Attention based Residual Network Architecture to Classify Pneumonia from Chest X-ray Images

1st Muneeb Zafar
NUST, SEECS
MSCS Student
mzafar.mscs19seecs@seecs.edu.pk

2nd AsadUllah
NUST, SEECS
MSCS Student
aullah.mscs17seecs@seecs.edu.pk

3rd Asad Rehman
NUST, SEECS
MSCS Student
arehman.mscs19seecs@seecs.edu.pk

***Abstract*—In 2017, 808,694 children died because of pneumonia around the world. More children in the world die from pneumonia than any other disease, like tuberculosis, HIV infection or malaria. Pneumonia accounts for 15% of child deaths. Computer aided diagnosis (CAD) is very helpful in diagnosis of different diseases like pneumonia, but it requires a great deal of training and knowledge of X-Ray images like most radiologists acquire. Even typical CAD methods are not enough since they require a lot of pre-processing and feature extraction which is very time consuming. To help radiologists and doctors, a deep neural network, called Residual Network (ResNet), is proposed in this work. ResNet has deep network structure since it can extract more informative features. Along with ResNet, Convolution block attention module (CBAM) is used to get fine features from X-Ray images. For the testing and training of this model, a chest X-Ray data set available on Keras is used in this work. The accuracy achieved by this proposed model is 94.31%, indicating that this model can be helpful in diagnosing pneumonia quickly and accurately. The data set, model implementations (in Keras), and evaluations, are all made publicly available for research community at https://github.com/MuneebZafar00713/CS-867-Term-Project**

***Index Terms*—Pneumonia, Chest X-ray, Attention based ResNet, Deep Learning, Computer Aided Diagnosis**

I. Introduction

Pneumonia is a severe lung infection that can be caused by viruses, bacteria or fungi. As a result of this infection, air sacs get filled with pus or some other liquid. Types of pneumonia are lober and bronchial. Lober can affect some sections of lungs (also known as lobes), and Bronchial can affect some patches of lungs. There are two main causes of pneumonia, namely bacterial and viral pneumonia. Bacterial pneumonia shows more severe symptoms than viral one. Bacterial pneumonia requires antibiotics for treatment whereas viral infections can get cured on their own [1]. Pneumonia is a leading cause of death in the US. In Pakistan around 58,000 children who were under five years of age, died in the year 2018 [2]. It affects people all over the world, but it is particularly devastating in South Asia and some African countries. Its prevention is not costly. All it needs is timely diagnosis. That's why computer aided diagnosis is needed in this respect for timely and accurate diagnosis of this disease.

There are different tests that can help in diagnosing pneumonia, namely CT of the lungs, needle biopsy of the lung, chest X-rays, chest MRI and chest ultrasound. Among all available methods, Chest X-ray is by far the most common and best pneumonia diagnosing method.The reason why X-ray is better than other methods is because X-ray machines are commonly available, even in under developed countries and also it requires less time to get an X-ray as compared to MRI and CT tests, which are expensive as well.

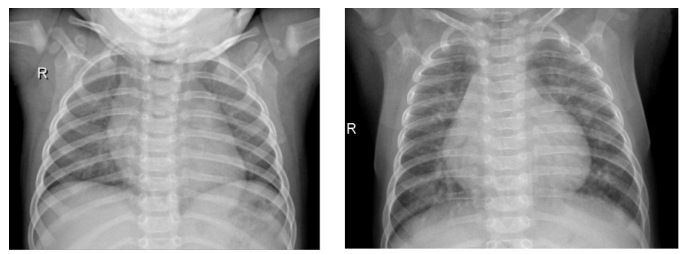

Fig. 1. Normal vs Pneumonia Chest X-ray

Figure 1 shows a side comparison of normal chest X-ray vs pneumonia chest X-ray. Sometimes, it is hard to discern the correct disease from X-rays for radiologists as well. Some radiologists mix some other disease features with pneumonia by looking at chest X-rays, which can make false diagnosis of disease. That is why computer aided diagnosis is required to reduce false prognosis. Conventional CAD approaches are not fruitful, which is why deep convolutional neural networks are used, since they are able to extract more fine grained features with high detection accuracy in image classification.

