Deep Learning Techniques to Predict Pneumonia in Chest X-ray (CXR) Images

Mathiazhagan Sampath (x18139973)

Abstract—In recent decade, the deep neural network model has achieved great height in analysis of medical images. Pneumonia is becoming one of the common diseases causing 50,000 deaths and a million to hospitalize each year in the world. Due to this crisis, physicians are undergoing a tedious process in analyzing the chest x-ray images of the pneumonia patients for further diagnosis. In the interest of eliminating this issue, deep learning classification models are proposed to predict the existence of pneumonia in chest x-ray images. Convolutional neural network classification model is implemented to detect pneumonia and the comparison of pre-trained models has been done. Substantial growth in accuracy of convolutional neural network based residual network models has been perceived. The prediction accuracy of 98% in correctly classifying the images has been accomplished with one of the pre-trained residual network models. Furthermore, the machine learning algorithms eludes the manual diagnostic errors and nurtures to strengthen the survival rate of pneumonia patients

Keywords—Pneumonia, Deep Neural Network, Convolutional Neural Network, Classification models, resnet models

I. Introduction

The recent development of medicine field is substantially impacted by machine learning. One of the primary advancements is image processing of clinical diagnosis in case of x-rays and scans. In medical stream, imaging technology of certain part of body to identify specific disease are practiced by doctors in last few decades. The result of the image is visualized by physicians to interpret the severity of the patient and treat them accordingly. One of such disease is pneumonia. Advancement of machine learning approach in the field of image classification assists in several possible ways to predict the pneumonia disease using chest x-ray images of the potential patients.

A. Background and Motivation

Pneumonia accounts for 15% of all deaths of children under five years old, killing 8,08,694 children in the year 2017. The economic cost in preventing pneumonia is estimated to be around $109 million US dollars per year which includes antibiotics and diagnosis. Analysis of medical images is an integral part of medical field by human radiologists [1]. Due to rise in the number of pneumonia patients in every hospital all over the world, the analysis of chest x-ray images becomes a complex process for radiologists. It requires time and financial effort in producing the proper radiologists to provide the report for pneumonia patients. Sometimes there exists a manual diagnosis error during the process leading to incorrect treatment of the patients. The machine learning models comes into the role of classifying the images correctly with intention of eradicating the above-mentioned serious issue.

B. Research Question

”How to predict pneumonia in chest x-ray (CXR) images with improved accuracy using deep learning technique?”

The primary purpose of the project can be summarised with the above-mentioned research question that targets to detect the presence of pneumonia disease in chest x-ray images of potential patients with improved accuracy using deep learning technique.

II. Related Works

Residual Representations in a image recognition, VLAD [2] is a representation that encodes by the residual vectors with respect to a dictionary, and Fisher Vector [3] can be formulated as a probabilistic version of VLAD. Both of them are powerful shallow representations for image retrieval and classification [4]. For vector quantization, encoding residual vectors [5] is shown to be more effective than encoding original vectors. In low-level vision and computer graphics, for solving Partial Differential Equations (PDEs), the widely used Multigrid method [6] reformulates the system as subproblems at multiple scales, where each subproblem is responsible for the residual solution between a coarser and a finer scale. An alternative to Multigrid is hierarchical basis preconditioning [7], which relies on variables that represent residual vectors between two scales. It has been shown [8] that these solvers converge much faster than standard solvers that are unaware of the residual nature of the solutions. These methods suggest that a good reformulation or preconditioning can simplify the optimization. Shortcut Connections. Practices and theories that lead to shortcut connections [9] have been studied for a long time. An early practice of training multi-layer perceptrons (MLPs) is to add a linear layer connected from the network input to the output. In [10], a few intermediate layers are directly connected to auxiliary classifiers for addressing vanishing/exploding gradients. The papers of [11] propose methods for centering layer responses, gradients, and propagated errors, implemented by shortcut connections. In [12], an inception layer is composed of a shortcut branch and a few deeper branches. Concurrent with our work, highway networks [13] present shortcut connections with gating functions. These gates are data-dependent and have parameters, in contrast to our identity shortcuts that are parameter-free. When a gated shortcut is closed (approaching zero), the layers in highway networks represent non-residual functions. On the contrary, our formulation always learns residual functions; our identity shortcuts are never closed, and all information is always passed through, with additional residual functions to be learned. In addition [14], high- 771 way networks have not demonstrated accuracy gains with extremely increased depth (e.g., over 100 layers).

