Pneumonia Detection from Chest Radiograph using Deep Convolutional Neural Network

Anmol Chopra **Ishika Goyal** **Ayushi** **Asmi** **Jai Mehta**

Department of Computer Science Engineering, Chandigarh University, Punjab 140413, India

Abstract: Pneumonia is a respiratory infection caused by bacteria or viruses; it affects many individuals, especially in developing and underdeveloped nations, where high levels of pollution, unhygienic living conditions, and overcrowding are relatively common, together with inadequate medical infrastructure. Pneumonia causes pleural effusion, a condition in which fluids fill the lung, causing respiratory difficulty. Early diagnosis of pneumonia is crucial to ensure curative treatment and increase survival rates. Chest Xray imaging is the most frequently used method for diagnosing pneumonia. However, the examination of chest X-rays is a challenging task and is prone to subjective variability. In this study, we developed a computer-aided diagnosis system for automatic pneumonia detection using chest X-ray images. We employed deep transfer learning to handle the scarcity of available data and designed an ensemble of three convolutional neural network models: GoogLeNet, ResNet-18, and DenseNet-121. A weighted average ensemble technique was adopted, wherein the weights assigned to the base learners were determined using a novel approach. The scores of four standard evaluation metrics, precision, recall, f1-score, and the area under the curve, are fused to form the weight vector, which in studies in the literature was frequently set experimentally, a method that is prone to error. The proposed approach was evaluated on two publicly available pneumonia X-ray datasets. The results were superior to those of state-of-the-art methods and our method performed better than the widely used ensemble techniques.

Keywords: Pneumonia, Prediction, Convolution Neural Network, Accuracy, Radiograph, Classification

I. INTRODUCTION

Pneumonia is a life-threatening disease, which occurs in the lungs caused by either bacterial or viral infection. It can be life-endangering if not acted upon at the right time and thus the early diagnosis of pneumonia is vital. Pneumonia is considered the greatest cause of child fatalities all over the world. Approximately 1.4 million children die of pneumonia every year, which is 18% of the total children died at less than five years old. Globally, overall, two billion people are suffering from pneumonia every year. Pneumonia is a lung infection, which can be caused by either bacteria or viruses. Luckily, this bacterial or viral infectious disease can be well treated by antibiotics and antivirals drugs. Nevertheless, faster diagnosis of viral or bacterial pneumonia and the consequent application of correct medication can help significantly to prevent deterioration of a patient's condition, which eventually leads to death. Chest X-rays are currently the best method for diagnosing pneumonia. X-ray images of pneumonia are not very clear and are often misclassified to other diseases or other benign abnormalities. Moreover, the bacterial or viral pneumonia images are sometimes miss-classified by the experts, which leads to wrong medication being given to the patients and thereby worsening the condition of the patients. There are considerable subjective inconsistencies in the decisions of radiologists reported in diagnosing pneumonia. There is also a lack of trained radiologists in low resource countries (LRC), especially in rural areas. Therefore, there is a pressing need for computer-aided diagnosis (CAD) systems, which can help the radiologists in detecting different types of pneumonia from the chest X-ray images immediately after the acquisition.

Machine learning and deep learning algorithms can be used for the early prediction of the disease. The purpose of this project is to implement one such technique to predict whether a patient has pneumonia or not. The literature review presents the method used for the said prediction. A dataset containing chest radiographical images have been used for the development of a deep learning model. Since a convolutional neural network(CNN) provides commonly used to analyse images, a convolutional neural network is developed in the project. The model is evaluated for its accuracy of prediction, loss, precision, recall, f-score and AUC-score. Two pre-trained CNN models: ResNet-50 and Inception-V3 are also used to compare the performance of the proposed model. The performance of the models is compared based on classification accuracy, loss, precision, recall, f-score and AUC-score.

A compiling factor is that the dataset used for training the model is typically class unbalanced, with a majority of images belonging to the normal class and a few belonging to the pneumonia class and viral pneumonia class. Therefore only the classification accuracy of the model is not enough to correctly evaluate the model. Other classification metrics thus used in the project for evaluating the

CNN model. These include the ROC value, precision and recall of each class of the dataset. F-score is also calculated for the model. These metrics are compared and the findings are presented in the paper. This paper thus presents the model exclusively developed for pneumonia prediction and the classification metrics used for the performance of the model.

This is an end-to-end deep learning project. The model developed was deployed to be used for early diagnosis of the disease using x-ray images. The findings of this project should allow future researchers to choose deep learning algorithms as an effective way of diagnosis the disease.

The rest of the paper is represented in the five sections. They are as follows: Section 2 represents the literature review. The theory is described in section 3. The methodology of the project is presented in section 4. Section 5 represents results and discussions and conclusions along with future works presented in section 6.

A. *Related Works*

Several studies have shown the use of deep learning and CNN for the prediction of pneumonia and its diagnosis.

Shaoqing Ren et al.[5] have presented a faster RCNN for real-time object detection with region proposal networks. They have presented RPNs (Region proposal networks) for efficient and accurate region proposal generation. They achieved a nearly cost-free region proposal step by sharing convolutional features with the downstream detection network. Their object detection system, called Faster R-CNN, is composed of two modules. The first module is a deep fully convolutional network that proposes regions, and the second module is the Fast R-CNN detector that uses the proposed regions. Their method enabled a unified, deep-learning-based object detection system to run at near-real-time frame rates. The learned RPN also improves region proposal quality and thus the overall object detection accuracy.

