Proceedings of 12th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2022 Month: October Volume: 12

Detection of Pneumonia Disease based on Digital Chest X-ray Image using Vgg16 Transfer Learning

Chandan Shah [a], Sanjeeb Prasad Panday [b], Tej Prasad Dhamala [c]

[a, b, c] *Department of Electronics & Computer Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*
[a] 076msice006.chandan@pcampus.edu.np

Abstract
Pneumonia is a life threatening lungs disease that is caused by viral or bacterial infection. Life may be at risk especially in case of newly born baby if not acted upon right time. Thus early diagnosis is vital. The aim of this research paper is to detect bacterial and Viral pneumonia using the digital chest x-ray images and analyze result under different data set after tuning different hyper-parameters . In this research paper, we have three classes identified by chest x-ray images: Bacterial Pneumonia chest x-ray, Normal chest x-ray, Viral Pneumonia chest x-ray. The classification accuracy achieved is 92 percent.It can be useful in quick diagnosing of pneumonia by the radiologist . Image classification has become one of the main tasks in the areas of computer vision technologies. Here, we have used a transfer learning-based approach that yields lower reconstruction error and higher classification accuracy.

Keywords
Bacterial Pneumonia, Chest X-ray, Deep CNN, VGG16, Viral Pneumonia, Image processing

1. Introduction

Pneumonia disease can be considered as one of the greatest cause of death of children in all over the world. Approximately 1.4 million children die of pneumonia every year. Pneumonia is a lung infection [1] caused by either bacteria or viral infection. Diagnosis of pneumonia at early phase and correct proper medication can help to prevent worsening of the patient condition. Diagnosing Chest X-ray Image is the best method for diagnosing pneumonia [2]. Deep learning is no doubt the future of technology. In this research paper, we are using technology in deep learning known as "Transfer Learning" which are considerably more improved than the standard deep learning algorithms.With this research, it can contribute to medical science in order to improve accuracy in recognizing various lung diseases. We believe this research will not only help medical professional but will open a new field for research and development. [3]

1.1 Problem Statement

To diagnosis chest X-ray image of diseased Patient is very much time consuming for medical experts to analyse chest xrays of different patients with multiple disease. Deep transfer learning [4] helps in analysis and categorization of chest-xrays. Sometimes chest x-ray image are wrongly classified by the domain experts, which may leads to false diagnosis and wrong medication. Hence health condition of the patient get worsened. Also there is a lack of trained radiologist in rural areas in developing SAARC countries .Therefore, moving towards computer aided diagnosis(CAD) system is important which can help the radiologists in detecting categories of pneumonia from the chest X-ray images. [5] [6].

1.2 Objective

1. The objective of this research paper is to recognize bacterial and Viral pneumonia from chest x-rays using Vgg16 Transfer learing , Pre-trained Imagenet model.

2. A customized VGG16 model consists of two parts, Feature extraction with a CNN layers and pneumonia category classification using fully convolutional network (FCN). Based on Chest X-ray image, We will create a deep learning model which will actually tells us the person having pneumonia or not.

2. Related Works

Transfer Learning with Deep Convolution Neural Network (CNN) for Pneumonia Detection using Chest X-ray [7] Data set – chest xray images consisting of 5247 frontal chest xray images.Data augmentation is used. Four Models are considered: Alex-Net, Resnet18, Dense-Net201, Squeeze-Net. Limitation: Medical history of the associated patient is not considered. Model needs high computational power. It is limited to only two types of disease category of lungs: namely Bacterial Pneumonia, Viral Pneumonia.

Efficient pneumonia detection in chest xray images using deep transfer learning.Diagnostics [8]. Dataset consisting of 5,156 frontal chest xray images belonging to two category: Normal and Pneumonia. Out of 5,156 images, Normal 1283 images and Pneumonia 3873 (Bacterial and Viral) are used for training. 700 images used for testing. Under data augmentation, rotation-45, vertical shift-0.2, horizontal shift-0.15, and Shear-16 is used. Four Models are considered: AlexNet, Resnet18, Inception v3, Xception, MobileNetv2DenseNet201, SqueezeNet Limitation: Medical history of the associated patient is not considered.

Multi-label Transfer Learning for identifying lung diseases using chest xrays. [?] Chest Radiography is used for classifying 14 thorax diseases.Transfer learning ResNet-50 pretrained model on Imagenet dataset used to improve the performance of diagnosis. Data Augmentation is used.

VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper “Very Deep Convolutional Networks for Large-Scale Image Recognition” [9].

CNN is also a type of neural network but in this network, output depends on the volume of the training set.CNN has a major drawback that they are not able to adjust to the viewpoint. If a particular image is inverted, CNN may not be able to identify the picture. CNN is a concept of a neural network which main attributes may consists of convolution layers, pooling layers , activation layers etc. [8]

VGG is a specific convolutional network designed for classification and localization.Like many other popular networks like Google-Net, Alex Net etc. [10].

