Efficient DNN Ensemble for Pneumonia Detection in Chest X-ray Images

V S Suryaa, Arockia Xavier Annie R, Aiswarya M S
Department of Computer Science and Engineering
College of Engineering, Guindy Anna University, Chennai, India

Abstract—**Pneumonia is a disease caused by a variety of organisms, including bacteria, viruses, and fungi, which could be fatal if timely medical care is not provided. According to the World Health Organization (WHO) report, the most common diagnosis for severe COVID-19 is severe pneumonia. The most common method of detecting Pneumonia is through chest X-ray which is a very time intensive process and requires a skilled expert. The rapid development in the field of deep learning and neural networks in recent years has led to drastic improvement in automation of pneumonia detection from analysing chest x-rays. In this paper, a pre-trained Convolutional Neural Networks (CNN) on chest x-ray images is used as feature extractors which are then further processed to classify the images in order to predict whether a person has pneumonia or not. The different pre- trained Convolutional Neural Networks used are assessed with various parameters regarding their predictions on the images. The results of pre-trained neural networks were examined, and an ensemble model was proposed that combines the predictions of the best pre-trained models to produce better results than individual models.**

Keywords—Deep neural networks; ensemble learning; pneumonia detection using x-ray images; transfer learning

I. INTRODUCTION

Pneumonia is an infection in one or both lungs. Bacteria, viruses, and fungi cause it. The infection causes inflammation in the alveoli. The alveoli fill with fluid or pus, causing difficulty in breathing. Pneumonia and lower respiratory tract infections like influenza and respiratory syncytial virus are the leading cause of death worldwide. WHO Child Health Epidemiology Reference Group reported that the median global incidence of clinical pneumonia is 0.28 episodes per child- year. This statistic converts to an annual incidence of 150.7 million new cases, of which 11-20 million (7-13%) are severe enough to require hospital admission [1]. Majority of the episodes of clinical pneumonia in young children worldwide occur in developing countries due to lack of proper timely diagnosis. In 2015, More than half of all global pneumonia cases were from India, Nigeria, Indonesia, Pakistan, and China alone [2]. Chest X-rays are used by radiologists to identify pneumonia among patients. They look for white spots in the lungs (called infiltrates) that identify an infection. Medical imaging accounts for more than 90% of the entire available medical. Radiologists are required to analyse large quantities of medical images which is a time-consuming and exhaustive process. With the development of deep learning methods, it would be possible to sift through the data and analyse medical exams more efficiently.

Machine learning algorithms like Logistic Regression, Support Vector Machines (SVM) do not learn any hidden representation in the images and directly use the image data provided. On the contrary, deep learning in computer vision has shown great success in decoding hidden representations and extracting features from them with the help of Convolutional Neural Networks. Convolutional Neural Networks (CNN) can extract and process data at very high speeds. The recent advancement in deep learning frameworks [3] have enabled faster and more accurate detection, while the increased CPU and GPU processing power available allows radiologists to improve their diagnostic efficiency.

In this work, pre-trained CNN architectures proposed in the past few years taken into consideration and try to assess their prediction on various parameters to identify the ideal one. All architectures used in the paper are CNNs which were pre-trained on ImageNet dataset previously. The CNNs were fine-tuned with the pneumonia chest x-ray dataset which was then used for feature extraction. The CNNs were connected with a common fully connected layer to assess their predictions from extracted features.

Lower image resolution helps lower training time on very large dataset and also decreases computational needs and device capacities due to relatively lesser number of parameters. With the advent of technology, it is easier to transfer image files over the phones which could be taken to advantage if the image is of lower resolution as it decreases the file size drastically. This also helps storing a large database of X-ray images over a device without much hassle.

The related works section discusses the various pre-trained networks. The third section discusses the strategies and hyper parameters considered for the comparison of these deep neural networks. The proposed ensemble model is discussed in the fifth section and it is followed by result analysis in the next section and finally the sixth section concludes.

II. RELATED WORK

In recent years, there has been significant research in automation of Pneumonia detection through deep learning and neural networks which has yielded impressive results.

The study in [4] contributed a voting ensemble (AlexNet, ResNet18, InceptionV3, DenseNet-121 and GoogleNet) classification approach to Pneumonia detection. In [5] proposes a customized VGG16 model for the detection with an accuracy of 96.2%. The work done in [6] asserts that DenseNet201

architecture outperforms AlexNet, ResNet18 and SqueezeNet with an accuracy of 98% and AUC score of 0.98.

Kermany et al. (2018) in [7] used InceptionV3 pre-trained model for classification. They have also worked on bacterial and viral pneumonia classification in addition to normal and pneumonia. Saraiva et al. (2019) in [8] used cross validation technique in a customized 4 fully connected layer neural network to obtain an accuracy of 94.4% and compared it with MLP network.

D.Varshini et al (2019) [9] used different pre-trained CNN feature extractors combined with classifiers SVM with RBF kernel, Naive Bayes, k-nearest neighbours and Random Forests. They conclude that DenseNet-169 for the feature extraction stage and SVM for the classification stage provide statistically better results. Quan et al (2021) [10] proposed DenseCapsNet, a capsule network for detection of Covid-19 from X-ray images. DenseCapsNet reached 90.7 % accuracy and 96% sensitivity in test set results.