Ross Girshick[6] in his paper, proposed a fast Region-based convolutional network(Fast R-CNN) for object detection. He proposed a new training algorithm that fixes the disadvantages of R-CNN and SPPnet, while improving their speed and accuracy. His experiments use three pre-trained ImageNet models. The first is the CaffeNet (essentially AlexNet) from R-CNN. The second network is VGG CNN is used. The final network used is the very deep VGG16 model.

Kaiming He[7] in their research have presented Mask R-CNN architecture. They have set hyperparameters following existing Fast/Faster R-CNN work. They performed a thorough comparison of Mask R-CNN to the state of the art along with comprehensive ablation experiments. They have used the COCO dataset for all experiments.

Zhaowei Cai et al.[8] in their paper have proposed a multi-stage object detection framework, the Cascade R-CNN, for the design of high-quality object detectors. They have designed this architecture to avoid the problems of overfitting at training and quality mismatch at inference. The solid and consistent detection improvements of the Cascade R-CNN on the challenging COCO and the popular PASCAL VOC datasets suggest the modelling and understanding of various concurring factors are required to advance object detection. The Cascade R-CNN was shown to apply to many object detection architectures. The authors believe that it can be useful to many future object detection research efforts.

Joseph Paul Cohen et al.[9] used chest xray14 image dataset which contains chest x-rays of patients having pneumonia and pneumonia. In their paper, they discuss the dataset and analysed its content. They mentioned tools that can be developed for the prediction of pneumonia from the dataset.

Kolla Bhanu Prakash et al.[10], in their paper have discussed various machine learning algorithms that can be applied for the prediction of pneumonia. They have used Pneumonia-India and Pneumonia-Data to build ML models for the prediction. They have used various machine learning algorithms like SVM, KNN, MLP, decision tree, an ensemble of decision tree and other algorithms and have compared their accuracy. They have concluded that random forests have outperformed other algorithms with an accuracy of 96.66%.

Luca Brunese et al.[11] have used KNN to analyse medical images of chest x-rays and predict the presence of pneumonia and other diseases with similar pulmonary symptoms. They have divided the dataset before training into a training set containing 90% of the data. They obtained an average precision and recall of 0.965. They have also calculated the ROC score and f-measure.

R. Sujath et al.[12], in their paper discussed the growing cases of pneumonia in India using the pneumonia dataset from Kaggle. They analysed the data and presented a visualization of the data. They built multilayer perceptron, linear regression and vector autoregression(VAR) models to predict the presence of pneumonia among patients. In the conclusion of their findings, they discussed that multilayer perceptron performed better than linear regression and VAR in predicting the disease based on the dataset used.

Linda Wang et al.[13] have built a CNN model for the prediction of pneumonia from the pneumonia dataset including 13,975 CXR images. They compared the performance of the proposed model with ResNet-50 and VGG-19 models. According to their conclusion, the model performed better than the other two architecture. Their model achieved an accuracy of 93.3% on the pneumonia dataset.

Xuehai He et al.[14] in their project have used CT scan images to develop deep learning methods for the prediction of pneumonia. They have integrated self-supervised and transfer learning methods to achieve higher accuracy. Due to the public unavailability of the dataset for this purpose, they have made a publicly available dataset for the same purpose. Their pneumonia samples contains 35K images for pneumonia positive samples. They have compared the performance of the different network and summarised the result. Their experimentation results have shown an F1 of 0.85 and AUC of 0.94. The above-related works indicate that the use of deep learning for the prediction of pneumonia using either chest x-ray or CT scans can be an efficient way of diagnosis. The models that the above authors have proposed have achieved accuracy to a great extent. Along with model accuracy, their models are also validated with precision, recall f-score and AUC. This paper thus discusses the performance of another deep learning model developed for the pneumonia prediction that has achieved even higher accuracy along with higher AUC score, f-score, precision and recall.

B. *Theoretical Background*

1) Convolutional Neural Network:

Convolutional Neural Network (CNN)[2] is a class of neural network in deep learning which is most commonly used for visual imagery analysis. They are also referred to as space invariant artificial neural networks or SIANN. It is based on the weight shared architecture of convolution kernels that slides through the input features to generate feature maps. CNN is a regularized version of MLP (Multilayer perceptron). CNNs are inspired by the biological processes of the neural connection of the visual cortex of animals. [3]

Architecture- The CNNs implements a mathematical operation known as convolution operation. Convolution neural networks are special types of neural network that uses the convolution function instead of matrix multiplication in their layers.

The general architecture of a typical CNN consists of input layers, hidden layers and output layers. A typical CNN architecture consists of:

- Convolution layers
- Pooling layers
- Fully connected layers
- Dropout layers
- Normalization layers
- Activation functions
- Receptive fields

The hidden layers in a typical CNN include the convolution layers. The input in a CNN is a tensor of shape (number of inputs)*(input height)*(input width)*(colour channels). A feature map is generated after the convolution kernel or filters is convoluted over the input tensor. Each convolution layer processes input data for its receptive fields.[4] Receptive fields are the restricted are of the previous layer which is used for obtaining desired pixels. The convolution layers include various hyperparameters such as height and width of the filters, colour channels, strides, padding and dilation. The CNN is advantageous over traditional neural networks for computer vision is because of the convolution operation due to which the number of trainable parameters reduces. Using regularized weights over fewer parameters solves the problem of vanishing and exploding gradients.

The computation of the output of a neuron in the convolution layer is shown in equation 1.

$$z_{ijk} = b_k + \sum_{u=0}^{f_h-1} \sum_{v=0}^{f_w-1} \sum_{k'=0}^{f_{n'}-1} x_{i',j',k'} \cdot w_{u,v,k',k} \qquad (1)$$

In equation 1, $i' = i \times s_h + u$ and $j' = j \times s_w + v$. $z_{i,j,k}$ is the output of the neuron located in row i, column j in the feature map k of the convolutional layer. $x_{i',j',k'}$ is the output of the neuron located in layer l-1. $w_{u,v,k',k}$ is the connection weight between any neuron in the feature map k of layer l. b_k is the bias term for feature map k in layer l. s_h and s_w are the vertical and horizontal strides. f_h and f_w are the height and width of the receptive fields and $f_{n'}$ is the number of feature maps in the previous layer.