Residual learning actually has been widely used in traditional image restoration methods. A lot of sparse representation based image super-resolution methods try to learn the image high frequency components. Specifically propose dual-dictionary to learn residual iteratively. Part of our work inspired by this multiple residual learning method. After the success of ResNet in image recognition, some works try to learn the residual instead of the undegraded image. Very recently, Dn CNN proposed to learn a CNN network with residual learning for image denoising. Compared with Dn CNN, our proposed method has three very significant advantages. On the one hand, there is no need of any operation of batch normalization that makes our network having fewer parameters that speed up the running time. On the other hand, hierarchical residual learning makes our method more robust.Multiple scale information. In order to avoid the tedious matter of training different model for different magnification image super resolution, train a multi-scale model. With this approach,parameters are shared across all predefined scale factors. However,their network structure does not contain any scale information. Their success maybe should attribute to the powerful learning ability of the convolutional neural network. Propose a Multi-Scale Guided convolutional network for depth map super resolution. But it can be classed as cascade opera- tion because it does not learn the scale information from the input low resolution image.

DCNN models provide a unified feature extraction-classification framework to free human users from the trouble some hand-crafted feature extraction for medical image classification.The author adopted a DCNN to minimize manual annotation and produced good feature representations for histopathological colon cancer image classification and proposed a multi-crop pooling strategy and applied it to a DCNN to capture object salient information for lung nodule classifi- cation on chest CT images.The author trained a DCNN using clinical images for diagnosing the most common and deadliest skin cancers and achieved the performance that matches the performance of 21board-certified dermatologists.The extracted output of the last fully connected layer in apre-trained ResNet-152 model and adopted them to train a custom network layer using the pseudo-inverse method.The integrated two different pre-trained DCNN architectures and combined them into a stronger classifier.Then presented an ensemble of multiple pre-trained ResNet-50 and VGGNet-16model and multiple fully-trained DCNNs by calculating the weighted sum of predicted probabilities. In our previous work[15], we jointly used deep and hand crafted visual features for medical image classification and found that hand crafted features were able to complement the image representation learned by DCNNs on small training datasets. Different from these networks, the proposed SDL model simultaneously takes multiple images as input, and thus enables multiple DCNN components mutually improve each other for learning better discriminative representation.

Research on identification of agricultural pests based on computer vision has been a hot topic. In recent years, many pest recognition systems were proposed.[16] proposed a SIFT-based feature learning method and constructed a feature histogram to classify stone fly larvae images. [15] studied image recognition of pests of sugarcane cotton aphids based on rough set and fuzzy C-means clustering.

The author [17] established an insect automatic classification system by analyzing both the color histogram and the gray level co-occurrence Matrices of the insect wing. proposed a plant pest identification system based on k-means cluster and correspondence filters. [18] used a spatial pyramid with sparse coding to identify farmland pest images. Compared to the early support vector machine and the neural network methods, the recognition accuracy of pest images with background has been improved. To further enhance the insect recognition ability, [19] developed an insect recognition method based on multi-task sparse representation and multi kernel learning techniques.

However, almost all of the above methods require a complex pretreatment of pest images, and the performance of the model is often influenced by characteristics of the selected features. Most of the pest image samples used in earlier studies are images with a uniform background or require removal of the background or binarization. Through the convolutional neural networks, end-to-end training of pest images with background can be achieved, thus greatly simplifying the training process.

III. Objectives

The core objective of the project is to classify the patients having the pneumonia disease in chest x-ray images using the deep machine learning algorithms with high accuracy. The pre-trained residual network models of convolutional neural network classification model are proposed to predict the pneumonia disease with improved accuracy. Convolutional neural network models are the best known and entirely appropriate machine learning algorithms in the field of image classification [20]. The datasets used to perform the proposed model is obtained from Kaggle website which contains the chest x-ray images of both normal and pneumonia affected patients. The datasets are categorised into test, train and validation set with two classes in each set. The train and validation set are used to train and build the proposed classification model. The test dataset is used to evaluate the performance of the proposed model in terms of accuracy and several other evaluation metrics. Further, the report is composed of methodology and implementation of the proposed classification model in the section 2. Followed by that, the evaluation of the model is provided and discussed in detail in the section 3. Succeeded by the conclusion of the project.