3. Methodology

One of the finest demonstrated strategy for typically the Convolutional Neural Network. CNN utilizes convolution strategy to extricate highlights from the input information to extend the number of highlights. With increment within the computational power of Graphical Preparing Units (GPU), CNN has accomplished exceptional cutting-edge comes about over a number of regions counting picture acknowledgment, scene acknowledgment, edge discovery and semantic division. Methodology is divided into no of stages illustrated further. a. Data Collection . b. Pre-processing. c. Features Extraction . d. Classification. e. Recognition.

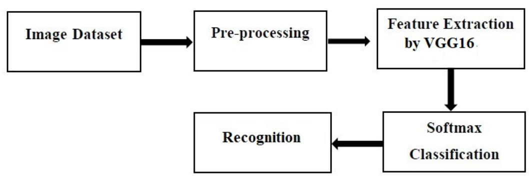


Figure 1: System Block Diagram

3.1 Dataset

At first dataset is collected from different sources and preprocessed is done. Here preprocessing refers to scaling of Image dataset to 224*224 pixels. No other Image processing techniques such as Histogram equalization, contrast stretching etc is done inorder to make model robust. Each dataset is labelled into three categories: Bacterial, Normal, Viral. [13].The Whole dataset is contained in single folder which contains total 6762 Images of chest xray belonging to three category: Bacterial pneumonia, Normal, Viral Pneumonia. Image category information of whole dataset is present in csv file. Out of whole dataset, 80percent of image i.e 5370 Images is used for training the model, 666 Images for internal validation of the model and 726 Image is used for testing the model externally.Testing dataset (726 Images) is not used as the part of training and validating the model.Three csv file is maintained for training, validation and testing purpose which contains Image file name and its category. These csv file are named as: train.csv, test.csv, valid.csv.

3.2 Tools

laptop with 8GB RAM, Jupyter Notebook (Anaconda3), Google colab, Python programming knowledge, Visual Studio Code, Keras API, TensorFlow, scikit-learn, Python's OpenCV library.

3.3 Implementation

Preprocessing: It is one of the important step in the data preparation whose objective is resizing of the X-ray images as input image data to different algorithms is different. For AlexNet and SqueezeNet, the images is resized to 227×227 pixels whereas for ResNet18, DenseNet201, VGG16 the images is resized to 224×224 pixels. All images are resized according to their pre-trained model standards. Data augmentation: It is seen that the data augmentation can improve the classification accuracy of the deep learning algorithms. The performance of the deep learning models can be improved by augmenting the existing data rather than collecting new data. In this research paper , Data augmentation is not done rather new data is collected and training dataset is increased to train the Model. In this paper, PyQt5 Gui widgets tool kit is also used which is most powerful cross platform GUI library. Creating a simple GUI application using PyQt5 involves importing Qtcore, QtGui and QtWidgets modules from PyQt5 package. PyQt5 API is large collection of classes and methods [11].

3.4 Model Evaluation

Evaluation Metric is a technique which is used to evaluate the performance of the model.Once the deep learning models are implemented, they need to be tested to find out the effectiveness of the model based on metrics and datasets. The performance metrics differ depending on the task handled by deep learning Algorithm. Classifcation Model performance metrics are: Accuracy, Precision, Recall, F1-score, ROC, AUC, Confusion Matrix.

A confusion matrix can also be referred to as an error matrix, is an special table structure. In each matrix row, the predicted outcome of class are expressed, whereas each column signifies different class or labels of image dataset.It evaluates quality of output of a classifier on the dataset. The diagonal elements represent the number of data for which predicted label is equal to true label. Off diagonal elements are wrongly classified by classifier. Higher the diagonal element values, greater is the accuracy of the model.

Accuracy Measure: True Positive(TP): The cases which are perfectly classified as Positive. True Negative(TN): The cases which are perfectly classified as Negative. False Positive(FP): The model predicts Positive but it is actually Negative. False Negative(FN): The model predicts Negative but it is actually Positive.

Accuracy = (TP+TN)/(TP+TN+FP+FN) ——Eq(3.4.1)

Eq(3.4.1) tells what proportion of samples both Positive and Negative were correctly classified. For Imbalanced Datasets, accuracy is not a good measure for assessing model performance.It is the function of sensitivity and specificity.

Misclassification Rate = 1 - Accuracy ——Eq(3.4.2) It is also known as "Error Rate".

True Positive Rate: Eq(3.4.3) tells what proportion of actual positives is correctly classified as Positive. It is also called "Sensitivity" or "Recall". It is a measure of completeness. TPR = TP/(TP+FN) ——Eq(3.4.3)

False Positive Rate: Eq(3.4.4) tells when class is actually Negative, how often does it predict Positive. FPR = FP/(FP+TN) ——Eq(3.4.4)

True Negative Rate: Eq(3.4.5) tells when class is actually Negative, how often it predict Negative. It is also called " Specificity".