The pooling layers in a CNN reduces the dimensions of the data to reduce the requirement of computational load, the usage of memory and the number of parameters thus limiting the risk of overfitting. A pooling neuron is connected to a limited number of neurons of the previous layers. The pooling neuron does not have any weights. There are two most common types of pooling layers: Average pooling and max pooling. The max-pooling takes the maximum value of each local cluster of neurons in the feature map while average pooling takes the average value.

Dropout and normalization layers in a CNN prevents the risk of overfitting. Dropout layers are used to drop a few neurons from the previous layers to prevent overfitting. Dropout layers are used in both hidden and visible layers. Normalization of images makes it more robust to variations. The most common normalization method used is batchnormalization. This layer allows the neurons to learn more independently. They are added to standardise the inputs or outputs. Fully connected layers in a CNN are the same as that of a traditional neural network. They connect every neurone in the previous layer to every neuron on the following layer.

The activation function in a neural network defines how the nodes of the output layer are transformed from the weighted sum of the input. Activation functions in a CNN are used in both hidden and output layers. The common activation function used in the hidden layers is rectified linear unit (ReLU), sigmoid. The most common activation function used in the hidden layers is ReLU. It is susceptible to a vanishing gradient. The ReLU function is calculated as shown in equation 2.

$$f(z) = \max(0, z) \qquad (2)$$

Softmax layers are used as a probabilistic output of the nodes of the previous layers. The output of a neuron in softmax layers depends upon the output of other neurons of the layer. The output of the softmax function will give the probability of each class in the target.

The convolutional neural network has huge applications in the field of computer vision, though they require preprocessing compared to other image classification algorithms. CNN has applications in image classification, image segmentation, image and video recognition and medical image analysis.

Different classification metrics such as classification accuracy, precision, recall f-score and ROC values are used to evaluate the CNN model. The use of recall, precision and ROC values are used as the training data is a class imbalanced data.

2) Accuracy:

The classification accuracy is a measure of how well the classifier can correctly predict cases into their correct category. Accuracy can be calculated using the following equation:

$$\text{Accuracy} = \left(\frac{TP+TN}{P+N}\right) \times 100\% \qquad (3)$$

Equation 3 is used to calculate the accuracy using the sample cases where P and N represent the number of positive and negative samples and TP and TN represents the True Positive and True Negative values respectively.

3) Precision:

It is a measure of the number of true positives and true negatives. It is also known as confidence. It is a measure of the cost of false positive. It is useful for medical diagnosis since a false positive is not a desirable result. It is calculated using equation 4

$$\text{Precision} = \left(\frac{TP}{TP+FP}\right) \times 100\% \qquad (4)$$

TP and FP represent the number of true positive and false-positive samples.

4) Recall:

The recall is the measure of the number of positive samples captured by positive predictions. It is also known as sensitivity. This is a desirable measure, especially in the medical field because of the number of observations that are correctly diagnosed. In this study it is more important to correctly identify the number of cases of pneumonia than the cases of healthy or viral pneumonia. It is calculated using equation 5

$$\text{Recall} = \left(\frac{TP}{TP+FN}\right) \times 100\% \qquad (5)$$

FN represents the number of false negatives.

5) F-Score:

Observing the value of only one of precision or recall will not provide the full picture. The f-score or f-measure is one way to summarize them, which is with the harmonic mean of precision and recall. It is calculated using equation 6

$$\text{fscore} = 2 \times \left(\frac{Precsion \times Recall}{Precsion + Recall} \right) \quad (6)$$

This particular variant is called the f1 -score. It can be a better measure than accuracy on imbalanced binary classification datasets as it takes precision and recalls into account.

6) AUC:

AUC provides an aggregate measure of performance across all possible classification thresholds. Value of AUC ranges in value from 0 to 1. AUC is a desirable metric because it is scale-invariant and classification-threshold-invariant. It measures the ranking of the predictions rather than their absolute values. It also measures the prediction quality irrespective of what classification threshold is chosen.

II. METHODOLOGY

The methodology of the project is discussed in this section. The overview of the project methodology is presented in Fig 1.

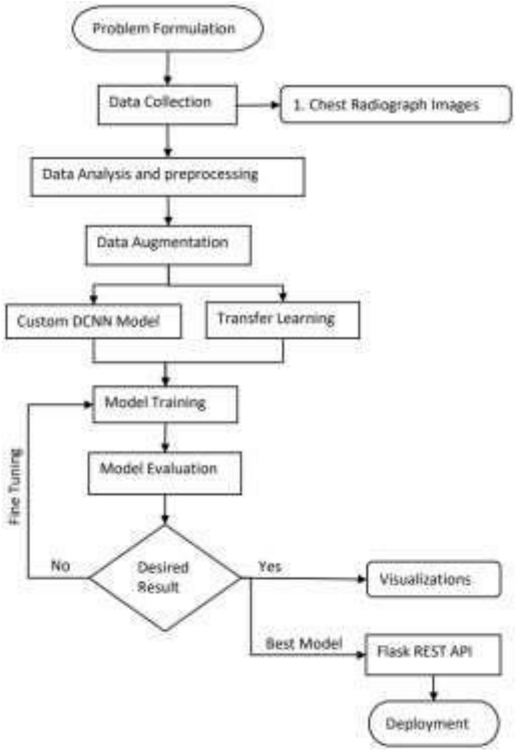

Fig. 1. Methodology for the proposed project

The details of the dataset, data pre-processing, proposed DCNN architecture and transfer learning is presented later in the section along with its implantation, training and deployment.