IV. Methodology

A popularly known Knowledge Discovery and Data Mining (KDD) methodology is used to build this model. Pattern and valuable insight can be withdrawn from the data using KDD. KDD methodology is of 5 steps.

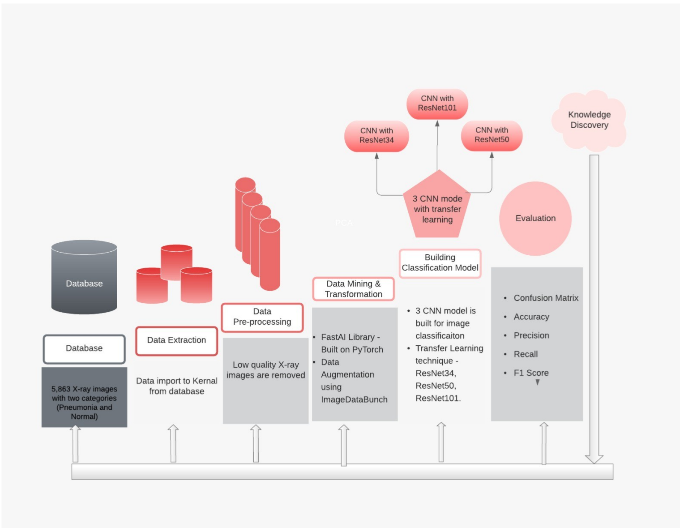

Fig. 1. KDD Methodology for building the Pneumonia image classification model using x-rays

Data selection, Data Pre-processing, Data Transformation, Data Mining, Data interpretation or Evaluation are the steps involved in KDD and Figure 1 illustrates the same. Figure 1 shows the steps involved in building the pneumonia prediction model.

A. Data Selection

The data set used in our study is from Kaggle which consist of Chest X-ray images of patients with pneumonia. There are 5,863 X-ray images are used for our study with 2 categories (Pneumonia/ Normal). Dataset is sliced into 3 folders as train, test and validation and each contains subfolder for pneumonia and normal category.

B. Data Pre-processing

As the data was checked by radiographers to screen the quality of images by removing the low quality x-rays and images are graded by two expert physicians before making the data available. So no pre-processing is required for our study to build our model.

C. Data Mining and Transformation

To transform the data FastAI library is used. FastAI library is built on PyTorch(a framework built on Python). The results obtained from this model are built on FastAI using the Kaggle Kernal with

- 5GB HADRD DISK.
- 16GB RAM, GPU ON to increase the speed of computation.
- Inbuilt data augmentation of FastAI is used

1) FastAI Library: FastAI library is built on PyTorch(a framework built on Python).

- Batch size
  FastAI library is imported and as the images will be processed and GPU will be used for computation, Batch size of 64 will be passed to model as a parameter.
- Image Size
  The recommended size of image for ResNet transfer learning is 224, so the image size is set to 224.
- ImageDataBunch
  After importing the data inside the kernel, data need to be split as train, test and validation. As our data is sliced into train, validation and test folder with pneumonia and normal as categorical subfolders we can use inbuilt ImageDataBunch.from_folder of FastAI library to read the label name from folder automatically.
  To view the images, show_batch function can be used from ImageDataBunch class.
- Data Augmentation
  In FastAI, Data augmentation can be done using ds_tfms. The value for ds_tfms is set as get_transforms(). get_transform perform flip, zoom, warp and lightning transform. Figure 2 shows the images after data augmentaion is done.

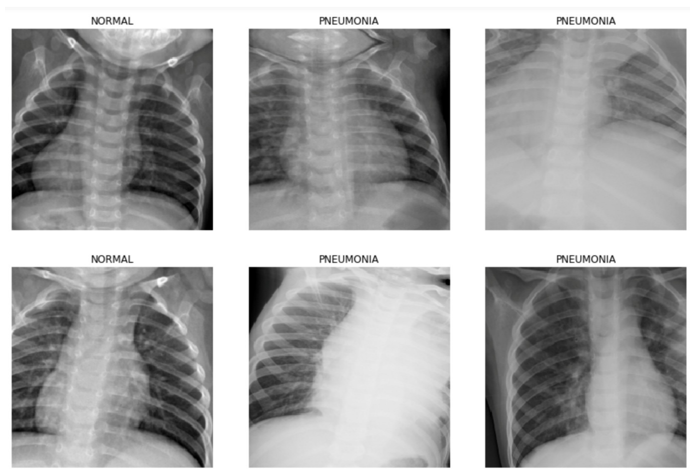

Fig. 2. Data Augmentation of X-ray images

- Normalize the data
  ImageNet parameter is used to normalize the data using .normalize

