f_k, which convey maximum information about the true label (class) and provide minimum redundant information when compared to other classifiers already present in the subset L. This is not entirely correct as when classifiers already have high accuracy, they convey more information about the true label (class), but at the same time they become more and more repetitive. Thus, a classifier which has higher accuracy and is repetitive should be penalized more as compared to a classifier which has lower

accuracy and is repetitive. We incorporate this logic using (13)

$$marginal\ contribution = I(f_k; class) - \frac{1}{l}\sum_{f_i \in L} weights * I(f_j, f_i) \quad (13)$$

We can either assign them a set of weights logically based on the validation accuracy they obtained or assign them a few different sets of weights and take the average result. We use the second technique as this is a heuristic method and choosing one set of correct weights may not be possible. Rather a range of weights may be considered suitable, and we need to use as many of them as possible. In our case, we aim to figure out three such sets of weights and averaged out their results.

The three sets of weights are based logically as follows. Since we want to assign higher mutual information values to the classifiers which have higher accuracy, we assign them lower weights. In the first logic, we subtract the validation accuracy of each of the pair of classifiers f_k and f_i from 1 and take the average. To normalize, we then divide using the maximum weight. In the second set of weights, we take the reciprocal of the validation accuracies of both the classifiers, average them and then divide by the maximum to normalize. The third weighting scheme is of taking the negative of the logarithm of the validation accuracies of the two classifiers and then taking the average. We then normalize the weight by dividing it by the maximum weight. Let V_k and V_i be the validation accuracies of the two classifiers f_k and f_i respectively. For weighting scheme 1, we have,

$$weight_1 = \frac{(1 - V_i) + (1 - V_k)}{2} \quad (14)$$

$$Normalized\ Weight_1 = \frac{weight_1}{Max(weight_1)} \quad (15)$$

For weighting scheme 2, we have,

$$weight_2 = \frac{(1/V_i) + (1/V_k)}{2} \quad (16)$$

$$Normalized\ Weight_2 = \frac{weight_2}{Max(weight_2)} \quad (17)$$

For weighting scheme 3, we have,

$$weight_3 = \frac{(-log(V_i)) + (-log(V_k))}{2} \quad (18)$$

$$Normalized\ Weight_3 = \frac{weight_3}{Max(weight_3)} \quad (19)$$

Thus we get three sets of fuzzy measures.

7) The Choquet Integral: For each set of fuzzy measures, we evaluate a different Discrete Choquet integral. The form of the Discrete Choquet integral in (20) was introduced in the paper [28]. The discrete Choquet integral with respect to fuzzy measure μ is given by (20), where $\mathbf{x} = (x_{(1)}, x_{(2)}, x_{(3)} \ldots x_{(n)})$ is a non-decreasing permutation of the input $\mathbf{x}$, which in our case are the confidence scores, $A_{(i+1)} = \phi$, $A_i = \{(1), (2), (3), \ldots .(n)\}$, which in our case is the subset of classifiers and $\mu(A_i)$ are the fuzzy measures for the subsets of classifiers which we have calculated using Shapley value and information theory.

$$\mathbf{C}_\mu(\mathbf{x}) = \sum_{i=1}^{n} x_{(i)}(\mu(A_i) - \mu(A_{i+1})) \quad (20)$$

After rearranging the terms and assuming that $x_0 = 0$, the Discrete Choquet integral [24] can be represented by (21), which we have used in our calculations.

$$\mathbf{C}_\mu(\mathbf{x}) = \sum_{i=1}^{n} (x_{(i)} - x_{(i-1)})\mu(A_i) \quad (21)$$

The Choquet integral creates a confidence score for the three classes after combining the confidence scores of the three classifiers for each of the classes. The image is classified into the class with the maximum confidence score for each of the Choquet integral with respect to the three sets of fuzzy measures.

The majority voting is used on the result obtained from the evaluation of the three Choquet integrals with respect to the three sets of fuzzy measures for making the final decision.