C. *The Dataset*

The dataset[15][16][17] used for this project contained 21,165 images. X-ray images of lungs with pneumonia, viral pneumonia, lung opacity and healthy lungs are contained in the dataset. The dataset used for the training and development of the CNN model for pneumonia prediction had been pre-processed and contains 15,153 images of chest radiograph belonging to pneumonia, viral pneumonia and healthy lungs. The images are of shape 299×299×3. Randomly selected samples of chest x-ray images belonging to the three classes are shown in Fig. 2.

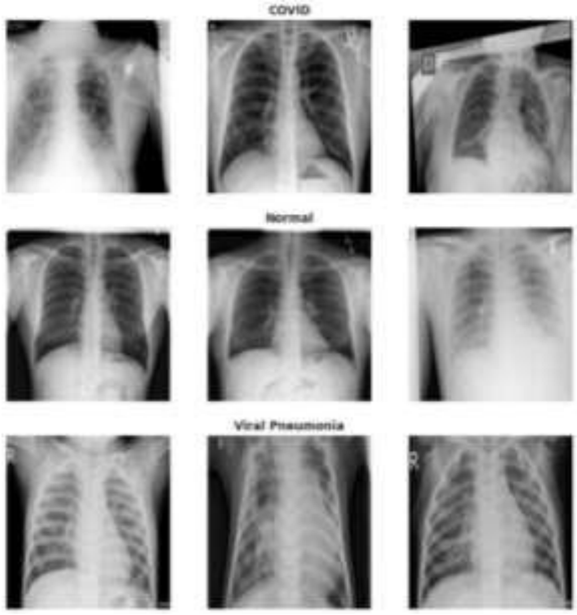

Fig. 2. Chest radiograph samples from the dataset belonging to 3 classes.

A complicating factor about the data is that it is a class imbalanced dataset. 67.3% of images in the dataset belongs to healthy lungs. 23.9% of images belong to pneumonia and only 8.9% of images belong to viral pneumonia. The percentage distribution of classes is shown in Fig. 3.

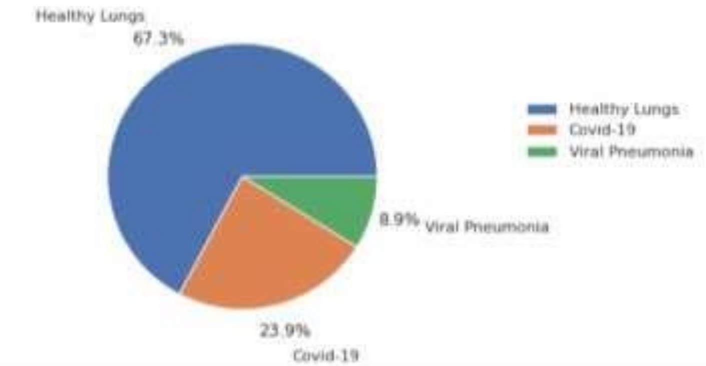

Fig. 3. Percentage distribution of the classes in the dataset.

The number of image samples of each class of the dataset is shown in figure 4. It can be seen that the number of samples diagnosed as 'healthy lungs' is 10,192 and is comparatively more than the other two classes. Due to this imbalanced dataset, only the classification accuracy of the model is not enough for the evaluation. Other classification metrics such as precision, recall, f-score, AUC have to be used to evaluate the performance of the predictive model.

D. *Data Preparation and Pre-processing*

The x-ray images in the dataset have a shape of 299×299×3. The shape of the images was changed to a size of 150×150×3. The training set shape is (15153, 150, 150, 3). All the three colour spaces i.e. red, green, blue are kept the same for the training. The classes to be predicted are 'Healthy Lungs', 'Pneumonia' and 'Viral Pneumonia' having 10192, 3616 and 1345 number of samples respectively. The number of samples in each class is visualized in Fig. 3. The dataset is distributed in a ratio of 80%-20%. 80% of data i.e. 12,123 samples are used for training and 3,031 samples are used for testing purposes. A validation set having 1,818 samples is used to validate the training accuracy and loss.

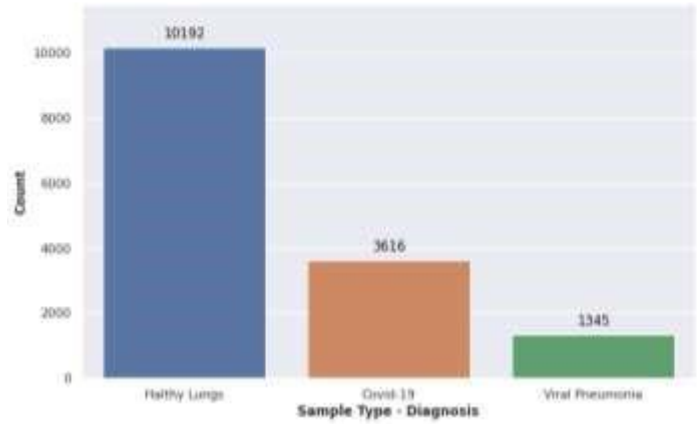

Fig. 4. Number image samples belonging to each class.

E. *Data Normalization*

Normalization of the training data is a crucial step in image processing. Normalization is necessary to make the data comparable across the entire record. Existing normalization techniques such as min-max normalization, z-score normalization and decimal scaling normalization are used predominantly. In this project, the image data are normalized by diving 255. The images in the training set have been converted to an array and normalized with range [0,1].