A. Transfer Learning

Transfer learning is a machine learning method that involves using an already developed model on a new problem task as the initial model in [11]. It is a popular approach used in computer vision and other arduous deep learning tasks.

When transfer learning is used, the model parameters start with good initial values due to the previous training and do not require huge modifications to be better adapted to the new task. The pre-trained model weights are treated as the initial values for the new task in hand, and updates are performed on that during training [12].

In proposed work, due to limited availability a new classifier is fitted on the top layers and fine tune only the last few convolutional layers in the model and use that for feature extraction purposes. The performance of some famous pretrained networks such as ResNet-50, ResNet-101 ResNet-152, VGG- 16, VGG-19, MobileNetV2 and DenseNet-201 also evaluated.

B. VGG Architecture

VGG16 is a convolutional neural network model proposed in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" in [13]. The model achieves 92.7% top-5 test accuracy in ImageNet dataset. The model was submitted to ILSVRC-2014 and gained huge popularity. It makes improvements over its previous works by replacing large kernel-sized filters with multiple 3×3 kernel-sized filters one after another. The network is characterized by stacking 3x3 convolutional layers on top of each other to increase network depth. The convolution layers increase the volume size which is handled by max-pooling. Max-pooling is performed over a 2×2-pixel window, with stride 2.

The VGG-19 Neural Network is defined as that type of neural network which is also specifically trained on more than a million images from the ImageNet database, but the difference between the VGG-19 Neural Network and the VGG-16 Neural Network is its network layer depth level as 19 and 16, respectively.

C. ResNet Architecture

The main idea behind ResNet is adding more layers without degrading the performance of the whole network due to vanishing gradient problems. Vanishing gradient problem was caused due to repeated application of chain rule during back propagation which made the gradient too small and eventually disappears. This led to no actual learning in the networks.

ResNet introduced, "Identity Shortcut Connection" that skips one or more layers [14]. It stacks up identity mappings which are initially skipped and the activations from the previous layers are used instead. This enables faster learning in the compressed network. Later when the network trains again, the skipped layers are included to understand the feature space. The main difference between ResNet-50, ResNet-101 and ResNet-152 are the number of layers in them. ResNet-50 has 3.8×10^9 floating point operations in total compared to 7.6×10^9 in ResNet-101 and 11.3×10^9 in ResNet-152.

D. DenseNet Architecture

The increasing depth of convolutional neural networks caused a problem of vanishing information about the input or gradient when passing through many layers. In order to solve this, authors introduced architecture [15] with a simple connectivity pattern to ensure the maximum flow of information between layers both in forward computation as well as in backward gradients computation. Each layer in the network adds its own feature-maps to the input received from its previous layers, which are then passed on to the subsequent layers in the network.

In DenseNet, H_i (*i* refers to the layer index) is a composite function of operations like ReLU, pooling, convolution and batch normalization. Each layer implements H_i (x_0, x_1, x_2, ..., x_{i-1}) where, [x_0, x_1, x_2, ..., x_{i-1}] refers to concatenation of the feature-maps produced in layers 0 to i-1. Variable size of feature-maps does not allow concatenation operation. Down-sampling layers are the most important aspect of convolutional neural network. To facilitate the downsampling in the architecture, the complete architecture has been divided into multiple densely connected dense blocks. The layers between dense blocks are transition layers which perform convolution and pooling. Convolution and pooling operations are performed in transition layers between the dense blocks.

The main advantages of DenseNet are decreasing vanishing gradient issue, improving feature propagation, both forward and backward, increase feature reuse and reducing the number of parameters.

E. MobileNetV2 Architecture

MobileNetV2 has three convolutional layers in a block: the first is a 1x1 convolution called the expansion layer. The main purpose of the expansion layer is to increase the number of channels in the data before it is passed on to the next layer. The expansion factor is 6 by default. The next layer is the depth wise convolution which is used to filter inputs. Finally, a 1x1 projection layer is used which projects data with higher dimension into a tensor with lower dimension and it reduces the amount of data that flows through the network.
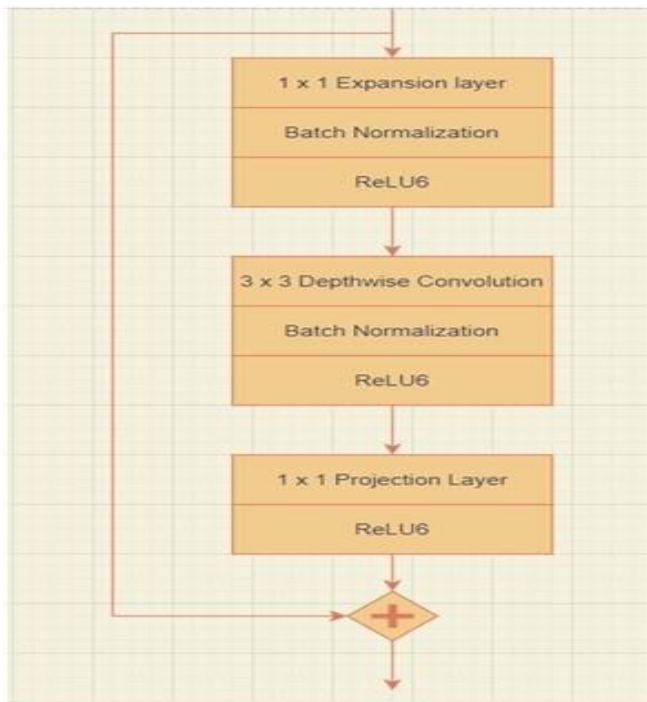
Fig. 1.   MobileNetV2 Architecture Description.