TABLE II
STATISTICS OF THE NUMBER OF CXR IMAGES USED FROM DIFFERENT PUBLICLY AVAILABLE REPOSITORIES TO FORM THE NOVEL COVID-19 CHESTXRAY REPOSITORY

Dataset	COVID-19	Pneumonia	Normal
COVID Chestxray set [34]	521	239	218
COVID-19 Radiography Database [33]	219	1345	1341
Actualmed COVID chestxray dataset [35]	12	0	80
Total Number of CXR images:	**752**	**1584**	**1639**

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Dataset Description

In this paper, by combining publicly available CXR image repositories, we have created a CXR image archive known as the Novel COVID-19 Chestxray Repository. The dataset can be publicly accessed on Kaggle [30]. Three different datasets derived from Github and Kaggle databases are used to do this. In our research, CXR frontal and lateral images are used because this view of radiography is widely used by radiologists for performing clinical diagnostic assessments. II offers a detailed description of the Novel COVID-19 Chestxray Repository and its parent image datasets. We have outlined how this dataset is generated in the following section.

- **COVID-19 Radiography Database**: There are 219 COVID-19, 1345 viral pneumonia and 1341 normal CXR radiographic images in the COVID-19 [29] Radiography Database. This dataset was developed, in collaboration with medical professionals from Pakistan and Malaysia, by a group of researchers from the University of Qatar in Doha, Qatar, and the University of Dhaka in Bangladesh. With the advent of new cases of COVID-19 patients internationally, this database is constantly updated. This database can be found publicly at [33].
- **COVID-Chestxray set**: A public image archive on Github [34] consisting of both CT scans and digital CXR files has been created by Joseph Paul Cohen, Paul Morrison, and Lan Dao. These were mainly collected from

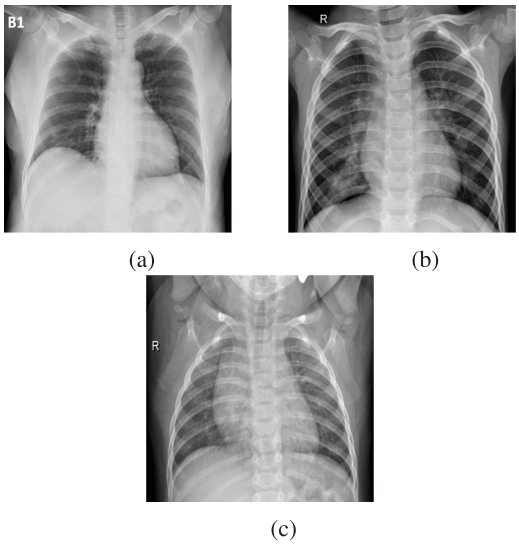

Fig. 3. (a) COVID-19 positive CXR, (b) Pneumonia infected CXR, and (c) Normal CXR.

retrospective cohorts of pediatric patients from the Women and Children of Guangzhou Medical Center. The dataset consists of 521 COVID-19 positive, 239 viral and bacterial pneumonia, and 218 normal radiographic CXR images of varying image resolutions. We have excluded the CT scan images while combining this dataset with the rest of the parent CXR image repositories.

- **Actualmed COVID chestxray dataset**: The Actualmed COVID chestxray dataset comprises 12 positive COVID-19 and 80 normal CXR images. The database is available at the Github link to the database at [35].

CXR images of COVID-19, Pneumonia, and Normal (healthy) classes are included in this newly created dataset, with a total of 752, 1584, and 1639 images respectively. Illustrations of CXR radiographic images of each class are given in Fig. 3.

B. Division of the Dataset

About 10% of the CXR images in the dataset have been used for testing. The rest of the samples are divided into a 4:1 ratio to form the training and validation sets. Thus, the number of test images is equal to 399, while the numbers of training and validation images are 2861 and 715 respectively. The test images have been randomly selected. After performing image augmentation protocols, the sizes of the training and validation sets are increased fourfold. Table III summarizes the dataset distribution which we have employed for performing the experiments presented in this paper.

C. Experimental Setup

To implement the proposed method, we have used Python programming language with the aid of Keras package with Tensorflow as the deep learning framework. We have run the codes on Google Colaboratory having the following system specifications: Nvidia Tesla T4 with 13 GB GPU memory, 1.59 GHz GPU Memory Clock, and 12.72 GB RAM.

TABLE III
CLASS DISTRIBUTION OF THE PROPOSED NOVEL COVID-19 CHEST XRAY REPOSITORY USED TO EVALUATE THE PROPOSED METHOD

Number of images per class after augmentation					
Data Percentage		COVID-19	Pneumonia	Normal	Total
Training and Validation Data (90%)	Training data (80%)	2163	4561	4719	11443
	Validation data (20%)	540	1139	1180	2859
Test data(10%)		76	159	164	399

D. Results

1) Visualization: In this section, we enhance the interpretability of the COVID-19 analysis with the aid of Explainable AI, with an attempt in overcoming the black-box dilemma, and making deep learning model predictions intuitive and understandable in the field of CAD-based diagnosis of COVID-19.