F. *Data Augmentation*

Data Normalization is a strategy which is used to increase the size dataset by adding slightly modified images of the existing images. This artificially increased size of the dataset using various modified images from the existing images helps the model to protect from the problem of overfitting. Various operations such as zooming, flipping, shifting, etc. are applied for augmentation of the data. In this project, different augmentation parameters are used on the image data to overcome the problem of overfitting. The augmentation parameters used on the training data is illustrated in Table I.

TABLE I DATA AUGMENTATION PARAMETERS AND VALUES USED

Data Augmentation Parameter	**Parameter Value**
Rotation_range	20
Height_shift_range	0.2
Width_shift_range	0.2
Shear_range	0.2
Zoom_range	0.2
Vertical_flip	True
Horizontal_flip	True

G. *The Deep Convolution Neural Network Architecture*

In this project, a convolutional neural network(CNN) architecture is constructed for pneumonia prediction. This model is constructed to predict one of the following three predictions: a) no infection (normal), b) viral pneumonia and c) pneumonia infection. The CNN model developed for pneumonia prediction is implemented using Keras deep learning library's sequential API [18]. The model contains 8 Convolution layers, 10 Batch Normalization layers,5 MaxPooling layers, 4 Dropout layers and 3 fully connected layers.

The CNN model consists of 1,331,779 parameters out of which 1,329,091 are trainable and 2,688 are non-trainable parameters. It is 32 layered architecture. The input shape of the model indicates that the chest x-ray images used for training are of shape 150×150×3 having all three additive colour spaces i.e. Red, green and blue. The output has a shape of 3 for the classification of 3 sample classes in the dataset. The small numbers of parameters of the model indicate that the proposed model is lightweight.

The proposed architecture comprises a heterogeneous mix of convolution layers with a kernel size of 3×3 and pooling layers of pool size 2×2. A varied filter size for the dimensionality of the output space is used in different convolution layers of the architecture. This architectural diversity of the CNN architecture provides a strengthened feature detection from the input radiographical images and enhances the representational capacity of the task of pneumonia prediction. A visual representation of the proposed DCNN architecture is presented in Fig. 5. It is visible from the figure that the network architecture contains 32 layers including input and output layers.

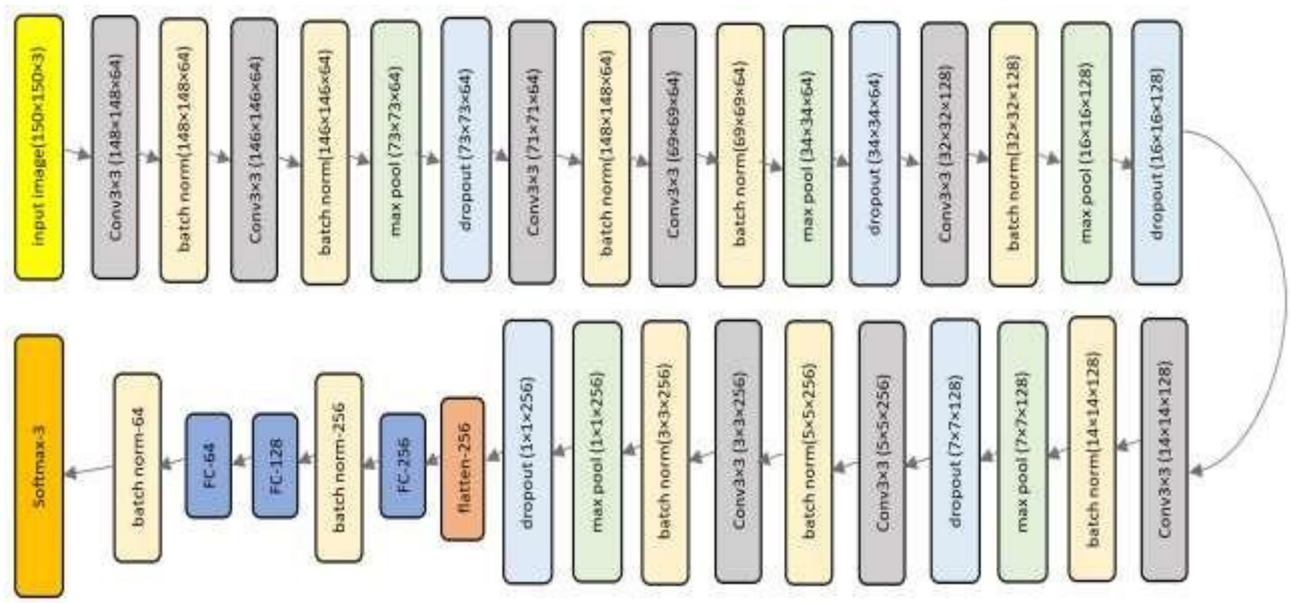

Fig. 5. Model architecture of the proposed DCNN.

H. *Model Training and implementation*

80% of the data that is 12123 images have been used to train the model. 1,818 images are used as a validation set to keep a check on overfitting. The images used for training the model is of the shape 150×150×3. The model is compiled and trained using the TensorFlow backend and Keras deep learning library.

The optimizer used for this model is Adam[19] which stands for adaptive moment optimizer. It is a combination of both momentum optimization and RMSProp optimizers. This optimizer keeps track of the exponentially decaying average of previous gradients and also keeps track of an exponentially decaying average of past squared gradients just like in momentum and RMSProp optimizers[20]. Adam works with the momentum of both first and second-order Adam includes bias corrections to the estimates of both the first-order moments (the momentum term) and the (uncentered) second-order moments to account for their initialization at the origin[21]. This optimizer is computationally costly but is very fast and converges to global minima very rapidly.