As observed in Fig. 1, the architecture introduces residual connection which helps the flow of gradients through the networks. All the layers have batch normalization and ReLU6 activation function (except for projection layer). The authors in [16] report that due to low dimensional data produced in the layer, using a non-linearity affects the information obtained.

The dataset used is from Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification" released by Daniel et al. in 2018 published on the Kaggle platform [17], which consists of 5,863 X-Ray images. Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou.

For this work, following the approaches from the past, the labels are treated as ground truth for the purpose of pneumonia detection. Out of the complete Radiographic image dataset, 4608 images were used as a training set which contains 1115 normal X-ray images and 3493 images labeled as Pneumonia. The test data and validation data both contains 624 X-ray images with 234 normal X-ray images and 390 images labeled as Pneumonia. The images were downscaled to 184 x 184 size for computational purposes before using it for analysis.

The training of all the models was done using a computer with 16 GB RAM and Nvidia Tesla P100 GPU. The figures shows sample chest X-ray images taken from dataset used for training, where, Fig. 2 is a chest X-ray of a normal patient and Fig. 3 is a chest X-ray of a patient affected by Pneumonia.

F. Pre-processing Stage

The images were scaled down to 3 channel 184 x 184 resolution and data normalization techniques were deployed to improve computational efficiency. The values were scaled between 0 and 1 with random zooming on the images for image augmentation [18].

G. Feature Extraction and Classification

The pre-trained neural networks considered for feature extraction in the experiment are ResNet-50, ResNet-101 ResNet-152, VGG-16, VGG-19, MobileNetV2 and DenseNet-201.

Further on the features extracted are classified using a fully connected layer fitted to the network. All the networks were fine-tuned by freezing earlier layers and unfreezing the rest for the network to adapt well to the dataset considered for the experiment.

1) Resnet pre-trained networks: All the Resnet networks considered for the experiment- ResNet50, ResNet101 and Resnet152 were networks pre-trained on the ImageNet dataset and were fed in with input of 184 x 184 x 3 chest x-ray images. All three had been fit with a fully connected layer for classification. The fully connected layer consists of a dense layer of 2048 units and another dense layer of 64 units with ReLU activation function, dropouts and batch-normalization.

2) VGG pre-trained networks: Both VGG-16 and VGG-19 taken for consideration were pre-trained networks on the ImageNet dataset. The output of the network without top layers was further classified with a fully connected layer containing two dense layers of 4096 units with ReLU activation. Then the fine-tuned network shown in Fig. 4, was later trained on the dataset to obtain the results on the test data

3) DenseNet201 pre-trained network: The feature extracted from the ImageNet pre-trained network was subjected to global average-pooling and fed into a fully connected layer consisting of 1024 unit dense layer and 512 units dense layer with ReLU activation function which was then classified with a Softmax function.

4) MobileNetV2 pre-trained network: The ImageNet pre-trained network was used for feature extraction which was then fed into a fully connected layer after global average-pooling with dense layers of 4096 units and 512 units and ReLU activation function. The output of that was classified with a Softmax function.

Fig. 2.   Chest X-Ray of a Normal Patient.

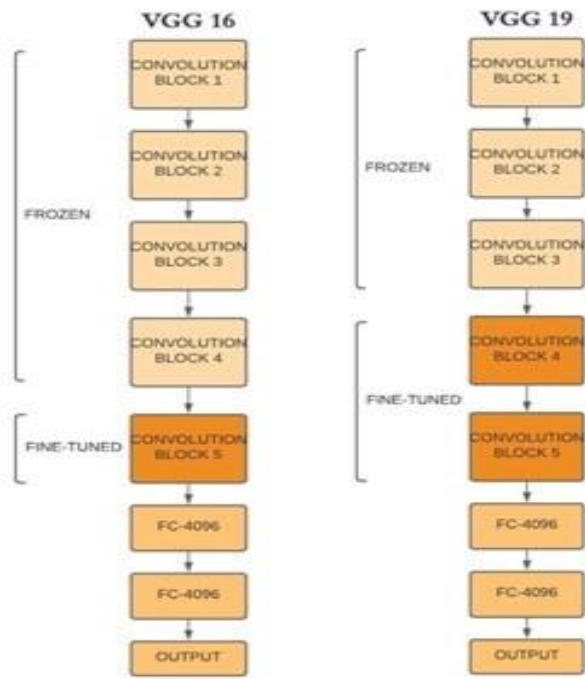Fig. 3.   Chest X-Ray of a Patient affected by Pneumonia.

Fig. 4. Transfer Learning Approach through Fine Tuning during Training for VGGNet.

III. STRATEGIES AND HYPER-PARAMETERS OF THE NETWORK

A. Learning Rate

The learning rate controls the ease of the model to adapt to the problem [19]. Smaller learning rates require more epochs during training given the smaller changes made to the weights each update as mentioned in Equation 1, whereas larger learning rates result in rapid changes and require fewer training epochs.

