Predict Pneumonia on X-ray using Deep Learning Techniques

Summer-2020: CAP 5610- Introduction to Machine Learning

Submitted By:
Fahad Almuqhim
Prasad Bhoite

Submitted to:
Dr. Miguel Alonso Jr., Visiting Associate Professor
(School of Computing and Information Sciences)
Florida International University
July 31, 2020

Table of Contents

Introduction

A) Introduction and Problem Statement:

Pneumonia is an infection that inflames the air sacs in one or both the lungs. It is one of the most common diseases of lungs. As per 2017 National Hospital Ambulatory MEdical Care Survey, there were 1.3 million emergency department visits with pneumonia as the primary diagnosis. In 2017, 49,157 reported deaths due to pneumonia in the U.S with 15.1 deaths per 1000,000 population. The diagnosis of pneumonia is usually based on interpretation of chest X-ray of a patient by the expert. The interpretation of chest X-ray is time consuming, varies by expert, and can be error prone.

B) Metrics:

Our aim was to predict/classify whether the X-ray image is of Pneumonia or No-pneumonia. We have used Convolutional neural network techniques to classify the images. We have used VGG16, InceptionV3, and ResNet50 algorithms for image classification. The best algorithm has been chosen using the accuracy scores on training datasets. Precision, recall, f1-score and support were also calculated.

Data Sources

We have used the “Chest X-ray Images (Pneumonia)” dataset from Kaggle to classify whether the X-ray image is of Pneumonia or No Pneumonia using image based deep learning. The aforementioned data set has data on 5856 X-ray images. Chest X-ray Images dataset is a labelled dataset with two categories. Viz. “Pneumonia” and “No Pneumonia”. The dataset contains 5,840 black and white X-ray images (5216 training images and 624 testing images).

4. Methodology

A) Tools and Libraries used during project:

a) Python (version 3.7):

Libraries:-

- o pandas: for data manipulation
- o matplotlib: Data Visualization
- o seaborn: Data Visualization
- o imblearn: To balance the unbalanced data sets
- o scikit-learn (sklearn): For model selection, confusion matrix

- keras: for deep learning algorithms
- Jupyter Hub: Integrated Development Environment for Python
- Docker: To run applications using containers.