D. Building the Image Classification Model

Different technique can be used for image classification, CNN outperforms other deep learning techniques in classify- ing the images [1]. So CNN is used to build our model. For transfer learning, ResNet outperformed well on comparing to other transfer learning techniques like VGG, AlexNet [20]. Three different ResNet model has been built using CNN.

1) Convolution Neural Network: Convolution Neural network is the best approach for classifying the images problem

[1]. CNN is built using create_cnn and predefined training model will be passed to train our data.

2) Residual Network (ResNet): By designing a deep neural network accuracy of model can be increased, but accuracy will get saturated and degradation of model will happen if the depth of neural network is increased. This degradation is commonly known as vanishing gradient problem. In worst case it may lead to stop the neural network from future training. To overcome this issue in CNN, Residual network can be used [19], where data from one layer will be passed to the later layers in the network. ResNet models are pre trained on ImageNet dataset.Three ResNet are built with CNN to classify the images. Figure 3 shows the layers of ResNet block and Figure 4 explains the ResNet Architecture.

- ResNet34 - Two Layer deep.
- ResNet50 - Three Layer deep.
- ResNet101 - Three Layer deep

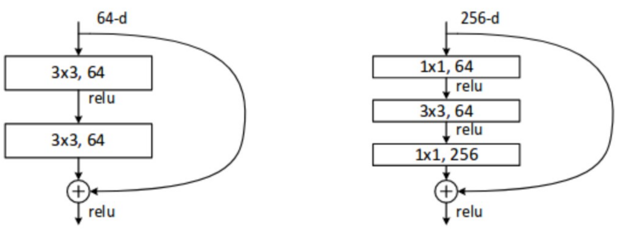

Fig. 3. ResNet 2 layer and 3 layer Block

layer name	output size	34-layer	50-layer	101-layer
conv1	112×112	7×7, 64, stride 2		
		3×3 max pool, stride 2		
conv2_x	56×56	$\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$
conv4_x	14×14	$\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times23$
conv5_x	7×7	$\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax		
FLOPs		3.6×10^9	3.8×10^9	7.6×10^9

Fig. 4. ResNet Architecture

V. Implementation

Three different CNN is created with ResNet transfer learning, Initial layers of the Network are frozen and last fully connected layer weight is learned. Classification Interpretation function of FastAI is used to predict the model classification accuracy is correct or not and top_losses is used to visualize the the image in prediction/actual/loss/probability(Figure 5).

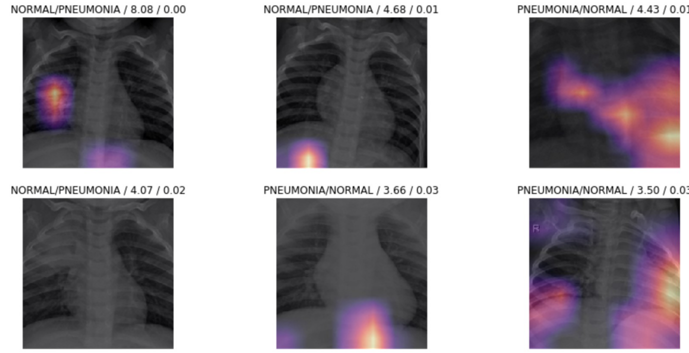

Fig. 5. Top loss function visualization of image

As of now we are concentrated on training the last classification layer now the initial layers need to be trained. To train the initial layers we need to find the ideal learning rate for that FastAI libray function lr__find will be used. lr__find function will run many subset of data for different learning rate and de- termines which learning rate will be best. Using recorder.plot function learning rate will be visulaised and manually learning rate can be determined. By unfreez() function the model can be unfreezed and full model can be trained using the determined learning rate using fit_one_cycle() method.