Large learning rates can cause the model to converge too quickly to a non-optimal solution or might lead to exploding gradient, whereas a small learning rate can cause the training to get stuck or lead to vanishing gradient issue.

$$w_i' = w_i - \gamma \frac{\partial L}{\partial w} \tag{1}$$

where

γ - Learning rate.

L - Loss function.

w_i – weights.

The equation shows as to how the weight gets updated. The learning rate used in training is 0.001, generally the learning rate ranges from 0.00001 to 1.

B. Optimization Algorithm

Batch gradient descent recomputes gradients for similar samples before each parameter update which leads to drastic increase in redundant computations for large datasets. Stochastic Gradient Descent [20] does away with this redundancy. Stochastic Gradient Descent frequently updates with a high variance that cause the objective function to fluctuate heavily, on the other hand, it allows it to find better local minima. With SGD estimating loss function using a small batch it potentially may not lead us in an optimal direction. Hence, usage of exponentially weighted averages can help us with a better estimate which is closer to the actual derivative. This explains the better performance of SGD with momentum than the classic SGD. SGD tends to find it difficult to find the optimum in cases of ravines (area where surface curves much more steeply in one dimension than other) and oscillates around them. This issue is overcome with momentum as it accelerates gradients [21] in the optimal direction. Hence, Stochastic Gradient Descent is used with momentum. This is shown in Equation 2 and 3.

$$V_t = \beta V_{t-1} + (1 - \beta)\nabla_w L(w, x, y) \tag{2}$$

$$w = w - \alpha V_t \tag{3}$$

where,

β - Coefficient of momentum.

L - Loss function.

V - Retrained gradient.

∇ - Learning rate.

w – weight.

x - Feature vector.

y – Output.

Using lower values of momentum implies averaging over much lesser values and more than 0.9 averages over a large data previously encountered. Hence, 0.9 is generally said to be a good estimate and is used in the experiments.

C. Dropout

Dropout involves removing some nodes so that the neural network does not overfit and can be implemented during the training process [22]. This enables the network to understand and distinguish redundant features. For each training stage, each node can be selected with probability P or drop it with probability (1-P). A dropout of 0.3 is applied on both the hidden layer units in the fully connected layers.

D. Batch Size

The batch size defines the number of samples that will be propagated through the network in one forward and backward propagation. Networks train faster with mini batches because the weights get updated after each propagation. It requires less memory. Since, the network is trained using fewer samples, the overall training procedure requires less memory. That's especially important when using large image datasets for training. A batch size of 32 is used for this work.

E. Activation Function

An activation function is a function that is added into an artificial neural network in order to help the network learn complex patterns in the data, it takes in the output signal from the previous cell and converts it into some form that can be taken as input to the next cell. It is like adding non-linear layers in between linear layers because non-linearity is required for

the network to understand complex data. In the fully connected layers, ReLU activation [23] is applied for the hidden dense layers and softmax function [24] in the prediction layer. Softmax is used in multiclass classification problems, as in equation 4.

$$\sigma(\overrightarrow{z_i}) = \frac{e^{z_j}}{\sum_{j=1}^{K} e^{z_i}} \tag{4}$$

where,

σ - softmax.

e^{z_j} - standard exponential function for output vector.

z - input vector.

K - no. of classes in multi-class classifier.

z_i values are the elements of the input vector and can take any real value. The denominator is the normalization term used so that the summation of the output values results in 1 and a valid probability distribution is maintained. ReLU (Rectified Linear Unit) is a type of activation function. ReLU is defined as y=max(0, x) in mathematical terms and the graphical representation is provided in Fig. 5. It is most commonly used activation function in convolutional neural network.

F. Loss Function

The loss function used in the experiment is sparse categorical cross-entropy. Cross-entropy is defined as the measure of the difference between two probability distributions for a given random variable or set of events.

A skewed probability distribution has lower entropy than a balanced probability distribution. In this work, sparse

categorical cross- entropy is considered because the classes are mutually exclusive from each other unlike the categorical cross-entropy which is used in the case of samples having multiple classes or soft probabilities.

G. Inference from Individual Models

The classification accuracy for each model has been plotted on a graph with respect to the number of epochs used for training. Both validation and training set accuracy has been plotted for each epoch. This helps us to see how the model has been fitting input data. The model loss for each of the models has been plotted for both validation and training set for every epoch.