The loss function used for evaluating the model and for the training of the parameters is categorical cross-entropy. It is also known as softmax loss. This function is a combination of cross-entropy loss and Softmax activation function. The use of this loss gives the prediction probability of each class. This loss function is used because of the multiclass classification problem of this project. The loss is implemented using Keras API's SparseCategoricalCrossEntropy() method [22].

A batch size of 64 is used to train the model. The model is trained for 100 epochs. A learning rate of 0.001 is used. Other hyperparameters used were factor=0.5 and patience =10. 321 samples are trained in each epoch. Accuracy, loss, precision, recall, f-score and AUC score of both training and validation sets are monitored during the model training. Tensor processing units(TPU) accelerators [23] is used to speed up the training of the model. GPU is used because of the large size of the training data containing 12,123 images each of shape 150×150×3. It minimized the time-toaccuracy during training of the CNN model.

I. *Transfer Learning*

The two pre-trained CNN models used for the comparison of the performance of the proposed models are ResNet-50 and Inception-V3.

ResNet-50[24] is a variant of the ResNet model. The architecture of ResNet-50 is similar to a bottleneck structure. The

ResNet-50 architecture consists of a total of 50 layers of which 48 are Convolution layers along with 1 MaxPool and 1 Average Pool layer. It has 3.8 x 10^9 Floating points operations. The ResNet-50 model achieved a top-1 error rate of 20.47 per cent and achieved a top-5 error rate of 5.25 per cent on the ImageNet validation set.

Inception-V3 [25] is a variant of the Inception model. This model is 48 layers deep. The Inception-V3 architecture consists of about 23 million parameters. Inception-V3 reaches 21.2%, top-1 and 5.6% top-5 error for single crop evaluation on the ILSVR 2012 classification. In Inception-V3, various techniques have been suggested for network optimization such as factorized convolutions, regularization, dimension reduction and parallelized computations. These techniques have been used to loosen the constraints for easier model adaptation.

J. *Deployment*

This project is an end-to-end deep learning project to predict pneumonia using chest radiographs. The deployment is of the model is necessary to be used to predict the presence of the disease.

The model is deployed through a web app on the Heroku server. Heroku is a platform as a service(PaaS). A flask API is created to be deployed on the Heroku cloud platform. This website[26] can be used for pneumonia prediction.

III. RESULT AND DISCUSSION

The performance of the proposed model in terms of classification accuracy, loss, precision, recall, f-score and AUC score is compared to that of ResNet-50 and Inception-V3. The classification accuracy achieved by ResNet-50, Inception-V3 and the proposed models is 98.92%, 97.67% and 98.86% respectively. All the classification metrics of the models over training and test are presented in table II and table III respectively.

TABLE II PERFORMANCE TESTED ON THE TRAINING SET USING THE THREE MODELS

Models	**Params(M)**	**Accuracy(%)**	**Precision(%)**	**Recall(%)**	**F-Score**	**AUC**
ResNet-50	25.6	98.92	98.41	98.36	0.9842	0.9987
Inception-V3	23.8	97.67	96.58	96.42	0.9649	0.9959
Proposed Model	1.3	98.86	98.29	98.28	0.9828	0.99.91

TABLE III PERFORMANCE TESTED ON THE TEST SET USING THE THREE MODELS

Models	**Params(M)**	**Accuracy(%)**	**Precision(%)**	**Recall(%)**	**F-Score**	**AUC**
ResNet-50	25.6	82.01	73.07	72.90	0.7290	0.8655
Inception-V3	23.8	66.36	49.53	49.26	0.4938	0.7244
Proposed Model	1.3	93.98	90.99	90.95	0.9096	0.9844

The results from the above two tables show that the lightweight proposed model performs at par on the training set and works comparatively better on the test set than ResNet-50 and Inception-V3 models. All the classification metrics of the proposed model on the test set is better than the rest two models. The results from table III prove that the proposed model performed better on unseen data than ResNet-50 and Inception-V3.

The proposed model is trained on the GPU for 100 epochs using a batch size of 64 and achieved a training accuracy of 98.86% and a validation accuracy of 95.46%. Accuracy, loss, precision, recall, f-score and AUC of both training and validation sets are visualized in Fig. 6.