VI. Results and Discussion

A. Performance Evaluation using Convolution Neural Network (CNN)

In order to find the best practices for Pneumonia detecting from chest X-rays by using CNN-architectures, we performed several experiments and analysed the results using ResNet101, ResNet50, ResNet34. We compared different CNN models in this research in terms of learning rate, accuracy and confusion matrix.

1) Learning Rate Comparison: In this section by taking the learning rate as the evaluation measure, we have compared all three Reset models which we have proposed in this research. Learning rate is consider as the most important hyperparameter which controls over the model each time when the weights are uploaded comparatively to the estimation error. When the value of learning rate is too small the model may get stuck as it ends in the long training process. When the values of the learning rate are too high the learning become unstable and too fast. The learning used in the neural network which basically has the positive value which vary from 0.1 to 1.0. The learning model is been experimented against all there Resnet models. The learning rate is being evaluated using different set of logarithmic scale vary from 1e-01 to 1e-07. In the Resnet34 model the loss is approximately 0.10 when the learning is trained in 1e-06. In the Resnet50 model the loss is seen less when the iteration of learning rate in is 1e-06 which account of 0.125. Comparatively the Resnet101 model shows the minimum loss rate when the model is trained in the learning rate of 1e-06.Figure 6,7 and 8 shows the graph of learning rate against loss for ResNet34, ResNet50, ResNet101.

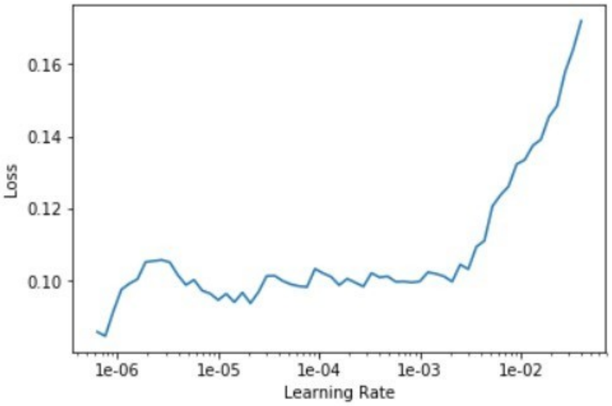

Fig. 6. Leaning rate vs Loss of ResNet-34

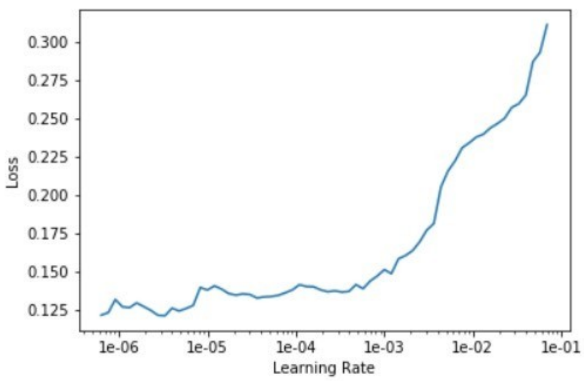

Fig. 7. Leaning rate vs Loss of ResNet-50

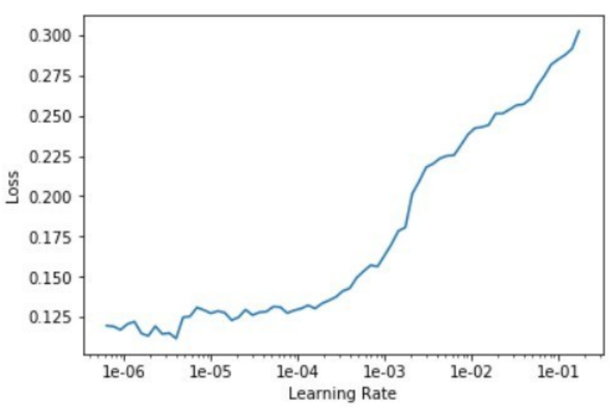

Fig. 8. Leaning rate vs Loss of ResNet-101

2) Batch Processed vs Accuracy: The batch size indicated the number of X-ray samples fed into the neural network. The algorithm works in the way that the first 100 samples will be taken from the training dataset and starts to process the network. Secondly it taken the next set of 100 samples which say 101 to 200 and process the network again. The network is trained until all the samples is been propagated through the network. In the Resnet34 model the accuracy varies when minimum range of sample is propagated to the network. And when the large set of samples is been fed to the network the accuracy became stable and shows the accuracy of 0.95. Over the Resnet50 model when the small sent of sample are propagated to the network the accuracy starts gradually increases and when the large set of samples are being fed the accuracy get decreased rapidly. Hence, we can infer from the graph that the Resnet50 model is not suitable for large set of samples and predicting the pneumonia.