Loss and Accuracy from the various Network models are tested and depicted as graphs in Fig. 6(a – g).The network models evaluated are ResNet-50, ResNet-101, ResNet-152, VGG-16, VGG-19, DenseNet-201 and MobileNet-V, respectively.
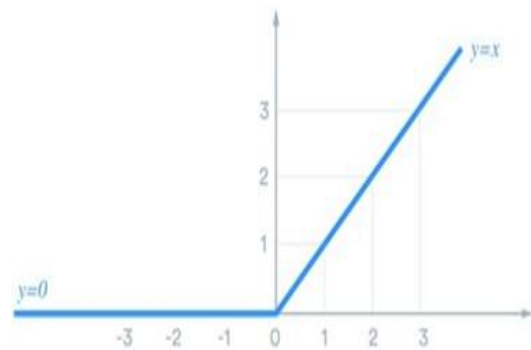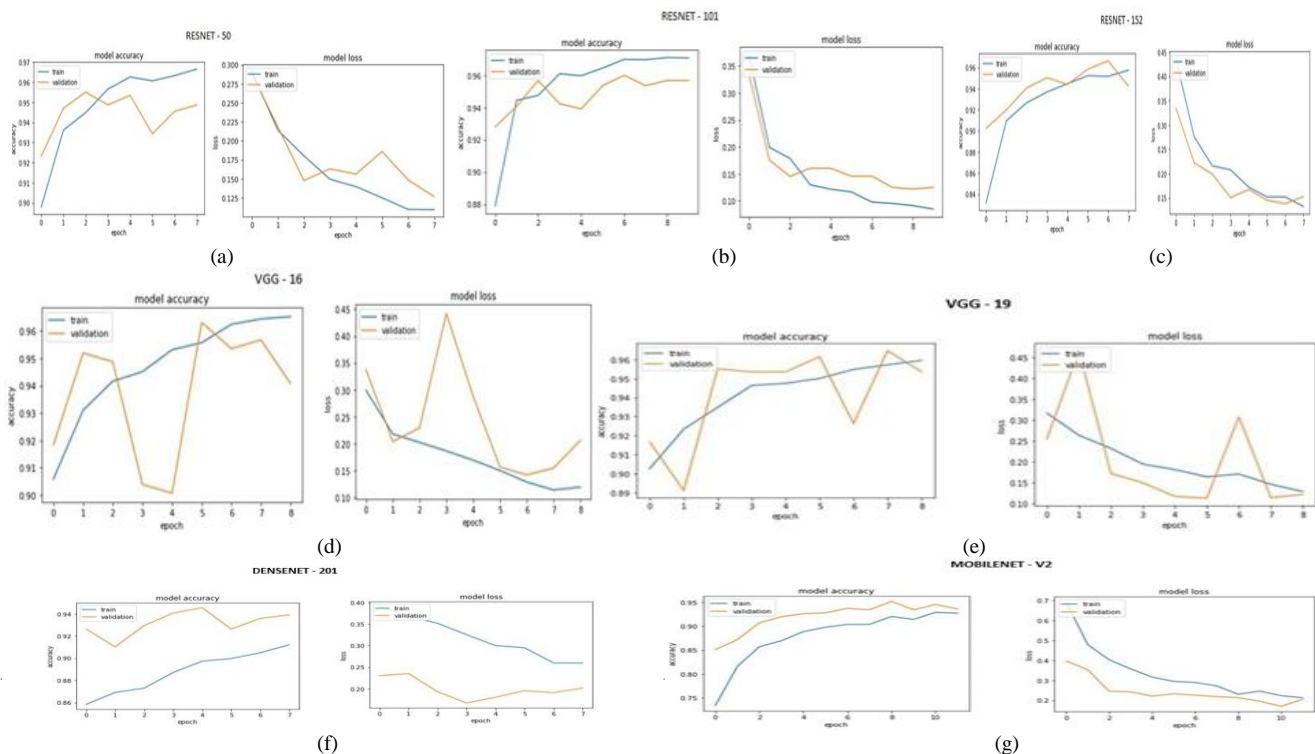

Fig. 5. ReLU Function.

(a)   (b)   (c)

(d)   (e)

(f)   (g)

Fig. 6. Loss and Accuracy from the Various Network Models (a) ResNet-50 (b) Resnet-101 (c) Resnet-152 (d) VGG-16 (e) VGG-19. (f) DenseNet-201 (g) MobileNet V2.

IV. ENSEMBLE NETWORK MODEL

The ensemble model works on combining predictions and decisions of different models to augment the overall performance and generalization of the models. The main idea is to reduce noise, bias and variance in different models. The ensemble model can employ simple methods like mode and weighted average or resort to advanced techniques like bagging and boosting depending on the requirements and constraints for the task in hand. Although it drastically increases the complexity of the model and the design time, it generally improves on the accuracy, stability and robustness of the model.

Weighted average predictions involving the multiple models are employed to obtain an ensemble model in the experiment. Ensemble model of which is proposed has the following models:

a) VGG-16

b) ResNet-50

c) ResNet-101

d) ResNet-152

e) VGG-19

f) DenseNet-201

g) MobileNet-V2

As shown in Fig. 7, the trained models are loaded with their respective weights and the prediction on the test set is obtained for all the models.

Then the model predictions are combined through weighted average, with 30% weightage VGG16 and 20% for VGG19 as they have a relatively higher accuracy and AUC score compared to others. All other models are given 10% weightage for computation of ensemble prediction value.