X-Ray images can be classified using convolution neural networks because they are also images and Convolution Neural Network (CNN) in the classification of images have shown excellent results. Because of this reason, CNNs are also very helpful in medical field for the diagnosis of such diseases [3]. Model proposed in this work is implemented using Attention based Residual Network with data augmentation technique, which extracts required features to train the model. The data imbalance present in the data set has been handled using class weights during model training and categorical cross entropy loss has been utilized to classify between two classes (i.e. Normal, Pneumonia) with high accuracy, as compared to binary classifier that shows biased results due to data imbalance towards pneumonia.

The rest of the paper is arranged in following order. Section II discusses related work. Section III discusses the various methods that have been employed within this paper, as well the proposed methodology that is finally utilized. A short portrayal of the techniques utilized within this paper are discussed as well. In Section IV, the results of the proposed methodology with different parameters are discussed, whereas Section V includes the discussion (methodology and results) as well the conclusion of the research paper.

II. Related Work

The area for detection and diagnosis of diseases has been an active one in recent times. Thus, the research area in medical image classification has garnered a lot of interest and attention.

Rajaraman et al. [4] used a set of Pediatric Chest X-ray images, and used a customized VGG16 model for prediction of Pneumonia from Chest Radio-graphs. The solution used a gradient-based localization using CAM for Visualization purposes.

Zhang X. et al. [5] explored the dataset "Breast Cancer Histopathological Image Classification (BreakHis)" for detecting and classifying breast cancer. The model used ResNet50 along with the CBAM for identifying the breast tumors from the mentioned data set.

Ansari et al. [6] explored the RNSA (Radiological Society from North America) and the CXI (Chest XRay Image) dataset for effective detection of Pneumonia using a ResNet model, and using the method of Transfer Learning approach for making the necessary predictions.

Hashmi et al. [7] described a method of pneumonia from Chest X-rays using Deep Transfer Learning. They used 5 CNNs, namely ResNet18, Xception, InceptionV3, DenseNet121, and MobileNetV3 for said problem.

Li et al. [8] used the RNSA (Radiological Society from North America) dataset, and proposed an improved CNN for pneumonia detection. They utilized the Squeeze and Excitation Networks along with ResNet to handle this problem, also made use of transfer learning to suppress background interference in the interest region of lungs.

Cha et al. [9] used CXI (Chest X-Ray Image) dataset for Pneumonia detection using Attention based Transfer Learning approach. They employed several pre-trained CNN models to extract feature vectors from Chest X-ray Images, then concatenated each feature vector and applied different attention mechanisms. The pre-trained models used were ResNet152, DenseNet121 and ResNet18, and the attention mechanisms of Self-Attention, Efficient Channel Attention as well as Squeeze and Excitation Networks were used.

III. Methodology

This sections deals with the methodology that was used in order to create the model and explains the different working modules. Different modules to be discussed are pre-processing, training the classifier, data augmentation, convolutional block attention module and classification using attention based Residual Network.

A. Pre-Prepossessing

This section deals with the pre-processing steps that are applied on the data set, so that the models used are able to correctly output the class during the training and testing phases. The pre-processing steps involved include changing the shape of the input image – This process involved changing the size of the images to 3x150x150 following the CxHxW convention where:

$$C = Number\ of\ Channels$$

$$H = Height\ of\ the\ Image\ Matrix$$

$$W = Width\ of\ the\ Image\ Matrix$$

This was done so that input images of different sizes could have the same shape, and this helped in faster training of the model as well.

B. Training the Classifier

In this part, the technique of transfer learning was used in the training phase. The problem that is being solved in this paper is the categorical classification problem, in which the input is the Image, and the output is labelled as either "0" (Normal) or "1" (Pneumonia).

The Transfer Learning technique is used in this work, which basically involves a pre-trained model, which is used for classification on one task, to be reused on another related task after fine-tuning.