Comparatively in Resnet101 the accuracy less in small set of samples and when the set of batch increases, the accuracy also increases significantly and shows the accuracy result of 0.970 over large batch of samples is propagated to the network. Here from the results we can conclude that the Resnet101 model is well suitable and shows highest accuracy over other models. Figure 9, 10, 11 shows the graph of image processed in batch to accuracy of ResNet34, ResNet50, ResNet101.

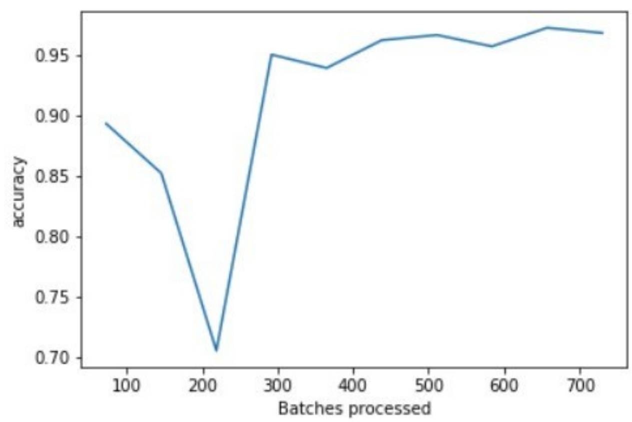

Fig. 9. Batch Processed to Accuracy of ResNet-34

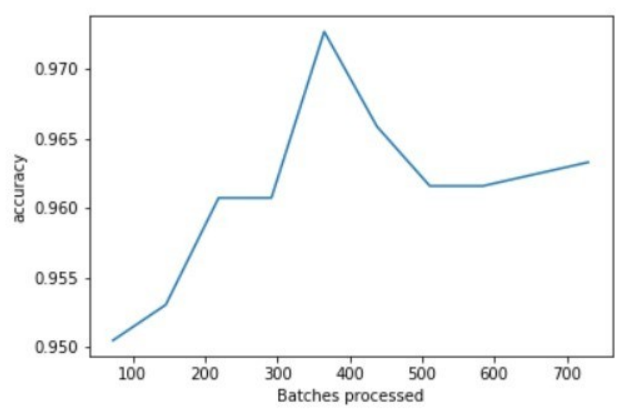

Fig. 10. Batch Processed to Accuracy of ResNet-50

3) Confusion Matrix:

- Accuracy

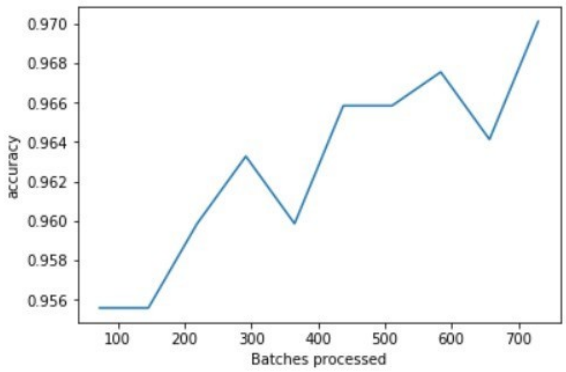

Fig. 11. Batch Processed to Accuracy of ResNet-101

In this research, the results are evaluated using the accuracy of the confusion matrix. Based of the output result of the confusion matrix the whole model is evaluated based on it. There are four different parameters used in confusion matrix which are True Positive (TP), False Positive (FP) which are the predicted class as rows. True Negative (TN) and False Negative (FN) are considered as the actual classes which are aligned in columns. The model is evaluated based on the accuracy from the confusion matrix. Figure 13, 14, 15 shows the confusion matrxi of ResNet34, ResNet50, ResNet101.