The proposed model has an accuracy of 95.03% and an AUC score of 0.9441 which is higher than any model considered by a good margin. The ROC curves used to compute the AUC scores have been compiled in Fig. 8. ROC graph [25] is a plot with the false positive rate on the X axis and the true positive rate on the Y axis. It is a visual approach for analysing the trade-off between the ability of a classifier to correctly identify positive cases and the number of negative cases that are incorrectly classified. ROC graph captures all information contained in the confusion matrix, since, FN is the complement of TP and TN is the complement of FP. Fig. 8 shows the different ROC curves for all the models.
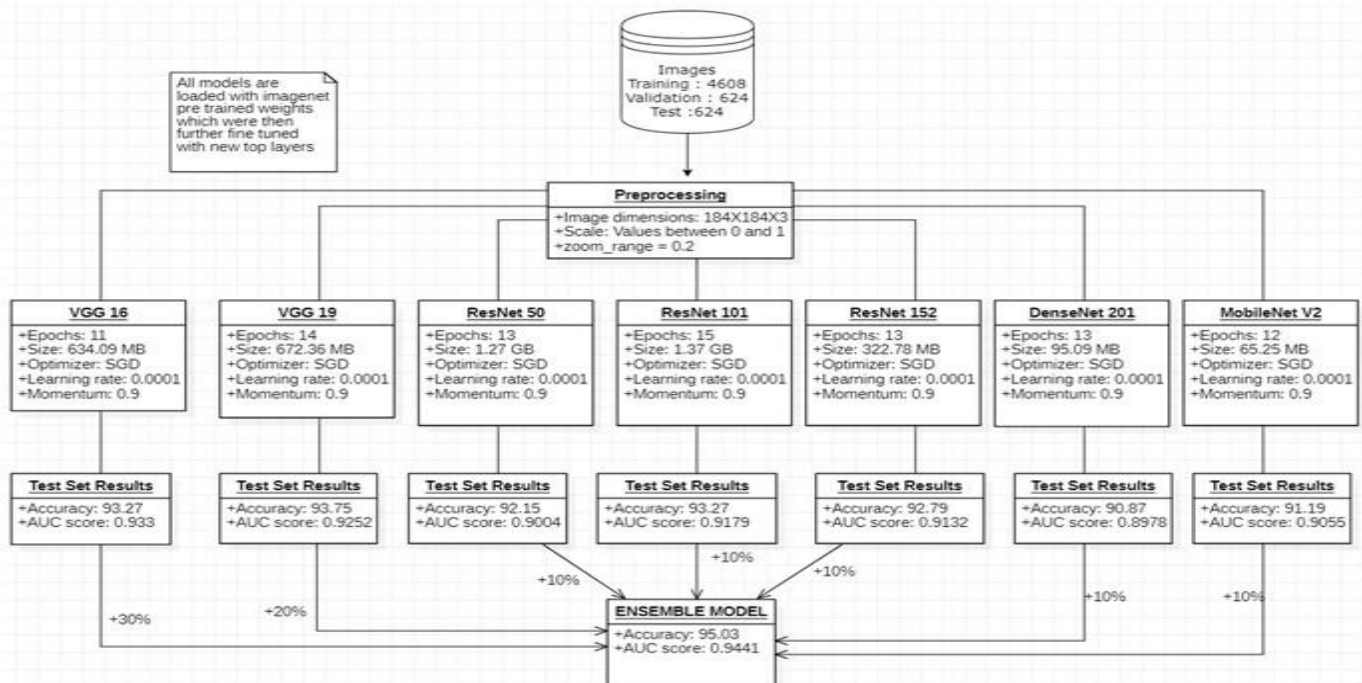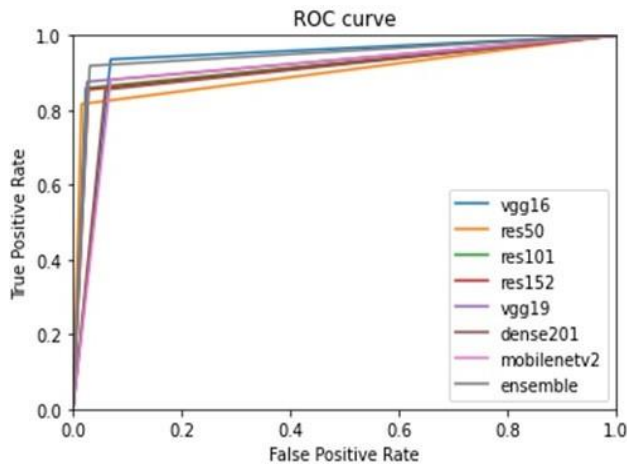

Fig. 7.   Proposed Ensemble Model.

Fig. 8. ROC Curve Comparison of all the Individual Models with Ensemble Model.

V. RESULT ANALYSIS

In this section, a comparison of the ensemble model and other models are done with different metrics. The following sections gives the performance metrics used, and provides discuss on how the Ensemble model provides better results than other models.

A. Performance Metrics for Evaluation

The Results have the study on Accuracy impact with various models as described in [26]. The various metrics used are as follows:

Confusion matrix, Classification accuracy, Precision, Sensitivity and Specificity, F1-Score and Area Under Curve AUC - ROC curve.

B. Result and Discussion

Table I shows that the ensemble model does better than all the models considering the different metrics of evaluation used and the ensemble network has an accuracy of 95.03 and AUC score of 94.41. The ensemble model is based on weighted voting, with each model's output being assigned a different weightage. The results assert that the combination of different results with appropriate weightage improves the overall prediction significantly. This can be attributed to the tendency of a particular model being biased towards a particular region of interest or feature. The bias can be avoided in an ensemble model. Moreover, it aids the evaluation by accounting different niche aspects of the image.