In this paper, ResNet-50 model, which has been trained on the ImageNet data set, is used as a baseline model for grouping Chest X-ray images into either pneumonia or normal labels. This pre-trained model would be helpful, since the ILSVRC (ImageNet Large Scale Visual Recognition Challenge) consists of a thousand-class classification data set, and since the model used is trained on such a large set of images containing thousand classes, the features which were found to be useful in the prior task can be used in this problem of pneumonia classification as well.

A ResNet model trained using the weights of the ImageNet data set is shown in Figure 2. The following points highlights the main steps used for model creation:

- The initial layers of the model are used to distinguish simple shapes, followed by the middle layers which are used to identify complex shapes and patterns, and the final layers are used to give the output prediction, as shown in Figure 2.
- In our approach, the pre-trained model Res-Net50 is untouched, except for the final layer, which uses a new prediction layer instead. The last layer consists of a vector which contains information that is used for prediction. The addition of new layer consists of two nodes, which is used for prediction of pneumonia ("Pneumonia/Normal").
- Table I shows the hyper parameters that are used during CNN training for Chest X-ray Images (CXI) data set. The most commonly employed loss function of categorical-cross entropy was used to solve the categorical classification problem. Stochastic Gradient Decent (SGD)

Loss Function	Categorical Cross Entropy
Optimizer	Stochastic Gradient Descent (SGD)
Initial Learning Rate	0.001
Number of Epochs	50
Batch Size	64

TABLE I
HYPER PARAMETERS FOR CNN TRAINING FOR CXI DATA SET

with exponential decay was used as an optimizer for the data set. The initial learning rate was set at 0.001 whereas the number of epochs were set to 50 with early stopping, if model did not improve validation accuracy for 3 consecutive epochs. The Batch Size was set to 64.

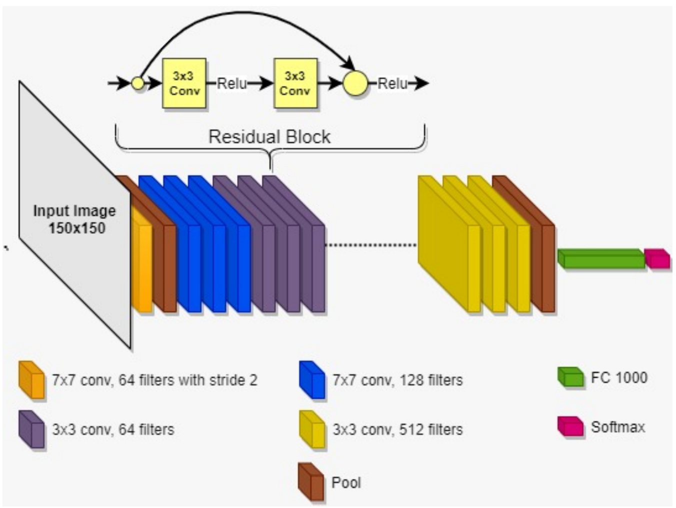

Fig. 2. ResNet-50 Base Model

C. Data Augmentation

The process of Data Augmentation was also applied to decrease loss by increasing the number of Images in the training data set, and modifying them by a re-scaling factor, so that the model can be trained to handle images with different variations as well.

In this paper, 20% of the training data has been used for validation, and hence, it is a part of the validation set. Thus, we're using the 80-20 approach, where 80% of training data becomes a part of the training set and the rest of the 20% becomes part of the validation set.

D. Convolutional Block Attention Module

To make the model more accurate, Convolutional Block Attention Module (CBAM) is also used. CBAM consists of Channel Attention Module(CAM) and the Spatial Attention Module (SAM). The sequential use of Channel Attention Module(CAM), which basically highlights the features in the feature map, and the Spatial Attention Module (SAM), which highlights the Object of Interest within the image, are used to correctly classify the anomalies faced in the input image, thus increasing the accuracy of the model. Besides Global Average Pooling (GAP), Global Max Pooling (GMP) is also used. All of these serve to increase the accuracy of the model, and hence that is why it is used in this model as well.