		Actual Class	
		Positive	Negative
Predicted Class	Positive	True Positive Count (TP)	False Positive Count (FP)
	Negative	False Negative Count (FN)	True Negative Count (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Fig. 12. Confusion Matrix

- Precision and Recall

The precision and Recall are the other metrices which are used to evaluate the result in this research. Precision measure are used to take the proportion of the positive images which are classified by the proposed model. Re- call are used to take the proportion of the actual positive cases which are classified by the individual models. Confusion matrix is been experiment using all three resent models and calculated the Accuracy, Precision and Recall evaluating the models. Here from the confusion matrix the accuracy, precision and recall has been cal- culated by using their respective formula. The model is been evaluated by comparing the results shown in table.

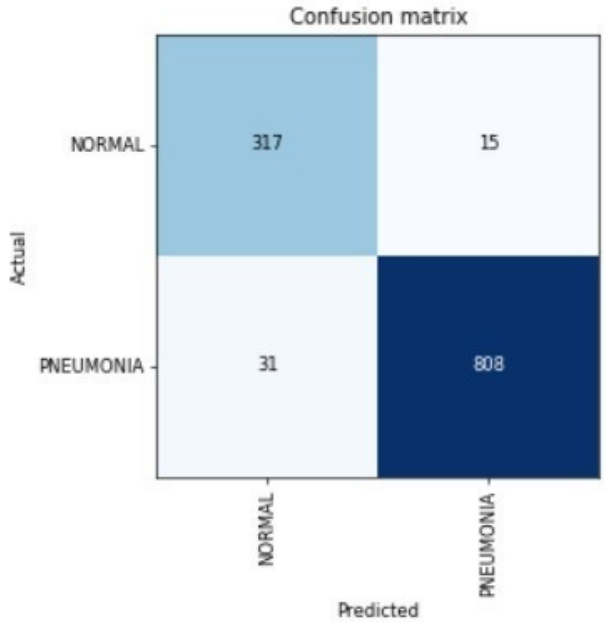

Fig. 13. Confusion Matrix ResNet 34

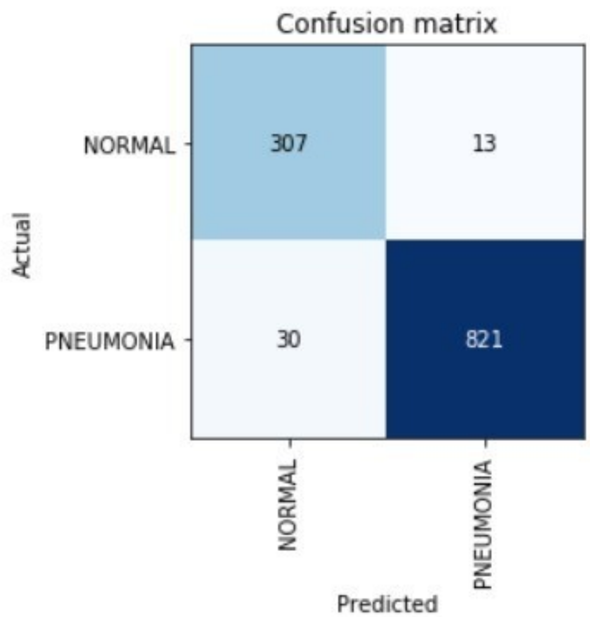

Fig. 14. Confusion Matrix ResNet 50

VII. CONCLUSION AND FUTURE WORK

In some areas there is shortage of radiologist to predict disease accurately, Technology can be use to reduce the time and effort in predicting the pneumonia. Using Deep learning technique in medical field, Human level prediction is achieved. ResNet34, ResNet50, ResNet 101 transfer learning models are used with Convolution Neural network (CNN). Three models had built and Chest X-ray with Pneumonia and Normal images are used to train the model. For training the model,