Using RISE [7], an approach to describe black-box models by estimating the significance/saliency of the input image regions for the model's prediction, we generate saliency maps for our model's predictions to illustrate the proposed study's results qualitatively. In this method, 500 randomly masked representations of a given CXR image are queried and a weighted mask corresponding to each output class is produced using their generated classification scores. Therefore, the mask regions that retain semantically important characteristics will result in a higher confidence score and thus a higher weight for the respective class in the final image mask. This heatmap produced is then superimposed over the CXR image of the input, making it intuitive for radiologists to examine the lung regions affected by the disease. The saliency maps can also be useful to the radiologist in localizing the X-ray areas for performing severity assessments. Fig. 4 shows the saliency maps generated using RISE [7] approach corresponding to COVID-19 positive, Pneumonia, and Normal CXR images respectively. The regions where colors tend towards infrared (red regions) depict the pulmonary areas which are severely affected. As we gradually tend towards the end of the blue color spectrum, the severity of lung infection decreases.

2) Classification Performance: We have performed a 3-class classification of the CXR images which are COVID-19 affected, Pneumonia affected, and Normal lungs. We have used three pre-trained models, namely, VGG16, Xception, and InceptionV3, and then ensembled the decision of these models using Choquet integral. The fuzzy measures are calculated using Coalition game theory and Lambda fuzzy approximation.

The parameters used for optimizing the DCNN model training procedure are given as follows. For training the MLP classifier using the extracted image descriptors, the Adam optimizer, with 0.001 as its base learning rate and hyperparameters β_1 and β_2 equal to 0.6 and 0.8 respectively, are employed. The learning rate and hyperparameter values are experimentally inferred to be the most optimal values obtained using the Grid search technique for

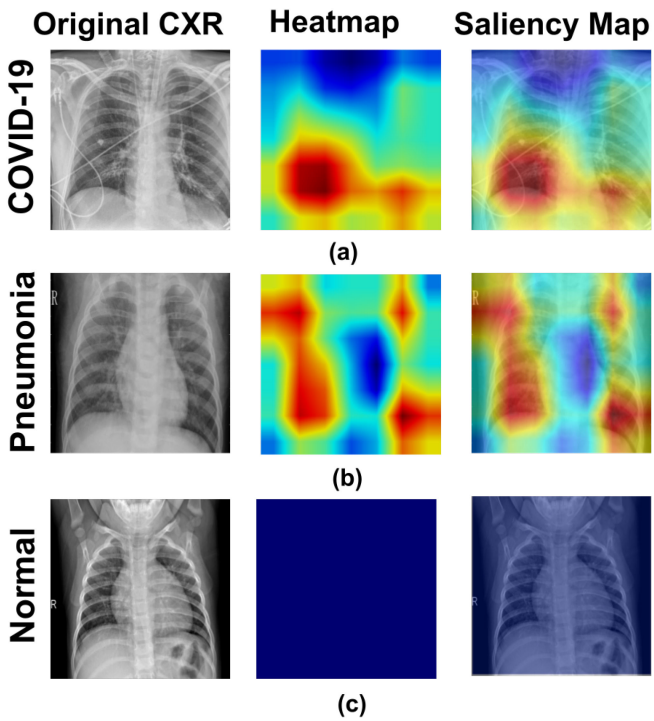

Fig. 4. Saliency maps of (a) COVID-19, (b) Pneumonia, (c) Normal CXR image samples. The first, second, and third columns of the figure illustrate the original CXR image of each class, heatmap, and saliency map generated using RISE approach respectively.

TABLE IV
VALIDATION AND TEST ACCURACIES (%), PRECISION, RECALL, AND AUC FOR THE 3-CLASS CLASSIFICATION OF CXR IMAGES FOR EACH OF THE 3 CLASSIFIERS AND THE ACCURACY OF THE PROPOSED ENSEMBLE METHOD

Cl/En	VA	TA	P (Avg)	R (Avg)	AUC
VGG16	96.71	91.22	0.92	0.92	0.92
Xception	97.02	92.98	0.93	0.93	0.92
Inception V3	97.49	93.48	0.94	0.94	0.94
Choquet Integral (WS-1)	97.74	94.23	0.94	0.94	-
Choquet Integral (WS-2)	98.24	94.23	0.94	0.94	-
Choquet Integral (WS-3)	97.49	93.73	0.95	0.95	-
Ensemble	**98.99**	**95.49**	**0.96**	**0.96**	**0.97**

where Cl stands for Classifier, En stands for Ensemble, VA stands for Validation Accuracy, TA stands for Test Accuracy, P stands for Precision, R stands for Recall, Avg stands for Average, WS-1 is Weighting Scheme-1, WS-2 is Weighting Scheme-2, WS-3 Weighting Scheme-3 and AUC is Area under the Curve.

model tuning and optimization. The batch size is set at 32, and model training is halted at 1000. Subsequently, the best model weights are saved.