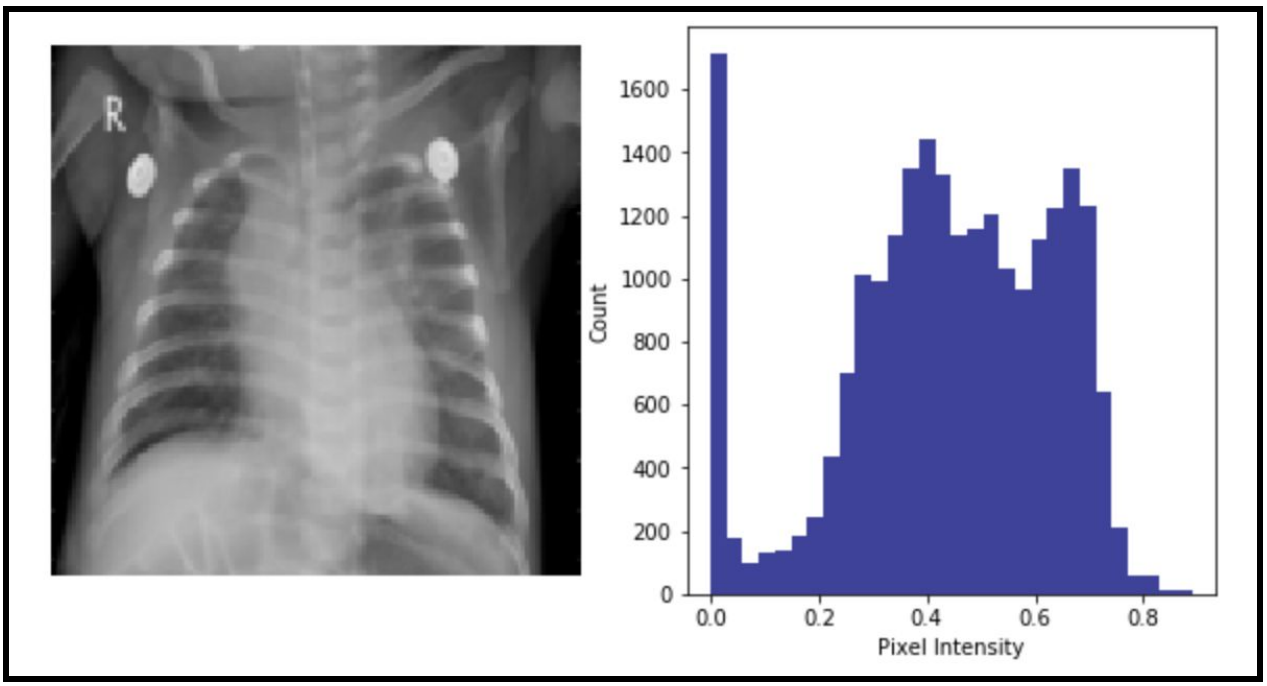
B) Analysis

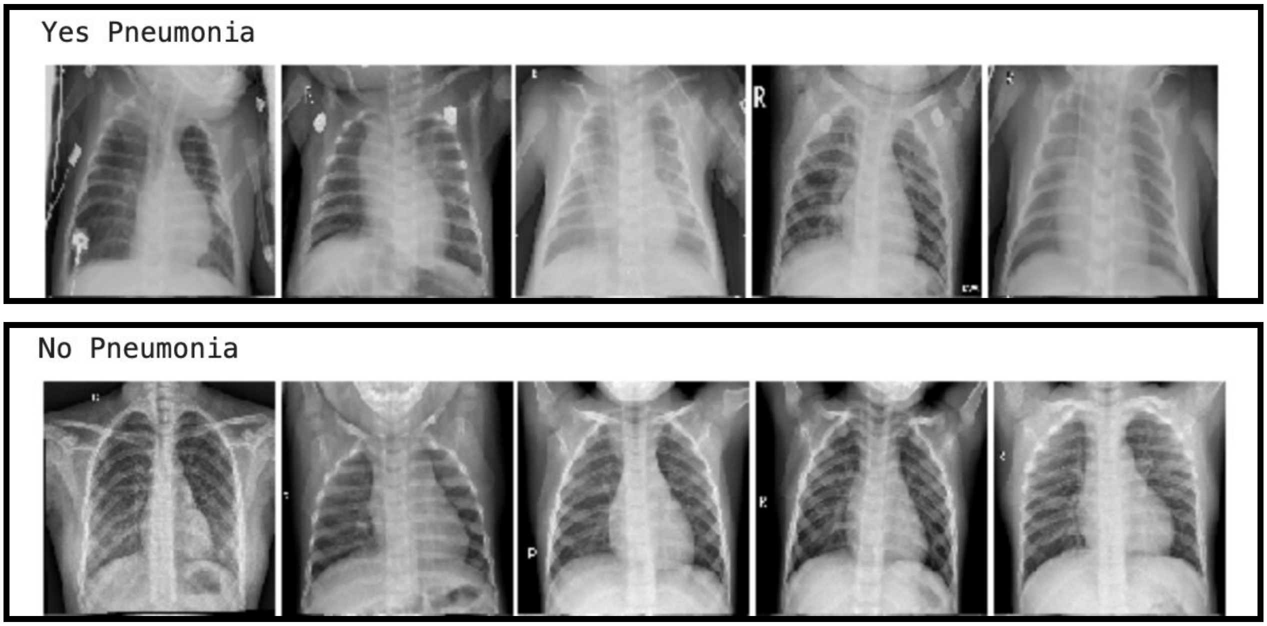
a) Data Preprocessing

All the images were downloaded on a local machine. The downloaded images were first resized into smaller size (150, 150, 3) to reduce the computational cost. As we have Resizing images into smaller sizes, it saves our memory and running time. All the data analyses were performed using Python, version 3.7. using Python. Exploratory data analysis was performed to examine the shape, head, tail, data types, features of the datasets. The training data was further divided into x_train and y_train. The testing data was further divided into x_test and y_test.