E. Attention based Residual Network

ResNet50 was used as our base model, because it had a bigger network than other models such as VGG16, and it uses 50 layers to make the network deep and uses residual learning to learn the features more accurately, and classify the anomalies found in the chest X-rays. The VGG16, a smaller network, had a poor performance as compared to ResNet50. Besides this, having a model with too many layers can also decrease the accuracy due to saturation, and also increases the model complexity as well as the training time – Hence, the optimum choice was ResNet50.

ResNet50 uses residual block to train very deep Neural Networks, without degrading the performance. The way ResNet handles this issue is to use a "shortcut" path, so that the input of one layer can directly be fed to another hidden layer, which is usually the second one, and thus skipping the first hidden layer. That's why it's also called a "skip connection". This helps to solve the "Vanishing/Exploding Gradient" problem, that is faced during training of Neural Networks and also helps in resolving the issue of degradation when many layers are used. The ResNet model thus performed better than other SOTA models, and has also won many classification competitions including ILSVRC 2015 classification competition. Figure 3 shows how ResNet50 is transformed into Attention based Residual Network that is used as our enhanced model to detect pneumonia in chest X-rays.

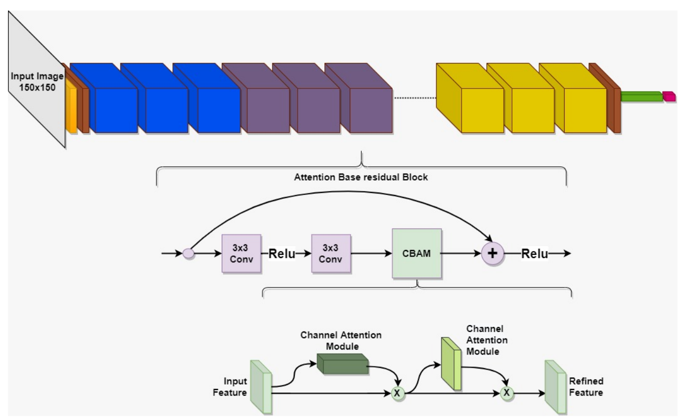

Fig. 3. Attention based Residual Network

IV. RESULTS AND DISCUSSION

The following section gives a brief explanation of the data set that is used and the results obtained from the base model and enhanced model, along with discussion section to highlight the comparison between the two.

Type	Accuracy	Loss
Training	60.37%	0.6872
Validation	77.70%	0.6725
Test	72.43%	0.6780

TABLE II
ACCURACY & LOSS (RESNET50 - BASE MODEL)

	Precision	Recall	F1-Score	Support
Normal	0.79	0.36	0.50	234
Pneumonia	0.71	0.94	0.81	390
Accuracy			0.73	624
Macro avg	0.75	0.65	0.66	624
Weighted avg	0.74	0.73	0.69	624

TABLE III
RESNET-50 CLASSIFICATION REPORT

A. Data Set

The data set that was selected to be utilized in this paper is as follows: "Chest X-ray Images (CXI) for pneumonia detection with deep learning" which is a publicly available data set. This data set was taken from Kermany et al. This data set was created in Guangzhou Women and Children's Medical Center, Guangzhou, China. It contains images of pediatric patients whose age range varies from one to five years old. 5,856 images of chest X-Rays are included in this data set that have been corroborated. The images are further divided into training as well as the testing set of independent patients. Images are named as (disease: NORMAL/BACTERIA/VIRUS)-(randomized patient ID)-(image number of a patient). The images are classified into two groups: either Pneumonia or Normal.

B. ResNet-50 (Base Model)

Table II summarizes the training, validation and test accuracy, as well as losses for ResNet50 base model. The same have also been plotted with respect to accuracy, loss per epoch for training and validation in Figure 4.