In Table IV, we have recorded the validation accuracy, test accuracy, precision, recall, and AUC of each of the three models, and the final results obtained after applying the ensemble methods (both tiers). In Table V it can be seen that the precision of our model is high for COVID-19 lung X-Ray images and so it is clear that our model works properly for the minority class and hence takes care of the unbalanced dataset. In Table VI, we have shown the Shapley value that we have obtained for the three classifiers, namely VGG-16, Xception, and InceptionV3 for the three weighting schemes. It is clear that for all the weighting schemes Inception V3 is assigned a higher value.

TABLE V
PERFORMANCE OF OUR PROPOSED METHOD ON THE NOVEL COVID-19 CHESTXRAY REPOSITORY [30]

Performance Metric	Proposed Method
Accuracy (%)	95.49
Normal Sensitivity	0.975
Pneumonia Sensitivity	0.931
COVID 19 Sensitivity	0.961
Normal (Precision)	0.950
Pneumonia (Precision)	1.000
COVID 19 (Precision)	0.890

TABLE VI
SHAPLEY VALUES CALCULATED OF THE THREE CLASSIFIERS VGG16, XCEPTION AND INCEPTION V3 FOR EACH OF THE WEIGHTING SCHEMES

Subset	S V (WS-1)	SV (WS-2)	S V (WS-3)
VGG16	0.2682	0.1753	0.3962
Xception	0.2855	0.1865	0.4157
Inception V3	0.3027	.1997	0.4345

where SV stands for Shapley Value, WS-1 stands for Weighting Scheme-1, WS-2 stands for Weighting Scheme-2 and WS-3 stands for Weighting Scheme-3

TABLE VII
CALCULATED FUZZY MEASURE VALUES FOR EACH OF THE $2^n - 1$ (IN OUR CASE 7) SUBSETS OF THE N (IN OUR CASE 3) CLASSIFIERS FOR EACH OF THE 3 WEIGHTING SCHEMES

Subset	FM (WS-1)	FM (WS-2)	FM (WS-3)
{1}	0.2682	0.1753	0.3962
{2}	0.2855	0.1865	0.4157
{3}	0.3027	0.1997	0.4346
{1,2}	0.5965	0.4743	0.8119
{1,3}	0.6163	0.4955	0.8307
{2,3}	0.6365	0.5143	0.8502
{1,2,3}	1.0000	1.000	1.0000

where FM stands for Fuzzy measure, WS-1 stands for Weighting Scheme-1, WS-2 stands for Weighting Scheme-2 and WS-3 stands for Weighting Scheme-3

This is intuitive as the accuracy of Inception V3 is higher as compared to those of Xception and VGG16. Values assigned by different weighting schemes can be found to be different. This is because, as stated above, since these weighting schemes focus differently on different intervals of accuracies, they differ in the assignment of the Shapley values and the relative importance of the classifiers. However, this difference is the main reason for the further increase of the accuracy when the results calculated from the three weighting schemes are combined. In Table VII we have shown the fuzzy measures obtained using the three different sets of weights. In this table, we refer to VGG-16 as 1, Xception as 2, and Inception-V3 as 3. Sets like $(1, 2)$ are essentially sets of classifiers like $\{VGG-16, Xception\}$. The values satisfy the monotonic behavior of the fuzzy measures. Besides, more compatible classifiers or sets of classifiers are assigned higher values as is expected by our idea. We give more importance to sets that have more compatible classifiers while calculating the final decision. Moreover, it can be seen that different weighting schemes give a varied relative importance to different subsets of classifiers. This again justifies our reason for a second ensemble

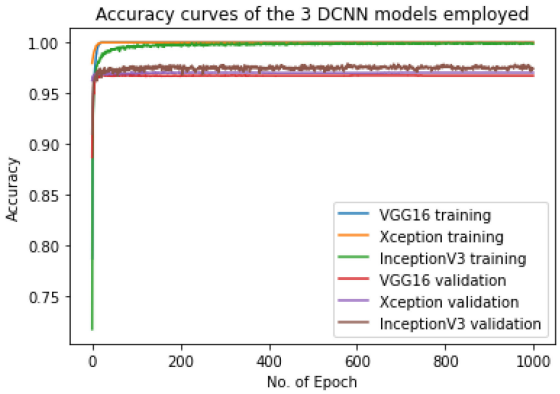

Fig. 5. Training and validation accuracy curves of each DCNN model employed before proposed ensemble method.