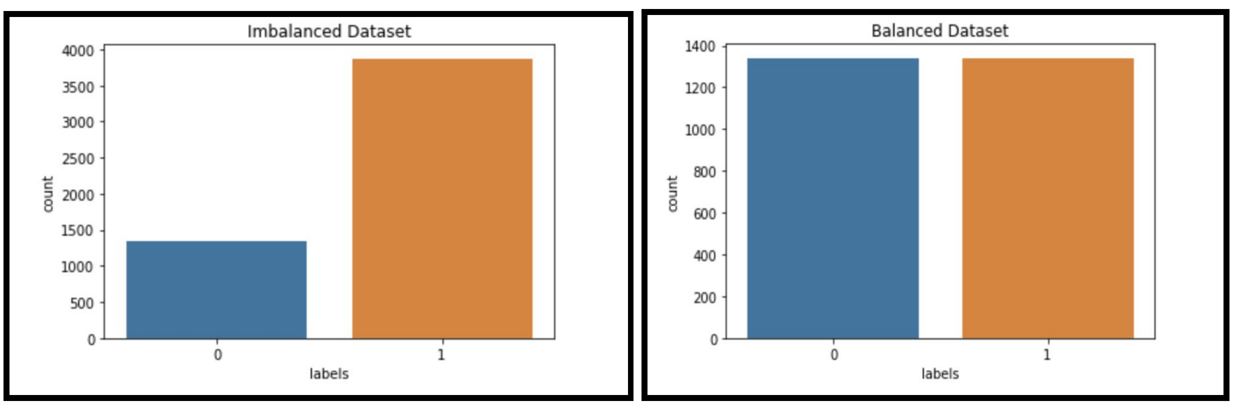
b) Exploratory Visualization

Exploratory data visualization was performed as per the specific aims of the project. Exploratory data visualization was done using different data visualization techniques such as bar graphs, histograms.





c) Balancing the imbalanced datasets:
Our dataset had class imbalance issues. Therefore, we balanced the datasets by under-sampling technique using an imblearn library in Python. The imbalanced dataset had 3875 “Yes-Pneumonia” images and “1341 No-Pneumonia” images. After balancing the dataset, each “Yes-Pneumonia” and “No-Pneumonia” have 1341 images respectively.



d) Implementation
After balancing the datasets, all the datasets were further split into training and testing sets. Pretrained VGG16, InceptionV3 and ResNet50 algorithms were created and applied on training sets independently. Model accuracy scores were measured for each pretrained model. The activation function for predicting the subject is softmax.

e) Model Evaluation, Selection, and Validation:

After we pretrained the models, we run the compile, fit and evaluate functions for each model. The number of epochs was set to 6 which represents the number of times the training data will be passed to the model for training and evaluating. The model is trained using training data and is validated using testing data. The result of each model will be presented in the next section.