TNR = TN/(TN+FP) ——Eq(3.4.5)

Precision: Eq(3.4.6) tells what proportion of predicted positive values is actually positive. It can be thought of as a "Measure of exactness".

Precision = TP/(TP+FP) ——Eq(3.4.6)

F1 Score: It is also evaluation performance metric preferable for imbalanced class distribution. For imbalance class distribution, we cannot use accuracy performance matrix. With F1 score, we will be able to judge model more accurately. It is preferred when we are looking for a balanced measure between precision and recall (Type I and Type II errors). F1 Score is suitable measure for classification problems on imbalanced data sets as it is highly sensitive to data distribution. In binary classification, we measure overall F1 score for entire data set but in Multi class classification, we measure F1 score for each class separately. F1 score is the harmonic between precision and recall given by Eq(3.4.7).

F1 Score = 2*(Precision * Recall)/(Precision+Recall) ——Eq(3.4.7)

AUC and ROC is Area Under Curve and Receiver

Operating Characteristics. It tells classification model capability of distinguishing between classes. Using AUC and ROC, we can visualize performance of the Multi class classification problem. AUC represents degree of separability and ROC represents probability curve . Higher the Area Under the Curve, better the model is. To find ROC and AUC, First the Model will predict accuracy measure (TP, TN, FP, FN) based on different threshold value of predicting particular class. This threshold value could have any random number. Hence, this is an arbitrary choice and should not affect the decision provided by our Model. A Method that helps us see how the threshold plays out the decision of the model is produced by ROC curve that is developed at various threshold settings by plotting True Positive Rate (Sensitivity or Recall) Vs False Positive Rate. Activation function: Soft-max function is used as activation function. Optimizer: ADAM Optimizer also known as Adaptive Moment Estimation in which learning rate is not constant.It is computationally efficient, appropriate for noisy or sparse gradient. It requires less memory for operation. Well suited for problems that are large in terms of data or parameters.

4. Results and Discussion

The system is able to classify the 3 different classes of lungs disease category of chest x-ray : Bacterial Pneumonia Chest X-ray, Normal Chest X-ray, Viral Pneumonia Chest X-ray

4.1 Three label Classification and Performance Evaluation under batch size 32 and 64

- Total no of training images: 5370/2803
- Total no of validation images: 666/933
- Total no of test images: 726/936
- Image Size: 224x224
- Weights: image net
- Total no of epoch: 14/15/20
- Batch Size: 32, 64
- Optimizer: Adam Optimizer
- Loss: Categorical Cross Entropy
- Metrics: Accuracy

4.2 Comparative performance of train vs validation accuracy and train Vs validation loss under batch size 32 and 64 under different data set

Initially we have considered training data set of 2803 images belonging to three category: Normal, viral, Bacterial and Finally training data set consisting of 5370 images belonging to same three category: Normal, Viral, Bacterial. New data set were added in initial data set and training dataset was increased from 2803 images to 5370 images.In first case, data set is highly unbalanced data set while in second case data set is nearly balanced data set. Batch Size 32 and Batch Size 64 is considered as good starting point for any deep learning Model.

The comparative performance of train vs validation accuracy and train vs validation loss for batch Size 32 and 64 for different epoch can be seen in Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, Figure 7, Figure 8 and Figure 9 respectively. Data set considered viz training images: 5370, validation images:666, test images: 726.

Figure 2 shows train vs validation accuracy under batch size 32 run for 15 epochs with balanced training data set of 5370 images.

Figure 3 shows train vs validation accuracy under Batch size 32 run for 20 epochs with highly unbalanced training data set of 2803 images.

Figure 4 shows train vs validation accuracy under Batch size 64 run for 14 epochs with balanced training data set of 5370 images.

Figure 5 shows train vs validation accuracy under Batch size 32 run for 20 epochs with highly unbalanced training data set of 2803 images.

Figure 6 shows train vs validation Loss under Batch size 32 run for 15 epochs with balanced training data set of 5370 images.

Figure 7 shows train vs validation Loss under Batch size 32 run for 20 epochs with highly unbalanced training data set of 2803 images.

Figure 8 shows train vs validation Loss under Batch size 64 run for 14 epochs with balanced training data set of 5370 images.

Figure 9 shows train vs validation Loss under Batch size 32 run for 20 epochs with highly unbalanced training data set of 2803 images .