Different models have been trained in accordance to the validation result analysis using early stopping technique as observed in Fig. 6. Among individual models, VGG-16 and VGG-19 have comparatively better results with accuracy of 93.27% and 93.75%, respectively, and with an AUC score of 93.30 and 92.52 for VGG16 and VGG19. AUC score is generally used for binary classification models and is taken as the main parameter to assess the models because of the class imbalance between normal and pneumonia affected X-ray images. AUC score is not in consistent with the accuracy which avoids evaluating the model from a narrow perspective.

TABLE I. COMPARISON OF METRICS OF ENSEMBLE MODEL WITH OTHER INDIVIDUAL NETWORK

Model	Accuracy	Sensitivity	Specificity	Precision	F1score	AUC
DenseNet-201	90.87	91.52	89.69	94.1	0.927	89.78
ResNet-50	92.15	89.93	96.95	98.46	0.94	90.04
MobileNet-V2	91.19	92.84	88.41	93.08	0.929	90.55
ResNet-152	92.79	91.77	94.79	97.18	0.944	91.32
ResNet-101	93.27	92.03	95.71	97.69	0.947	91.79
VGG-19	93.75	92.91	95.35	97.44	0.951	92.52
VGG-16	93.27	96.03	89.02	93.08	0.946	93.3
Ensemble	95.03	95.21	94.71	96.92	0.961	94.41

TABLE II. COMPARISON OF ENSEMBLE MODEL WITH BASELINE MODELS

Models from Literature	Accuracy	Precision	Sensitivity	AUC
Rajaraman et al. [5]	96.2	97.7	96.2	99.3
Rahman et al. [6]	98	97	99	98
Chouhan et al. [4]	96.39	93.28	99.62	99.34
Enes Ayan et al. [27]	84.5	91.3	89.1	87
Saraiva et al. [8]	94.4	94.3	94.5	94.5
Kermany et al. [7]	92.8	-	93.2	-
Ensemble Model - Proposed	95.1	96.92	95.21	94.5

On the comparison study for previous works in Table II, the works of Kermany et al. in [7], Rajaraman et al. in [5] and Rahman et al. in [6] have been significant in automation of pneumonia detection. They have also compared their works with other authors who have published their results in the same problem. Rahman reported 98% classification accuracy and AUC score of 98 for an input image size of 224 x 224 for ResNet18 and DenseNet201 and 227 x 227 for AlexNet and SqueezeNet architectures. Rajaraman et al. [5], provided a study of the better results of cropped ROI (Region of Interest) data when compared to baseline data and proposed a customized VGG 16 model for the problem. Vikash et al proposed an ensemble model with an AUC score of 99.34 and 96.39 accuracy. Saraiva et al. in [8] evaluated a CNN network with an input image size of 150 x 150 pixels and reported 94.4% accuracy and 94.5 AUC score.

The slight dip in both [8] and proposed Ensemble results can be attributed to low training image resolution which affects the classification performance of the CNN. This has been supported and reasoned in the works of Sarkar et al. in [28], Koziarski et al. in [29], Dodge et al. in [30] and Kannojia et al. in [31]. The work in [29], reports that models that achieved high accuracy on the original, undistorted images were also more resilient to low image resolution and the pattern was observed across almost all the architectures. In our work, the relative dip in performance on testing with images of lower resolution would be less compared to the previous works due to the lower training set image resolution. This is also confirmed by the works of [31] in the reported values for MNIST and CIFAR-10 datasets where we could observe the drastic lowering of performance in models when the resolution of training and testing images vary vastly. They also state that improved results on low quality images need models trained with lower quality images which is the core idea behind our work.

For a 224 x 224-pixel image of 8-bit depth, file size is 6.125 KB whereas a 184 x 184-pixel image of 8-bit depth is 4.132 KB, which is nearly 150% increase in file size. In addition to it, it adds on to the computation time due to the increased number of parameters in the CNN leading to poor efficiency. With COVID-19 pandemic causing damages on a global scale, deep learning solutions using X-ray images are being actively proposed by researchers. Sridhar et al. in [32] evaluated a ResNet model to identify the similar regions between the X-rays of different lung disease and reported Atelectasis, Consolidation, Emphysema, and Pneumonia are most similar in nature to COVID-19. The result augments the need for more extensive research in Pneumonia detection to help distinguish the disease from COVID-19 which would help provide timely and appropriate treatment.

VI. Conclusion

The study presents a transfer-learning based ensemble model to automate Pneumonia detection using Chest X-rays. Different CNN architectures were fine-tuned, trained and the results analyzed to finally propose an ensemble model. The other core idea focused on the work is to lower the resolution and size of the images used while balancing the trade-off with the performance of the model. The final ensemble model

evaluated had an accuracy of 95.03 and AUC score of 94.5 with a precision of 96.92. With significant research in the problem yielding promising results, such models can be deployed in real life to reduce the workload on physicians and bring down human error levels. With lower storage capacity and computing device needs, the implementation can be taken to remote rural areas globally that lack proper diagnoses and treatment for such illness due to lack of skilled doctors and radiologists, poor connectivity and lack of infrastructure. Although it cannot replace a physician, it can aid the diagnosis process and reduce the crucial time taken.