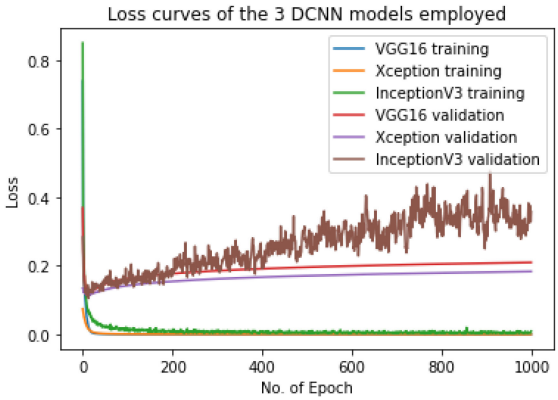

Fig. 6. Training and validation loss curves of each DCNN model employed before proposed ensemble method.

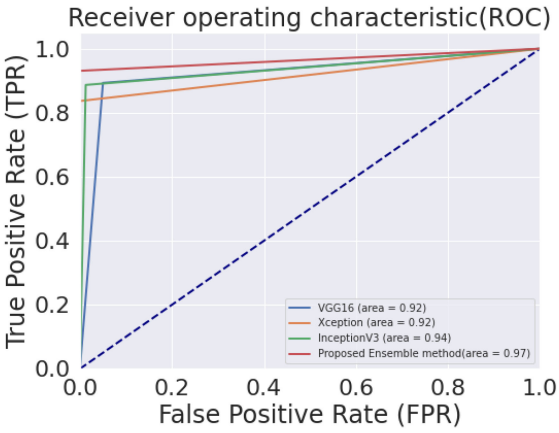

Fig. 7. ROC of the 3 DCNN models and proposed ensemble method.

and selecting three different weights which have different views of the classifiers and hence produces a more generalized result.

We have plotted the accuracy and loss curves of the three DCNN models used, before performing the proposed ensemble method, for both training and validation sets, shown by Fig. 5 and Fig. 6 respectively. In Fig. 7, we have plotted the Receiver operating characteristic (ROC) graph for the 3 classifiers and

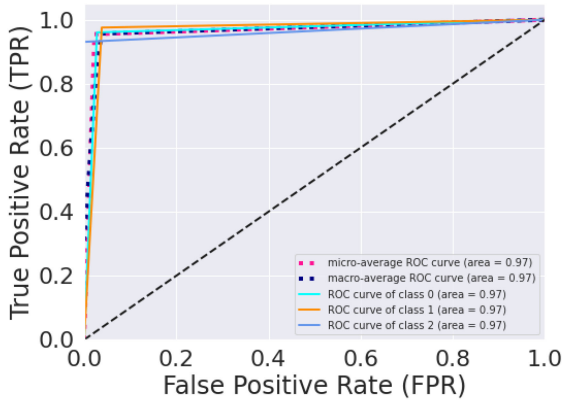

Fig. 8. Multi-labelled ROC of the proposed ensemble method.

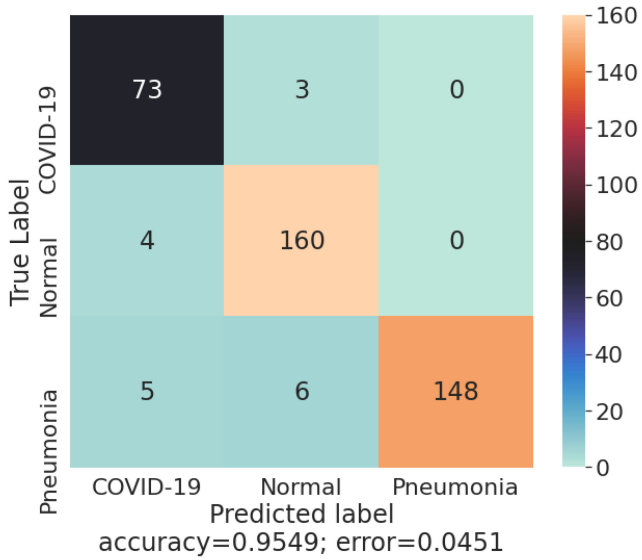

Fig. 9. Confusion matrix of the proposed ensemble method.

the ensemble method. In Fig. 8 we plot the multi-labeled ROC curve for the 3 classifiers and the proposed ensemble method. In Fig. 9 we have plotted the Confusion Matrix for the proposed ensemble method.