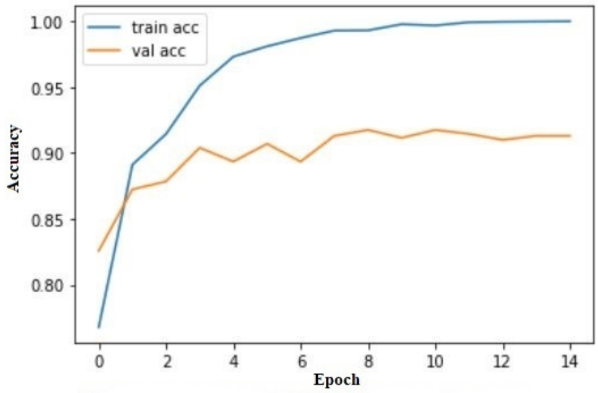


Figure 2: Train Vs Validation Accuracy under Batch Size 32 (TS:5370)

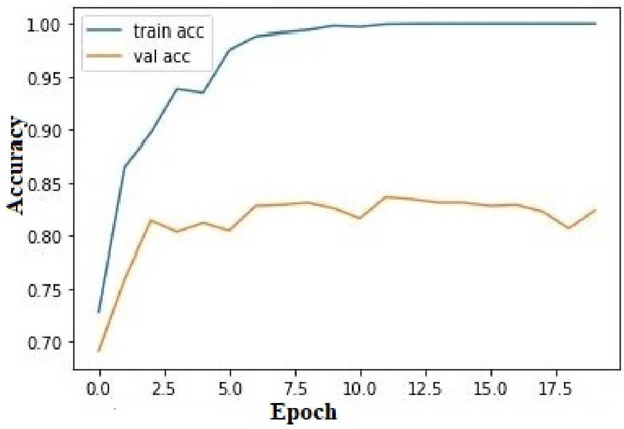


Figure 3: Train Vs Validation Accuracy under Batch Size 32 (TS:2803)

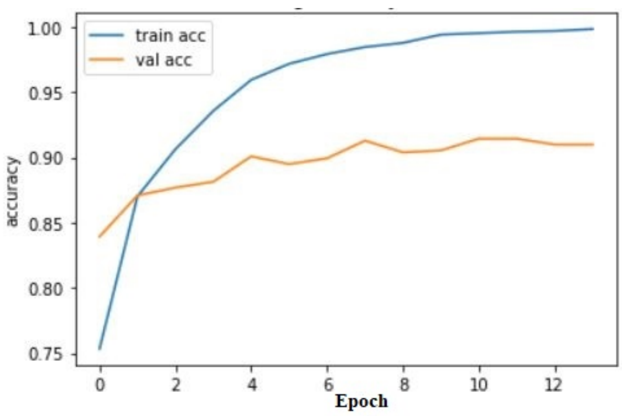


Figure 4: Train Vs Validation Accuracy under Batch Size 64 (TS:5370)

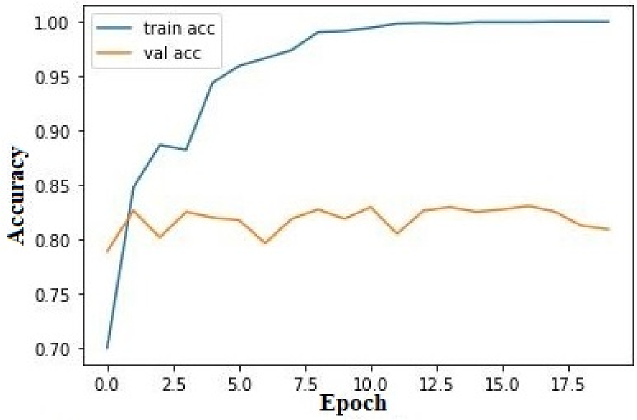


Figure 5: Train Vs Validation Accuracy under Batch Size 64 (TS:2803)

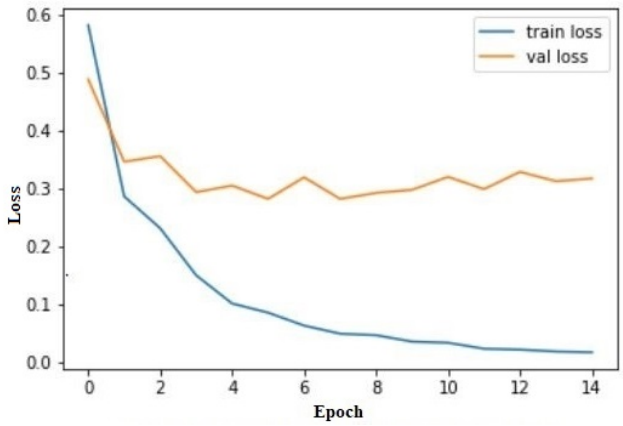


Figure 6: Train Vs Validation Loss under Batch Size 32 (TS:5370)