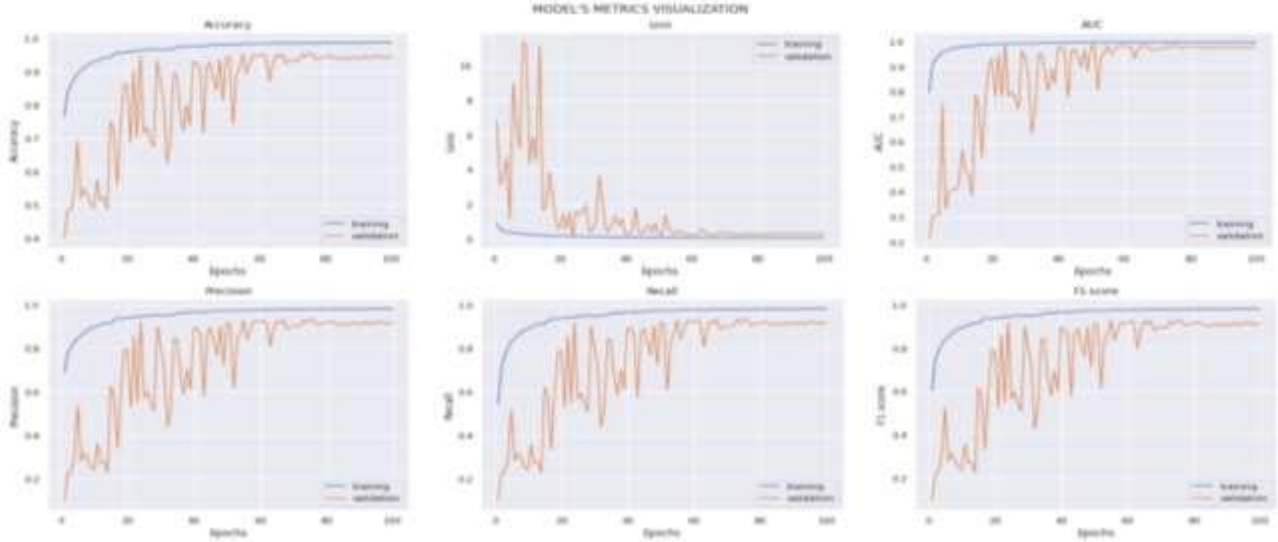

Fig. 6. Comparison of performance of proposed DCNN on training and validation using classification accuracy, loss, AUC, precision, recall and f-score during 100 epochs of training.

It is visible from figure 6, that both the training and validation metrics converge at the 100th epoch. It is clear from the accuracy and loss values of the training and validation set from figure 5 that the model does not overfit. The fluctuations in the accuracy and loss validation data is not a big problem since the model has the freedom of choosing ways to classify the training data.

The precision, recall and f-score of the training and validation set converge at the 100th epochs which shows that the proposed model correctly identifies images in the class-imbalanced dataset and correctly classifies and unseen data into the desired classes. The AUC value of both sets converges at the 100th epoch as well. This proves that the proposed model's prediction quality is excellent.

The model achieved an accuracy of 94% on the held-out test set which is a good measure. Table IV shows the confusion matrix of the trained model on the test set.

TABLE IV CLASSIFICATION MATRIX OF THE MODEL ON THE TEST SET

	Pneumonia	**Normal**	**Viral Pneumonia**
Pneumonia	612	13	0
Normal	190	1524	16
Viral Pneumonia	7	7	207

The test set contains 1,730, 221 and 625 image samples belonging to Normal, Viral Pneumonia and Pneumonia classes respectively. Out of 1,730 samples of the Normal class, the model classified 1,524 samples correctly. Out of 221 samples belonging to Viral Pneumonia, 207 samples have been correctly classified by the model and out of 625 pneumonia samples, 612 samples have been correctly classified by the model. In total, out of all 3,576 image samples, the proposed model correctly classified 2,343 samples which give an accuracy of 93.98% on the test set as mentioned above. The data used for training as well as testing the model is class unbalanced data with the majority of samples belonging to normal class. Due to this reason, the calculation of classification accuracy is not enough. Other classification metrics should be monitored as well. The AUC value achieved by the model is 0.9844 which is a good value.

Since, this is a classification of diagnosis of a medical condition, the true negative and false positive cases should be taken care of. Precision, recall and f-score are also calculated for the test set.

The model achieved an average precision value of 91%. The model achieved an average recall value of 91%. An average f-score of 91% is achieved by the model. The average Precision value, average recall value and average Fscore is summarized in Table V.

TABLE V SPECIFICITY, SENSITIVITY AND F-MEASURE OF THREE CLASSES

	Specificity(%)	**Sensitivity(%)**	**F-Score**
Pneumonia	76	98	0.85

Normal	99	88	0.93
Viral Pneumonia	93	94	0.93
Weighted Average	0.91	0.91	0.91

It is evident from Table V that the precision of the normal class is the greatest 99% while the recall value of the pneumonia class is the largest 98% which is a desirable value for this data. The f-score is highest for the second and third classes. As mentioned above, the accuracy of the held-out test set is 93.98% which is a good value given that symptoms of pneumonia should exhibit certain clear patterns.

A. *Discussion:*

The results presented above show that the proposed model has a very good performance on the test set. The value of classification metrics presented in table 2 and table 3 shows that the proposed DCNN model has a higher chance between the nor and mal, viral pneumonia and pneumonia cases. The selection of appropriate optimizers and loss functions has resulted in higher values in the classification accuracy, precision, recall, AUC and f-score. The model performed better on unseen data than its pre trained counterparts. The use of GPU to train the model has decreased the training time. The model is very lightweight and achieved desirable accuracy. Better accuracy of the model can be achieved using more training images. The enhanced sensitivity and specificity of the model will enhance clinical diagnosis and prognosis. More data and better pre-processing can be used in future to increase the performance of the model. The project can be accessed at [26].

IV. CONCLUSION

In this paper, a deep convolutional neural network architecture is proposed for pneumonia prediction which is developed using TensorFlow backend and Keras deep learning library. The CNN architecture is lightweight and contains 32 layers and 1.3 million parameters. It is trained on 80% of 15,153 images of chest radiographs.

The effectiveness of the model is evaluated based on classification accuracy, sensitivity, specificity and f-measure and AUC score. The model achieved a staggering accuracy of 98.86%. The model also achieved a precision of 98.29%, a recall of 98.28% and an f-score of 0.98 during training. Along with this the model also achieved an AUC of 99.91. The precision, recall and f-score value of all three classes have been presented in the paper. A comparison of the accuracy of the training and validation and test set is presented. The model achieved a satisfying result in this multiclass classification problem.

The performance of the proposed DCNN model is presented in the paper concerning ResNet-50 and Inception-V3 on classification accuracy, recall, precision, f-score and AUC score. The proposed showed better performance on unseen data than ResNet-50 and Inception-V3.