From Table IV, Fig. 7 and Fig. 8 it is clear that the Test Accuracy, Precision and Recall all three improve a lot after Fuzzy ensemble. The accuracy increases around 2% from the accuracy of the best performing deep learning model, while it increases by almost 4.5% from the accuracy of the worst performing model.

Again, from Fig. 9 it is clear that our model classifies all Normal CXRs correctly, while it misclassifies only 2 COVID-19 infected cases as Normal Pneumonia and misclassifies only 2 Normal Pneumonia CXRs as Normal. The reason for the first error is that COVID X-Ray images often look similar to that of Normal Pneumonia X-Ray images. The reason for the other set of mistakes is that the patients were having very mild symptoms of Pneumonia and hence the X-Ray looked almost like the X-Ray of a Normal person.

TABLE VIII
COMPARISON OF OUR PROPOSED METHOD WITH THE METHODS REPORTED IN [32] AND [31] ON THE COVIDX DATASET

Performance Metric	Wrok ref. [32]	Work ref. [31]	Proposed Method
Accuracy (%)	93.3	-	93.81
Normal Sensitivity	0.950	0.900	0.923
Pneumonia Sensitivity	0.940	0.930	0.958
COVID 19 Sensitivity	0.910	1.000	1.000
Normal (Precision)	0.913	0.957	0.920
Pneumonia (Precision)	0.938	0.903	0.960
COVID 19 (Precision)	0.889	0.769	0.895

E. Comparison With Other Methods

Since in this classification problem there is no standard dataset that has been used by most of the past researchers, we cannot compare with other methods directly as they each evaluate their method on a different dataset. Thus each method has used its dataset and to compare with each of the methods, we need to use the respective datasets, each of which is different in composition and have different sets of images. We choose one such dataset which has been used by two methods for evaluation or comparison. For the comparison, with the papers [31] and [32] (COVIDNet), we evaluate using the COVIDx dataset. In Table VIII we have depicted the comparison with the said methods.

V. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have proposed an ensemble of deep learning models for the screening of COVID-19 from CXR images. In doing so, we have used Choquet integral for aggregation, which takes into consideration the decisions made by the subset of classifiers along with individual classifiers unlike in other aggregation functions. We have also proposed a novel method for the calculation of the Fuzzy measures using Coalition Game theory (Shapley value), Information Theory, and Lambda fuzzy approximation. As feature extraction, we have used three pre-trained DCNN models namely, VGG-16, Xception, and Inception V3 while for the classification task we have used MLP.

In this paper, we have considered that all the classification models are important for classification to some extent. We may consider a number of different classifiers all of which may not be important, but only a subset of them may provide valuable information. However, the basic consideration of Shapley value is that the players (here classifiers) form a *Grand Coalition* even if some of the classifiers do not convey useful information. Thus the selection of useful classifiers from a set of different classifiers which may or may not convey useful information must be done based on experimentation. This can be considered as a shortcoming of the proposed method and we intend to take care of this problem in our future work.

VI. CODE AVAILABILITY

The python code implementation is publicly available at https://github.com/subhankar01/Covid-Chestxray-lambda-fuzzy.

VII. DATA AVAILABILITY

The proposed Novel COVID-19 Chestxray Repository can be accessed at https://www.kaggle.com/subhankarsen/novel-covid19-chestxray-repository.

ACKNOWLEDGMENT

Source code of this work is available at https://github.com/subhankar01/Covid-Chestxray-lambda-fuzzy.