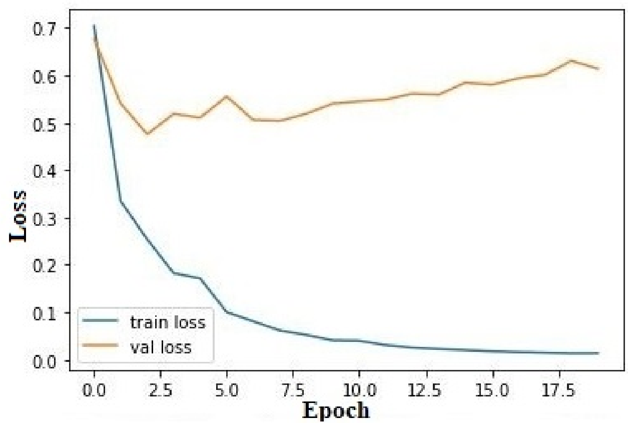


Figure 7: Train Vs Validation Loss under Batch Size 32 (TS:2803)

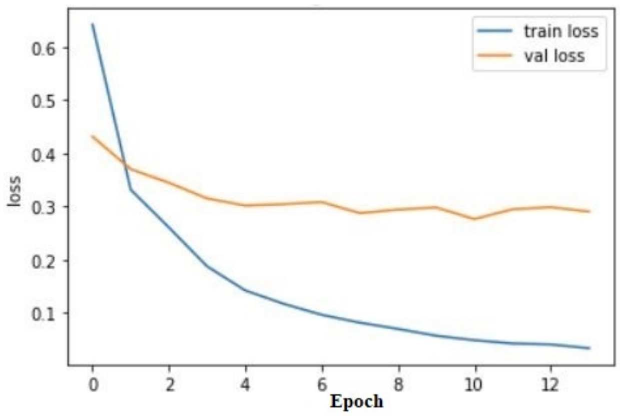


Figure 8: Train Vs Validation Loss under Batch Size 64 (TS:5370)

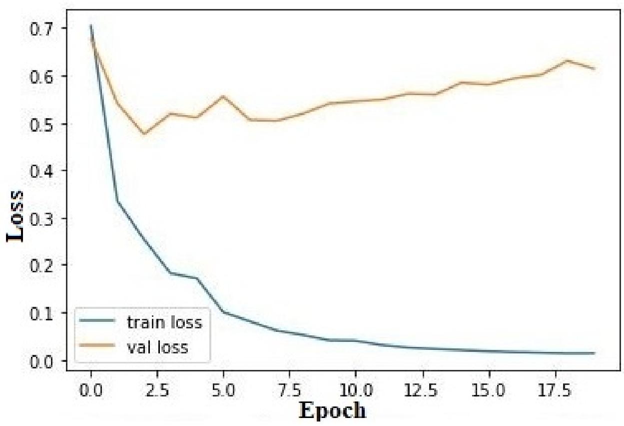


Figure 9: Train Vs Validation Loss under Batch Size 64 (TS:2803)

4.3 Comparison of ROC Curve under Batch Size 32 and 64 for different dataset

In Figure 10 the area under the curve (AUC) /receiver-operating characteristics (ROC) can be seen under Batch size 32 run for 15epochs with balanced training data set of 5370 images. In Figure 11 the area under the curve (AUC) /receiver-operating characteristics (ROC) can be seen under Batch size 32 run for 20 epochs with highly unbalanced training data set of 2803 images. In Figure 12 the area under the curve (AUC) /receiver-operating characteristics (ROC) can be seen under Batch size 64 run for 14 epochs with balanced training data set of 5370 images . Figure 13 the area under the curve (AUC) /receiver-operating characteristics (ROC) can be seen under Batch size 32 run for 20 epochs with highly unbalanced training data set of 2803 images. It can also be called as "AUROC" (area under the receiver operating characteristics)) for different classes, which is one of the most important evaluation metrics for checking any classification model's performance.

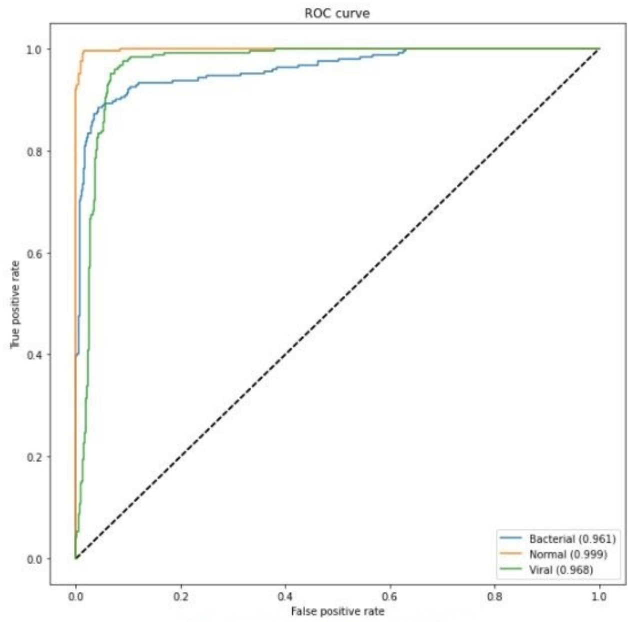


Figure 10: Comparison of ROC Curve under Batch Size 32 (TS:5370)