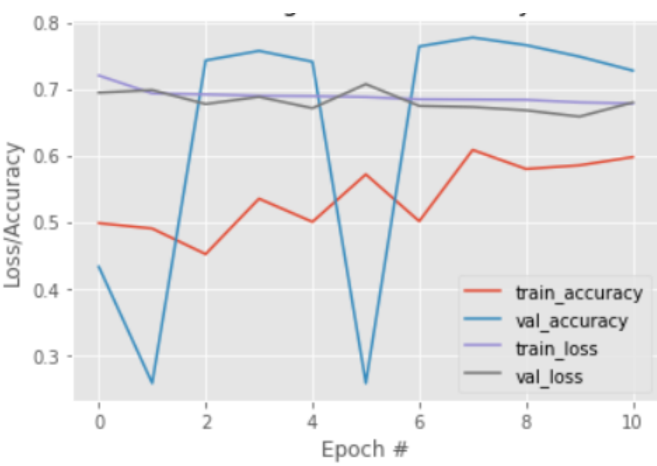

Fig. 4. ResNet-50 Accuracy & Loss (Train, Validation)

Table III summarizes the classification report for the ResNet50 Base Model, whereas the prediction results for some test data images are shown in Figure 5

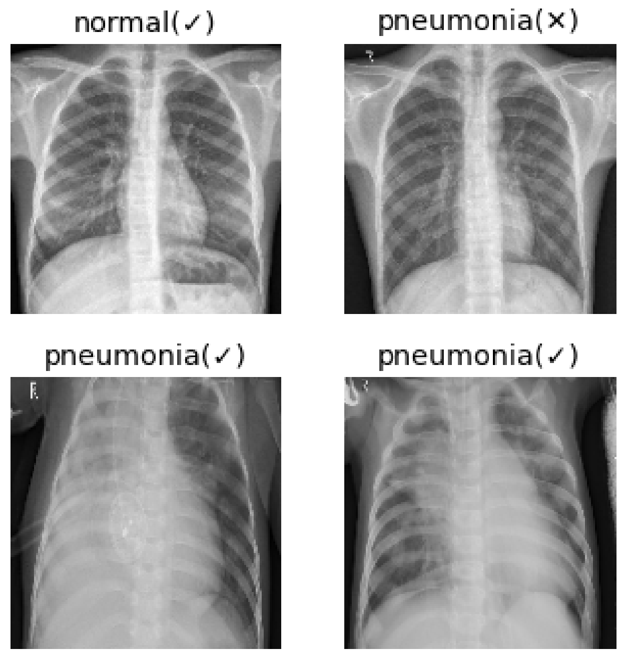

Fig. 5. ResNet-50 Prediction Visual Results

C. Attenion based ResNet50 Model

In a similar fashion, Table IV summarizes the training, validation and test accuracy, as well as losses for Attention based ResNet50 enhanced model. The same have also been plotted with respect to accuracy, loss per epoch for training and validation in Figure 6.

Table V summarizes the classification report for the Attention based ResNet50 Enhanced Model, whereas the predictions results for some test data images are shown in Figure 7

Type	Accuracy	Loss
Training	93.89%	0.6764
Validation	97.27%	0.5968
Test	94.32%	0.6687

TABLE IV
ACCURACY & LOSS (ATTENTION BASED RESNET50 - ENHANCED MODEL)

	Precision	Recall	F1-Score	Support
Normal	0.94	0.76	0.84	234
Pneumonia	0.87	0.97	0.92	390
Accuracy			0.89	624
Macro avg	0.90	0.86	0.88	624
Weighted avg	0.89	0.89	0.89	624