REFERENCES

[1] J. Born *et al.*, "POCOVID-Net: Automatic detection of COVID-19 from a new lung ultrasound imaging dataset (POCUS)," *arXiv:2004.12084*, 2020.
[2] B. Sahiner *et al.*, "Deep learning in medical imaging and radiation therapy," *Med. Phys.*, vol. 46, no. 1, pp. e1–e36, 2019
[3] A. Zahedi *et al.*, "Deep analysis of mitochondria and cell health using machine learning," *Sci. Rep.*, vol. 8, no. 1, pp. 1–15, 2018.
[4] R. Poplin *et al.*, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nature Biomed. Eng.*, vol. 2, no. 3, pp. 158–164, 2018.
[5] R. Adalarasan and R. Malathi, "Automatic detection of blood vessels in digital retinal images using soft computing technique," *Materials Today: Proc.*, vol. 5, no. 1, pp. 1950–1959, 2018.
[6] M. Marz *et al.*, "Challenges in RNA virus bioinformatics," *Bioinformatics*, vol. 30, no. 13, pp. 1793–1799, 2014.
[7] V. Petsiuk, A. Das, and K. Saenko, "Rise: Randomized input samplingfor explanation of black-box models," *arXiv:1806.07421*, 2018.
[8] M. Z. Islam, M. M. Islam, and A. Asraf, "A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images," *Informat. Med. Unlocked*, vol. 20, 2020, Art. no. 100412.
[9] M. Turkoglu, "COVIDetectioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble," *Appl. Intell.*, vol. 51, no. 3, pp. 1213–1226, Mar. 2021.
[10] A. T. Sahlol, D. Yousri, A. A. Ewers, M. A. Al-Qaness, R. Dama-Services, and M. Abd Elaziz, "COVID-19 image classification using deep features and fractional-order marine predators algorithm," *Sci. Rep.*, vol. 10, no. 1, pp. 1–15, 2020.
[11] P. Khosla *et al.*, "Supervised contrastive learning," *arXiv:2004.11362*, 2020.
[12] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *Proc. Int. Conf. Learn. Representations,* vol. 10, p. 3, 2017.
[13] F. Ucar and D. Korkmaz, "COVIDiagnosis-Net: Deep bayes-squeeze net-based diagnostic of the coronavirus disease 2019 (COVID-19) from X-ray images," *Med. Hypotheses*, 2020, Art. no. 109761.
[14] S. Roy *et al.* "Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2676–2687, Aug. 2020.
[15] J. Too and S. Mirjalili, "A hyper learning binary dragonfly algorithm for feature selection: A COVID-19 case study," *Knowl.-Based Syst.*, vol. 212, 2020, Art. no. 106553.
[16] M. A. Elaziz, K. M. Hosny, A. Salah, M. M. Darwish, S. Lu, and A. T. Sahlol, "New machine learning method for image-based diagnosis of COVID-19," *PLoS One*, vol. 15, no. 6, 2020, Art. no. e0235187.
[17] P. Remy, *Keract: A Library for Visualizing Activations and Gradients.* [Online]. Available: https://github.com/philipperemy/keract
[18] R. Babukarthik, V. A. K. Adiga, G. Sambasivam, D. Chandramohan, and J. Amudhavel, "Prediction of COVID-19 using genetic deep learning convolutional neural network (GDCNN)," *IEEE Access*, vol. 8, pp. 177647–177666, 2020.
[19] O. Russakovsky *et al.* "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[21] F. Chollet, "Xception: Deep Learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1251–1258.
[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.
[23] R. Mesiar, "Generalizations of K-order additive discrete fuzzy measures," *Fuzzy Sets Syst.,* vol. 102, pp. 423–428, 1999.
[24] G. Beliakov, S. James, and J. Wu, *Discrete Fuzzy Measures*. Springer, International Publishing, 2020.
[25] K. Lee and H. LeeKwang, "Identification of λ-fuzzy measure by genetic algorithms," *Fuzzy Sets Syst.*, vol. 75, pp. 301–309, 1995.
[26] C. Li, G. Zeng-Tai, and D. Gang, "Genetic algorithm optimization for determining fuzzy measures from fuzzy data," *J. Appl. Math.* vol. 2013, 2013, Art. no. 542153.