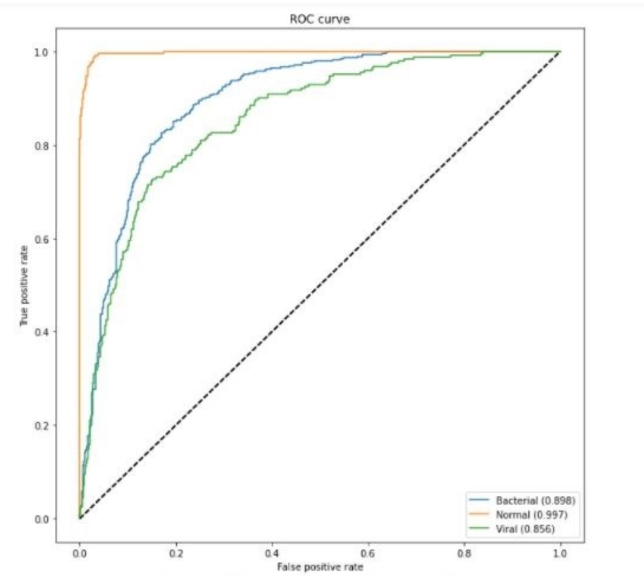


Figure 11: Comparison of ROC Curve under Batch Size 32 (TS:2803)

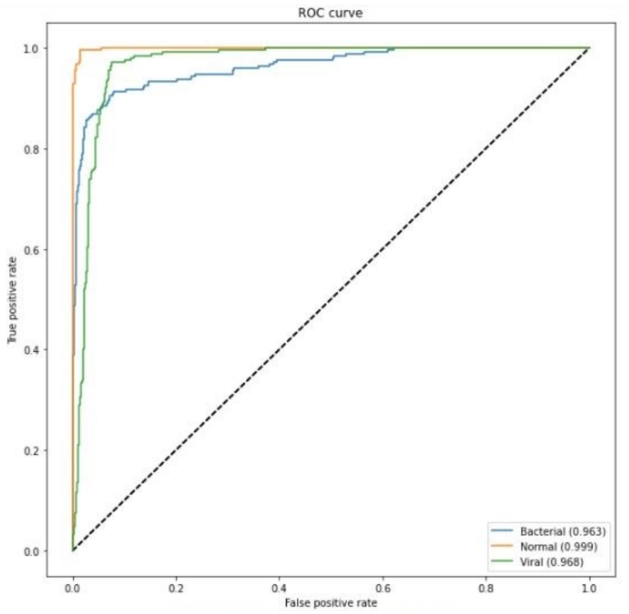


Figure 12: Comparison of ROC Curve under Batch Size 64 (TS:5370)

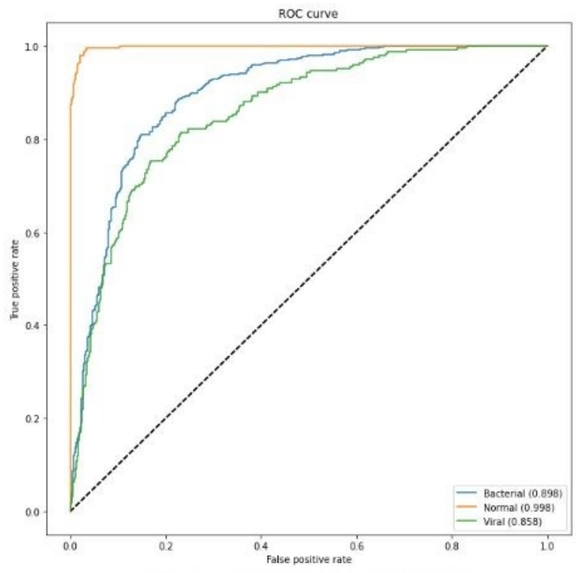


Figure 13: Comparison of ROC Curve under Batch Size 64 (TS:2803)

4.4 Confusion Matrix under Batch Size 32 under Batch Size 32 and 64 for different data set

Table 1 shows the confusion matrix under batch size 32 performed on test data with test data set Image of 726 run for 15epochs with balanced training data set of 5370 images. Table 2 shows the confusion matrix under batch size 32 performed on test data with test data set Image of 936 run for 20 epochs with highly unbalanced training data set of 2803 images . Table 3 shows the confusion matrix under batch size 64 performed on test data with test data set Image of 726 run for 14 epochs with balanced training data set of 5370 images . Table 4 shows the confusion matrix under batch size 32 performed on test data with test data set Image of 936 run for 20 epochs with highly unbalanced training data set of 2803 images.