TABLE V
ATTENTION BASED RESNET-50 CLASSIFICATION REPORT

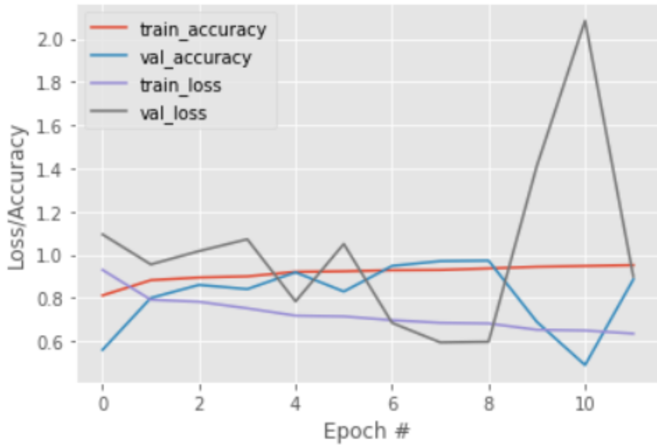

Fig. 6. ResNet-50 Accuracy & Loss (Train, Validation)

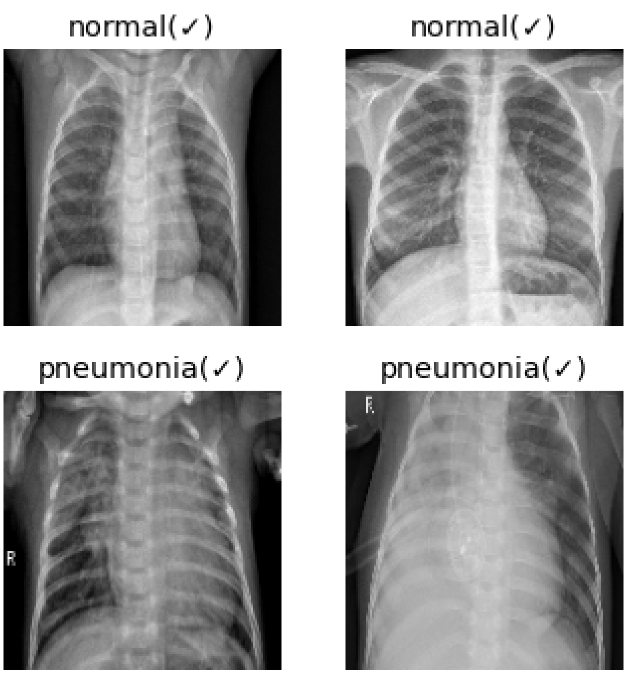

Fig. 7. Attention based ResNet-50 Prediction Visual Results

D. Discussion

In order to create an effective model for detection of pneumonia, various tests were performed on the data set. The split between the training set and validation set were adjusted to 80% and 20% of the training data set, whereas 100% of the test data set was used for the test set. The graph for loss and accuracy for ResNet50 is given in Figures 4, whereas for Attention based ResNet50, it is given in Figure 6.

"Chest X-ray Images (CXI) for pneumonia detection with deep learning" includes 5,856 images, and the best accuracy that was achieved for both ResNet50 and Attention based ResNet50 models were 72.43% and 94.31 % respectively (Tables II and IV. As can be seen from the accuracy stated

Metric	ResNet50	Attention based ResNet50
Loss	0.67	0.66
Accuracy	72.43%	94.31%
True Positives	368	378
False Positives	22	12
True Negatives	85	177
False Negatives	149	57

TABLE VI
METRIC COMPARISON ACROSS TWO MODELS (BASE, ENHANCED)

above, the Attention based ResNet50 performed better than the ResNet50 model with 20% gain in accuracy.

Table VI also shows the comparison between different metrics for both ResNet50 and Attention based ResNet50. Attention based ResNet50 performs better in every metric than ResNet50 model.

Also, in order to cater for the class imbalance between the two classes: Pneumonia and Normal, we assigned class weights during the training of the model in the model.fit() method. The weights that were assigned to the "Normal" images was 1.93, whereas that assigned to the "Pneumonia" class was 0.67. This helped the model to balance the classes, and make no bias, when making the classifications.

Besides this, the process of Data Augmentation was used to increase the overall efficiency of the model. Other networks like ResNet18 and ResNet34 were tested and used initially, but due to the networks being too shallow, the accuracy was not up to the mark. ResNet50 gave considerably high accuracy, and hence, that is why this network model was selected for this problem.