The project has significant value in the field of pneumonia prediction and diagnosis. Further, the deployment of the model can help in the ease of use for clinical diagnosis. The results shown above will allow understanding and can help further research in this field. The use of deep learning for the diagnosis of the disease can decrease the use of other techniques for the pneumonia test such as RT-PCR tests and can speed up the process of diagnosis. The automated prediction using deep learning models can reduce human intervention and thus can reduce the spread of pneumonia infection.

REFERENCES

[1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, ''ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 3462–3471.

[2] D. S. Kermany, G. Michael, C. Wenjia, C. V. Carolina, L. Huiying, L. B. Sally, and M. E. A. Alex, ''Identifying medical diagnoses and treatable diseases by image-based deep learning,'' Cell, vol. 172, no. 5, pp. 1122–1131, 2018.

[3] WHO | World Health Organization. Accessed: Mar. 31, 2020. [Online]. Available: https://www.who.int/

[4] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, ''An efficient deep learning approach to pneumonia classification in healthcare,'' J. Healthcare Eng., vol. 2019, pp. 1–7, Mar. 2019.

[5] R. Siddiqi, ''Automated pneumonia diagnosis using a customized sequential convolutional neural network,'' in Proc. 3rd Int. Conf. Deep Learn. Technol. (ICDLT), 2019, pp. 64–70.

[6] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, ‘‘Deep generative adversarial neural networks for compressive sensing MRI,’’ IEEE Trans. Med. Imag., vol. 38, no. 1, pp. 167–179, Jan. 2019.
[7] A. Oates, K. Halliday, A. C. Offiah, C. Landes, N. Stoodley, A. Jeanes, K. Johnson, S. Chapman, S. M. Stivaros, J. Fairhurst, A. Watt, M. Paddock, K. Giles, K. McHugh, and O. J. Arthurs, ‘‘Shortage of paediatric radiologists acting as an expert witness: Position statement from the british society of paediatric radiology (BSPR) national working group on imaging in suspected physical abuse (SPA),’’ Clin. Radiol., vol. 74, no. 7, pp. 496–502, Jul. 2019.
[8] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, ‘‘Learning to diagnose from scratch by exploiting dependencies among labels,’’ 2017, arXiv:1710.10501. [Online]. Available: http://arxiv.org/abs/1710.10501
[9] G. Deng and L. W. Cahill, ‘‘An adaptive Gaussian filter for noise reduction and edge detection,’’ in Proc. IEEE Conf. Rec. Nucl. Sci. Symp. Med. Imag. Conf., Oct. 1993, pp. 1615–1619.
[10] X. Zhuang and R. M. Haralick, ‘‘Morphological structuring element decomposition,’’ Comput. Vis., Graph., Image Process., vol. 35, no. 3, pp. 370–382, Sep. 1986.
[11] J. Canny, ‘‘A computational approach to edge detection,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
[12] J. A. K. Suykens and J. Vandewalle, ‘‘Least squares support vector machine classifiers,’’ Neural Process. Lett., vol. 9, no. 3, pp. 293–300, Jun. 1999.
[13] L. Breiman, ‘‘Random forests,’’ Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
[14] B. Yegnanarayana, Artificial Neural Networks. New Delhi, Delhi: PHI, 2009.
[15] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, ‘‘Handling imbalanced datasets: A review,’’ GESTS Int. Trans. Comput. Sci. Eng., vol. 30, no. 1, pp. 25–36, 2006.
[16] A. Krizhevsky, S. Ilya, and E. H. Geoffrey, ‘‘ImageNet classification with deep convolutional neural networks,’’ in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097–1105.
[17] I. Goodfellow, P.-A. Jean, M. Mehdi, X. Bing, W.-F. David, O. Sherjil, C. Aaron, and B. Yoshua, ‘‘Generative adversarial nets,’’ in Proc. Adv. Neural Inf. Process. Syst., 2014.
[18] A. Radford, L. Metz, and S. Chintala, ‘‘Unsupervised representation learning with deep convolutional generative adversarial networks,’’ 2015, arXiv:1511.06434. [Online]. Available: http://arxiv.org/abs/1511.06434
[19] T. Karras, S. Laine, and T. Aila, ‘‘A style-based generator architecture for generative adversarial networks,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 4401–4410.
[20] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, ‘‘Unpaired image-to-image translation using cycle-consistent adversarial networks,’’ in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2223–2232.
[21] L. L. G. Oliveira, S. A. E. Silva, L. H. V. Ribeiro, R. M. de Oliveira, C. J. Coelho, and A. L. S. S. Andrade, ‘‘Computer-aided diagnosis in chest radiography for detection of childhood pneumonia,’’ Int. J. Med. Informat., vol. 77, no. 8, pp. 555–564, Aug. 2008.
[22] J. Kim, O. Sangjun, Y. Kim, and M. Lee, ‘‘Convolutional neural network with biologically inspired retinal structure,’’ Procedia Comput. Sci., vol. 88, pp. 145–154, Jan. 2016.
[23] R. Mehra, ‘‘Breast cancer histology images classification: Training from scratch or transfer learning?’’ ICT Exp., vol. 4, no. 4, pp. 247–254, Dec. 2018.
[24] V. Badrinarayanan, A. Kendall, and R. Cipolla, ‘‘SegNet: A deep convolutional encoder-decoder architecture for image segmentation,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[25] O. Ronneberger, F. Philipp, and B. Thomas, ‘‘U-Net: Convolutional networks for biomedical image segmentation,’’ in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2015, pp. 234–241.
[26] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, ‘‘CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning,’’ 2017, arXiv:1711.05225. [Online]. Available: http://arxiv.org/abs/1711.05225