Table 1: Confusion Matrix under Batch Size 32 (TS: 5370)

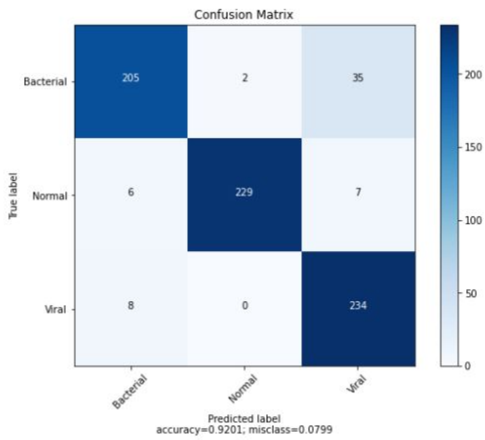


Table 2: Confusion Matrix under Batch Size 32 (TS:2803)

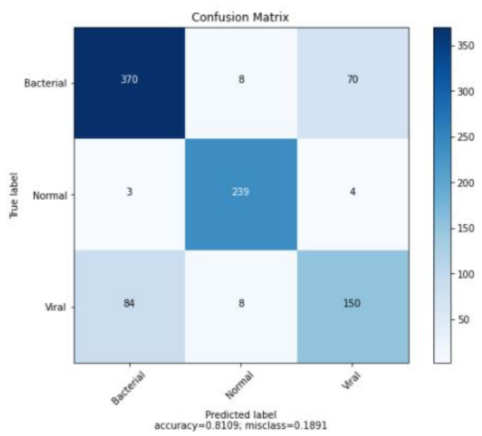


Table 3: Confusion Matrix under Batch Size 64 (TS:5370)

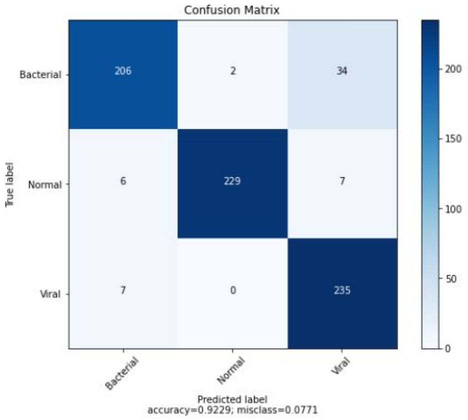
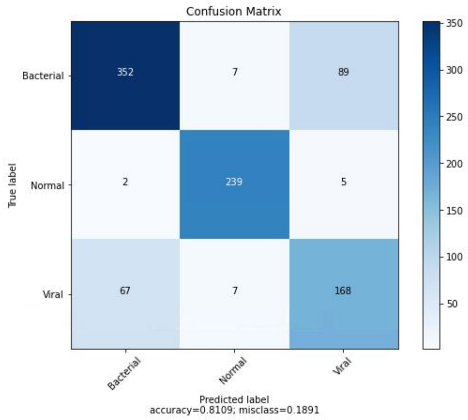


Table 4: Confusion Matrix under Batch Size 64 (TS:2803)



4.5 Classification Report under Batch Size 32 under Batch Size 32 and 64 for different data set

Table 5 shows the classification Report under batch size 32 performed on test data with test data set Image of 726 run for 15epochs with balanced training dataset of 5370 images. Table 6 shows the classification Report under batch size 32 performed on test data with test data set Image of 936 run for 20 epochs with highly unbalanced training data set of 2803 images . Table 7 shows the classification Report under batch size 64 performed on test data with test data set Image of 726) run for 14 epochs with balanced training data set of 5370 images . Table 8 shows the classification Report under batch size 32 performed on test data with test data set Image of 936 run for 20 epochs with highly unbalanced training data set of 2803 images Different performance metrics is obtained from classification report that is shown in Table 5, Table 6, Table 7 and Table 8.

Table 5: Classification Report under Batch Size 32 (TS: 5370)

	0	1	2	accuracy	macro avg	weighted avg
precision	0.936073	0.991342	0.847826	0.92011	0.925080	0.925080
recall	0.847107	0.946281	0.966942	0.92011	0.920110	0.920110
f1-score	0.889371	0.968288	0.903475	0.92011	0.920378	0.920378
support	242.000000	242.000000	242.000000	0.92011	726.000000	726.000000

Table 6: Classification Report under Batch Size 32 (TS:2803)

	0	1	2	accuracy	macro avg	weighted avg
precision	0.809628	0.937255	0.669643	0.810897	0.805509	0.806978
recall	0.825893	0.971545	0.619835	0.810897	0.805757	0.810897
f1-score	0.817680	0.954092	0.643777	0.810897	0.805183	0.808569
support	448.000000	246.000000	242.000000	0.810897	936.000000	936.000000

Table 7: Classification Report under Batch Size 64 (TS:5370)

	0	1	2	accuracy	macro avg	weighted avg
precision	0.940639	0.991342	0.851449	0.922865	0.927810	0.927810
recall	0.851240	0.946281	0.971074	0.922865	0.922865	0.922865
f1-score	0.893709	0.968288	0.907336	0.922865	0.923111	0.923111
support	242.000000	242.000000	242.000000	0.922865	726.000000	726.000000