V. Conclusion and Future Work

Thus, the recognition of pneumonia from chest X-rays is an intricate and convoluted problem. Due to this, there is a requirement for a framework for the precise understanding of radio-graphic pictures. Due to the astronomical progress in the field of Deep Learning, it has become feasible to ease the burden of radiologists in detection of diseases, and help them in making better decisions.

The main objective of the research is to create a practical model with improved efficiency to detect and classify pneumonia, which has been completed. According to the Literature Review, "Chest X-ray Images (CXI) for pneumonia detection with deep learning" data set was chosen to build the model for pneumonia detection. In order to get the middle ground between too shallow a network, and too deep a network, ResNet50 was chosen as the baseline architecture.

In order to enhance the model, the addition of Attention based Residual module was used, which further increased the accuracy, due to the Attention module being able to correctly isolate areas where the pneumonia infection occurs within the lungs, and also highlighting that region of pneumonia, thus increasing the overall accuracy of the model.

The reason for Stochastic Gradient Descent over other optimizers has also been explained, wherein a paper published by UC, Berkeley [10] stated that the SGD performed much better in general over other optimizers. The accuracy of 94.31% was achieved using this enhanced model, as compared to the 72.43% accuracy achieved by the baseline model. The accuracy achieved is equivalent to the models in the State of the Art (SOTA) research papers.

It is also important to note that accuracy was computed for Categorical Classification problem, and not for Binary Classification problem, thus it catered the accuracy for both "Pneumonia" detection and "Normal" detection, whereas in many SOTA papers, the accuracy computed was for just "Pneumonia" detection, and hence, it was a binary problem.

The model has achieved satisfactory results in the classification and prediction of pneumonia, as well as identifying and predicting the "normal" cases as well. Some methods to further improve upon this work had been noticed as well. One method could be to use even deeper Neural Network models, like ResNet-101 or ResNet-152 but bigger data sets would be required for that. Since in this paper, we were detecting and classifying pneumonia from chest X-rays, this work can further be extended to detect cancer, and other diseases along with pneumonia, as they also relate to chest-X-ray images.

REFERENCES

[1] "Pneumonia, causes, symptoms and diagnosis." https://www.hopkinsmedicine.org/health/conditions-and-diseases/pneumonia.

[2] "Pneumonia – a preventable disease – kills more children than any other infection." https://www.unicef.org/pakistan/press-releases/one-child-dies-pneumonia-every-39-seconds-agencies-warn, 2019.

[3] S. Naicker, J. Plange-Rhule, R. C. Tutt, and J. B. Eastwood, "Shortage of healthcare workers in developing countries–africa," *Ethnicity & disease*, vol. 19, no. 1, p. 60, 2009.

[4] 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," *Applied Sciences*, vol. 8, no. 10, p. 1715, 2018.

[5] X. Zhang, Y. Zhang, B. Qian, X. Liu, X. Li, X. Wang, C. Yin, X. Lv, L. Song, and L. Wang, "Classifying breast cancer histopathological images using a robust artificial neural network architecture," in *International Work-Conference on Bioinformatics and Biomedical Engineering*, pp. 204–215, Springer, 2019.

[6] N. Ansari, A. Faizabadi, S. Motakabber, and M. Ibrahimy, "Effective pneumonia detection using resnet based transfer learning," *In: Test Engineering and Management*, vol. 82, pp. 15146–15153, 2020.

[7] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, "Efficient pneumonia detection in chest xray images using deep transfer learning," *Diagnostics*, vol. 10, no. 6, p. 417, 2020.

[8] B. Li, G. Kang, K. Cheng, and N. Zhang, "Attention-guided convolutional neural network for detecting pneumonia on chest x-rays," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4851–4854, IEEE, 2019.

[9] S.-M. Cha, S.-S. Lee, and B. Ko, "Attention-based transfer learning for efficient pneumonia detection in chest x-ray images," *Applied Sciences*, vol. 11, no. 3, p. 1242, 2021.

[10] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, "The marginal value of adaptive gradient methods in machine learning," *arXiv preprint arXiv:1705.08292*, 2017.