Conflicts of Interest

All the authors declare that they have no conflicts of interest to report regarding this study.

References

[1] I. Rudan, L. Tomaskovic, C. Boschi-Pinto, H. Campbell, and WHO Child Health Epidemiology Reference Group, "Global estimate of the incidence of clinical pneumonia among children under five years of age," Bull. World Health Organ., vol. 82, no. 12, pp. 895–903, 2004.

[2] D. A. McAllister et al., "Global, regional, and national estimates of pneumonia morbidity and mortality in children younger than 5 years between 2000 and 2015: a systematic analysis," Lancet Glob. Health, vol. 7, no. 1, pp. e47–e57, 2019.

[3] J. Gu et al., "Recent advances in convolutional neural networks," Pattern Recognit., vol. 77, pp. 354–377, 2018.

[4] V. Chouhan et al., "A novel transfer learning based approach for pneumonia detection in chest X-ray images," Appl. Sci. (Basel), vol. 10, no. 2, p. 559, 2020.

[5] S. Rajaraman, S. Candemir, I. Kim, G. Thoma, and S. Antani, "Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs," Appl. Sci. (Basel), vol. 8, no. 10, 2018.

[6] T. Rahman et al., "Transfer learning with deep Convolutional Neural Network (CNN) for pneumonia detection using chest X-ray," Appl. Sci. (Basel), vol. 10, no. 9, p. 3233, 2020.

[7] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, pp. 1122-1131.e9, 2018.

[8] A. Saraiva et al., "Models of learning to classify X-ray images for the detection of pneumonia using neural networks," in Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies, 2019.

[9] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan and A. Mittal, "Pneumonia Detection Using CNN based Feature Extraction," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2019, pp. 1-7, doi: 10.1109/ICECCT.2019.8869364.

[10] Quan H, Xu X, Zheng T, Li Z, Zhao M, Cui X. DenseCapsNet: Detection of COVID-19 from X-ray images using a capsule neural network. Comput Biol Med. 2021; 133:104399. doi:10.1016/j.compbiomed.2021.104399.

[11] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, 2010.

[12] S. Rajaraman et al., "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," PeerJ, vol. 6, p. e4568, 2018.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv [cs.CV], 2014.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

[17] Kermany, Daniel, Kang Zhang, and Michael Goldbaum. "Labeled optical coherence tomography (oct) and chest X-ray images for classification." Mendeley data vol.2, 2018.

[18] S. Lakshmanan, "Towards AI, How, When, and Why Should You Normalize/Standardize/Rescale Your Data?" Towards AI-The Best of Tech, Science, and Engineering.

[19] D. R. Wilson and T. R. Martinez, "The need for smaller learning rates on large problems," in IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222),IEEE, 2001, p. vol. 1, pp. 115–119.

[20] S. Ruder, "An overview of gradient descent optimization algorithms," arXiv [cs.LG], 2016.

[21] I. Sutskever, J. Martens, G. Dahl and G. Hinton, "On the importance of initialization and momentum in deep learning," in International conference on machine learning ,May 26 2013, PMLR, pp. 1139–1147.

[22] N. Srivastava , G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting" The journal of machine learning research. ;15(1):1929-58, 2014.

[23] A. F. Agarap, "Deep learning using rectified linear units (relu)," arXiv preprint arXiv:1803.08375., 2018.

[24] B. Gao and L. Pavel, "On the properties of the softmax function with application in game theory and reinforcement learning," arXiv [math.OC], 2017.

[25] Fawcett and Tom, "ROC graphs: Notes and practical considerations for researchers," Machine learning. 2004 Mar 16;31(1):1-38, 2004.

[26] G. R. Cogalton and M. Story M, "Accuracy assessment: a user's perspective." Photogrammetric Engineering and remote sensing.52(3):397-9, 1986.

[27] E. Ayan and H. M. Ünver, "Diagnosis of pneumonia from chest x-ray images using deep learning," Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science (EBBT), pp. 1–5, 2019.

[28] R. Sarkar, A. Hazra, K. Sadhu, and P. Ghosh, "A novel method for pneumonia diagnosis from chest X-ray images using deep residual learning with separable convolutional networks," in Computer Vision and Machine Intelligence in Medical Image Analysis, Singapore: Springer Singapore, 2020, pp. 1–12.

[29] M. Koziarski and B. Cyganek, , Impact of low resolution on image recognition with deep neural networks: An experimental study, vol. 28, no. 4. International Journal of Applied Mathematics and Computer Science, 2018.

[30] S. Dodge and L. Karam, "Understanding how image quality affectsdeep neural networks," Eighth International Conference on Quality of Multimedia Experience(QoMEX), vol. 19, pp. 1–6, 2016.

[31] S. P. Kannojia and G. Jaiswal, "Effects of varying resolution on performance of cnn based image classification: An experimental study," International Journal of Computer Sciences and Engineering, vol. 6, no. 9, pp. 451–456, 2018.

[32] S. Sridhar, S. Mootha, R. Seetharaman and A. X. A. Rayan, "A Study on Co-occurrence of various Lung Diseases and COVID-19 by observing Chest X-Ray Similarity using Deep Convolutional Neural Network," in 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), 2020, pp. 1–8.