Table 8: Classification Report under Batch Size 64 (TS:2803)

	0	1	2	accuracy	macro avg	weighted avg
precision	0.836105	0.944664	0.641221	0.810897	0.807330	0.814250
recall	0.785714	0.971545	0.694215	0.810897	0.817158	0.810897
f1-score	0.810127	0.957916	0.666667	0.810897	0.811570	0.811877
support	448.000000	246.000000	242.000000	0.810897	936.000000	936.000000

Acknowledgments

The first author is thankful to Prof. Dr. Shashidhar Ram Joshi, Prof. Dr. Ram Krishna Maharjan, Prof. Dr. Subarna Shakya, Assoc. Prof. Sharad Kumar Ghimire, Asst. Prof. Dr. Babu R. Dawadi, and entire Electronics and Computer Engineering department for their valuable guidance, motivation and support for this research. The first author is also thankful to all the friends for their views and ideas regarding this research paper.

5. Conclusion

This research paper presents transfer learning approach under VGG16 for detection of two types of Pneumonia disease.Initially it was trained and tested on highly imbalanced data set containing total training image 2803 belonging to three category.No satisfactory result was obtained. Later on training data set size was increased to 5370 images almost balanced data set containing three category of images: Bacterial, Normal, Viral. VGG16 algorithms were used to trained model under batch size 32 and 64 then tested for classifying bacterial and viral pneumonia patients using chest x-ray images. It was observed that using nearly balanced data set for two category: Bacterial and Viral pneumonia, model trained under batch Size 64 performs better than the model which is trained on batch size 16 and 32. Overall accuracy of model achieved is 92percent. The classification accuracy, precision and recall of bacterial, viral and Normal class is shown in Table 5, 6, 7 and 8 for Batch Size 32 and 64 for different training and test data set respectively.

References

[1] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Inf-net: Automatic covid-19 lung infection segmentation from ct images. *IEEE Transactions on Medical Imaging*, 39(8):2626–2637, 2020.

[2] Ophir Gozes, Maayan Frid-Adar, Nimrod Sagie, Huangqi Zhang, Wenbin Ji, and Hayit Greenspan. Coronavirus detection and analysis on chest ct with deep learning. 4 2020.

[3] Young-Gon Kim, Kyungsang Kim, Dufan Wu, Hui Ren, Won Young Tak, Soo Young Park, Yu Rim Lee, Min Kyu Kang, Jung Gil Park, Byung Seok Kim, et al. Deep learning-based four-region lung segmentation in chest radiography for covid-19 diagnosis. *Diagnostics*, 12(1):101, 2022.

[4] Dipak Kumar Nidhi, Khushboo Srivastav, Basanta Joshi, and Shashidhar Ram Joshi. Convolutional neural networks for the assessment of fetal echocardiography. 2021.

[5] Tawsifur Rahman, Amith Khandakar, Yazan Qiblawey, Anas Tahir, Serkan Kiranyaz, Saad Bin Abul Kashem, Mohammad Tariqul Islam, Somaya Al Maadeed, Susu M Zughaier, Muhammad Salman Khan, et al. Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. *Computers in biology and medicine*, 132:104319, 2021.

[6] Tawsifur Rahman, Amith Khandakar, Muhammad Abdul Kadir, Khandaker Rejaul Islam, Khandakar F Islam, Rashid Mazhar, Tahir Hamid, Mohammad Tariqul Islam, Saad Kashem, Zaid Bin Mahbub, et al. Reliable tuberculosis detection using chest x-ray with deep learning, segmentation and visualization. *IEEE Access*, 8:191586–191601, 2020.

[7] Tawsifur Rahman, Muhammad EH Chowdhury, Amith Khandakar, Khandaker R Islam, Khandaker F Islam, Zaid B Mahbub, Muhammad A Kadir, and Saad Kashem. Transfer learning with deep convolutional neural network (cnn) for pneumonia detection using chest x-ray. *Applied Sciences*, 10(9):3233, 2020.

[8] Mohammad Farukh Hashmi, Satyarth Katiyar, Avinash G Keskar, Neeraj Dhanraj Bokde, and Zong Woo Geem. Efficient pneumonia detection in chest xray images using deep transfer learning. *Diagnostics*, 10(6):417, 2020.

[9] Very deep convolutional networks for large-scale image recognition. 9 2014.

[10] Chirath Dasanayakaa and Maheshi Buddhinee Dissanayake. Deep learning methods for screening pulmonary tuberculosis using chest x-rays. 12 2020.

[11] Vivian Siahaan and Rismon Hasiholan Sianipar. *LEARNING PyQt5: A Step by Step Tutorial to Develop MySQL-Based Applications*. Sparta publishing, 2019.