[27] A. Tehrani, W. Cheng, K. Dembczyński, and E. Hüllermeier, "Learning monotone nonlinear models using the Choquet integral," *Mach. Learn.*, vol. 89, pp. 183–211, 2020.
[28] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure," *Fuzzy Sets Syst.*, vol. 29, no. 2, pp. 201–227, 1989.
[29] M. E. Chowdhury *et al.*, "Can AI help in screening viral and COVID-19 pneumonia?," *arXiv:2003.13145*, 2020.
[30] S. Sen, P. Bhowal, R. Sarkar, J. H. Yoon, Z. W. Geem, *Novel COVID-19 Chest Xray Repository*. Accessed: 2 Feb. 2021. [Online]. Available: https://www.kaggle.com/subhankarsen/novel-COVID19-chestxray-repository
[31] Y. Oh, S. Park, and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2688–2700, Aug. 2020.
[32] L. Wang, Z. Q. Lin, and A. Wong, "Covid-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020
[33] M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, M. T. Islam, *COVID-19 Radiography Database*. Accessed: 26 Feb. 2021. [Online]. Available: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database
[34] J. P. Cohen, P. Morrison, L. Dao, and K. Roth, T. Q. Duong, and M. Ghassemi, *COVID-Chest Xray Set*. Accessed: 26 Feb. 2021. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset
[35] L. Wang, A. Wong, and A. Chung, *Actualmed COVID-19 Chest X-Ray Dataset Initiative*. Accessed: 26 Feb. 2021. [Online]. Available: https://github.com/agchung/Actualmed-COVID-chestxray-dataset
[36] *ImageNet dataset*. Accessed: 26 Feb. 2021. [Online]. Available: http://www.image-net.org/
[37] Keras Applications. Accessed: 26 Feb. 2021. [Online]. Available: https://keras.io/api/applications/
[38] Z. Wang, L. Zhao, and Q. Dou, "Contrastive cross-site learning with redesigned net for COVID-19 CT classification," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2806–2813, Sep. 2020.
[39] S. Liu and C. Kao, "Fuzzy measures for correlation coefficient of fuzzy numbers," *Fuzzy Sets Syst.,* vol. 128, pp. 267–275, 2002.
[40] G. Beliakov and J. Wu, "Learning fuzzy measures from data: Simplifications and optimization strategies," *Inf. Sci.*, vol. 494, pp. 100–113, 2019.
[41] M. Grabisch, "A new algorithm for identifying fuzzy measures and its application to pattern recognition," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 1995, vol. 1, pp. 145–150.
[42] J. Murillo, J. Guillaume, E. Tapia, and P. Bulacio, "Revised HLMS: A useful algorithm for fuzzy measure identification," *Inf. Fusion*, vol. 14, pp. 532–540, 2013.
[43] J. Zhang *et al.*, "Viral pneumonia screening on chest X-ray images using confidence-aware anomaly detection," *arXiv:2003.12338*, 2020.
[44] K. Panetta, F. Sanghavi, S. Agaian, and N. Madan, "Automated detection of COVID-19 cases on radiographs using shape-dependent fibonacci-p patterns," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 1852–1863, Jun. 2021.
[45] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," *Appl. Intell.*, Springer, vol. 51, pp. 854–864, 2021.
[46] Y. Song *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, Mar. 2021.
[47] R. Kundu, H. Basak, P. K. Singh, A. Ahmadian, M. Ferrara, and R. Sarkar "Fuzzy rank-based fusion of CNN models using Gompertz function for screening COVID-19 CT-scans," *Sci. Rep,* vol. 11, 2021, Art. no. 14133.
[48] S. Dey, R. Bhattacharya, S. Malakar, S. Mirjalili, and R. Sarkar, "Choquet fuzzy integral-based classifier ensemble technique for COVID-19 detection," *Comput. Biol. Med.*, Jun. 2021.
[49] Y. Karbhari, A. Basu, Z. W. Geem, G. T. Han, and R. Sarkar, "Generation of synthetic chest X-ray images and detection of COVID-19: A deep learning based approach," *Diagnostics*, vol. 11, no. 5, p. 895, May 2021.
[50] S. Das, S. D. Roy, S. Malakar, J. D. Velasquez, and R. Sarkar, "Bi-level prediction model for screening COVID-19 from chest X-ray images," *Big Data Res.*, vol. 25, 2021, Art. no. 100233.
[51] A. Garain, A. Basu, F. Giampaolo, J. D. Velasquez, and R. Sarkar, "Detection of COVID-19 from CT scan images: A spiking neural network-based approach," *Neural Computing and Applications*, Apr. 16 2021.
[52] S. Chattopadhyay, A. Dey, P. K. Singh, Z. W. Geem, and R. Sarkar, "COVID-19 detection by optimizing deep residual features with improved clustering-based golden ratio optimizer," *Diagnostics*, vol. 11, no. 2, p. 315, Feb. 2021.
[53] S. Sen, S. Saha, S. Chatterjee, S. Mirjalili, and R. Sarkar, "A bi-stage feature selection approach for COVID-19 prediction using chest CT images," *Applied Intelligence*, Apr. 2021.