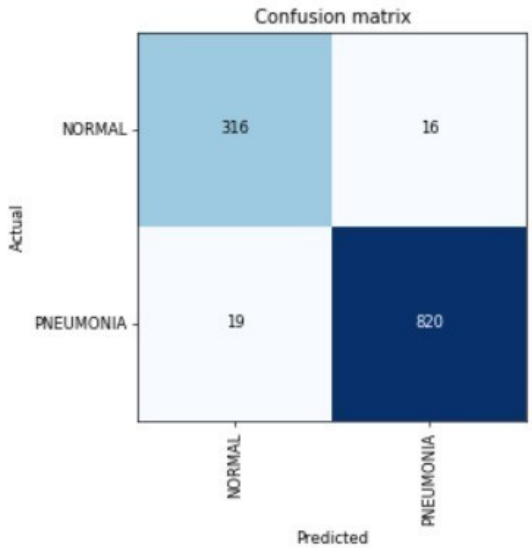

Fig. 15. Confusion Matrix ResNet 101

Model	Precision	Recall	F1-Score	Accuracy
Resnet34	0.98	0.96	0.97	0.96
Resnet50	0.98	0.96	0.97	0.96
Resnet101	**0.98**	**0.98**	**0.98**	**0.97**

Fig. 16. Results of ResNet models

different learning rates are used to improve the performance of the model and it is evaluated using Accuracy, Precision, Recall and F1-score.

From the three models, ResNet101 outperformed ResNet34 and ResNet50. ResNet101 has an accuracy of 97%. This study concludes that deep learning techniques performance is closer to human level precision and these techniques can be used in relevant fields to reduce the time for consulting the radiologist and to get proper treatment.

In future, Different transfer learning model other than ResNet can be used to predict pneumonia or other disease and accuracy can be increased.

REFERENCES

[1] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," Ieee Access, vol. 6, pp. 9375–9389, 2017.

[2] H. Jegou, F. Perronnin, M. Douze, J. Sa´nchez, P. Perez, and C. Schmid, "Aggregating local image descriptors into compact codes," IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 9, pp. 1704–1716, 2011.

[3] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in 2007 IEEE conference on computer vision and pattern recognition, pp. 1–8, IEEE, 2007.

[4] K. Chatfield, V. S. Lempitsky, A. Vedaldi, and A. Zisserman, "The devil is in the details: an evaluation of recent feature encoding methods.," in BMVC, vol. 2, p. 8, 2011.

[5] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in Proceedings of the 18th ACM international conference on Multimedia, pp. 1469–1472, ACM, 2010.

[6] H. Jegou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 1, pp. 117–128, 2010.

[7] W. L. Briggs, S. F. McCormick, et al., A multigrid tutorial, vol. 72. Siam, 2000.

[8] R. Szeliski, "Locally adapted hierarchical basis preconditioning," in ACM Transactions on Graphics (TOG), vol. 25, pp. 1135–1143, ACM, 2006.

[9] R. Szeliski, "Fast surface interpolation using hierarchical basis functions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 6, pp. 513–528, 1990.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9, 2015.

[11] W. N. Venables and B. D. Ripley, Modern applied statistics with S-PLUS. Springer Science & Business Media, 2013.

[12] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in Artificial intelligence and statistics, pp. 562–570, 2015.

[13] T. Vatanen, T. Raiko, H. Valpola, and Y. LeCun, "Pushing stochastic gradient towards second-order methods–backpropagation learning with transformations in nonlinearities," in International Conference on Neural Information Processing, pp. 442–449, Springer, 2013.

[14] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," in Advances in neural information processing systems, pp. 2377–2385, 2015.

[15] Z.-Q. Zhao and D.-S. Huang, "A mended hybrid learning algorithm for radial basis function neural networks to improve generalization capability," Applied Mathematical Modelling, vol. 31, no. 7, pp. 1271–1281, 2007.

[16] N. Larios, H. Deng, W. Zhang, M. Sarpola, J. Yuen, R. Paasch, A. Moldenke, D. A. Lytle, S. R. Correa, E. N. Mortensen, et al., "Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects," Machine Vision and Applications, vol. 19, no. 2, pp. 105–123, 2008.

[17] L.-Q. Zhu and Z. Zhang, "Auto-classification of insect images based on color histogram and glcm," in 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, vol. 6, pp. 2589–2593, IEEE, 2010.

[18] C. Xie, R. Li, W. Dong, L. Song, J. Zhang, H. Chen, and T. Chen, "Recognition for insects via spatial pyramid model using sparse coding," Transactions of the Chinese Society of Agricultural Engineering, vol. 32, no. 17, pp. 144–151, 2016.

[19] C. Xie, J. Zhang, R. Li, J. Li, P. Hong, J. Xia, and P. Chen, "Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning," Computers and Electronics in Agriculture, vol. 119, pp. 123–132, 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.