Results & Discussion

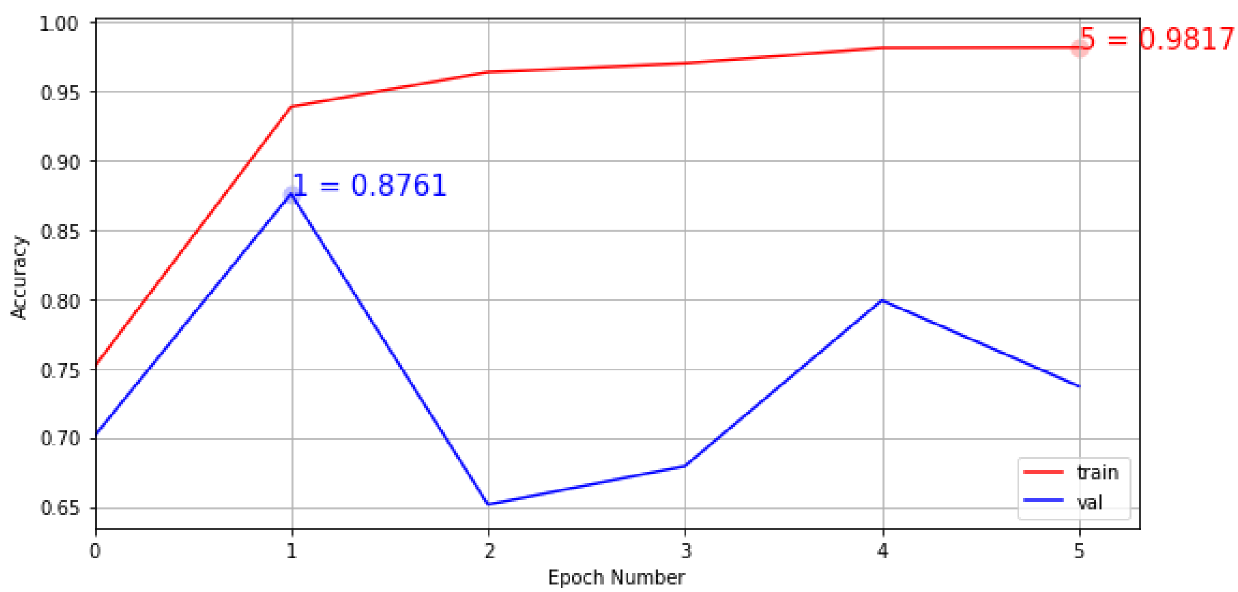
This section is to show the result for each trained model. We will divide our results and discussion based on the trained model as follows

a) VGG16 Model

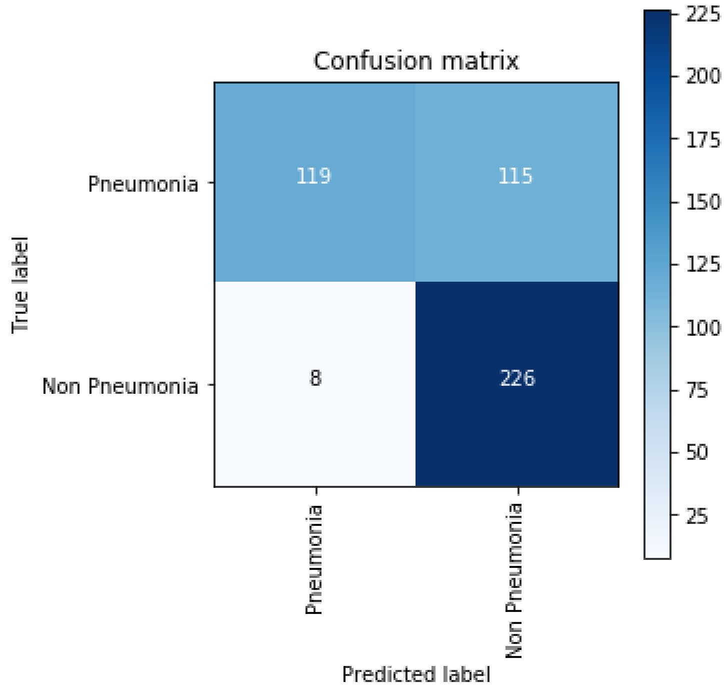
As mentioned above, this model used a CNN model for training. The number of epochs is 6. The accuracy of the model is ranging from 75% to 98%. However, the validation accuracy is ranging from 65% to 87% with an average accuracy of 74%. The difference between the training accuracy and testing accuracy shows that the model will need more data to be built in order to train the model so that will prevent the overfitting. The model summary result is as follows:

	Precision	Recall	f1-score	Support
Pneumonia	0.94	0.51	0.66	234
Non Pneumonia	0.66	0.97	0.79	234
Accuracy			0.74	468
Macro avg	0.80	0.74	0.72	468
Weighted avg	0.80	0.74	0.72	468

The following chart shows the accuracy of the training model and the validation accuracy for each epoch:



From the above chart, we can not get a clear view to decide if we should increase the number of epochs to get better accuracy. Thus, the only solution in hand would be to maximize the number of normal X-rays. This kind of benchmark is usually available since any kind of chest problem that is diagnosed by X-ray is common.
The confusion matrix of this model is as follows:



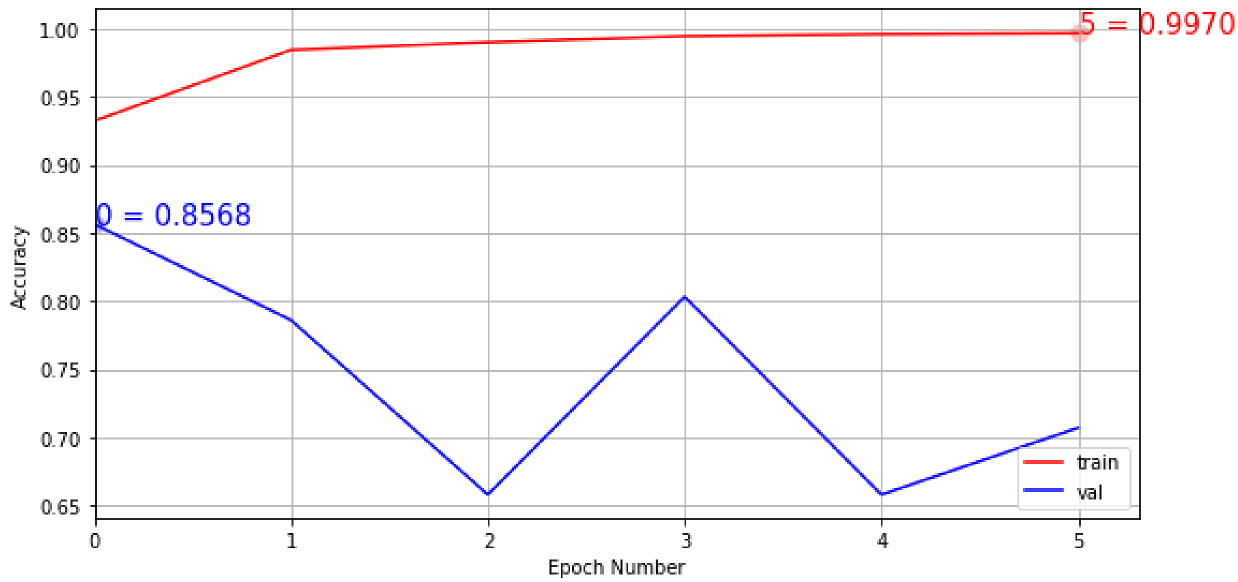
From the confusion matrix, it can be seen that the accuracy is affected by mislabeling non pneumonia subjects. As a solution for this would be to increase the dataset especially for non pneumonia class. This increase will give the model a chance to learn more about non pneumonia subjects.

b) InceptionV3 Model
As mentioned above, this model used a CNN model for training. The number of epochs is 6. The accuracy of the model is ranging from 93% to 99%. However, the validation accuracy is ranging from 65% to 85% with an average accuracy of 70%. The difference between the training accuracy and testing accuracy shows that the model will need more data to be built in order to train the model so that will prevent the overfitting. The model summary result is as follows:

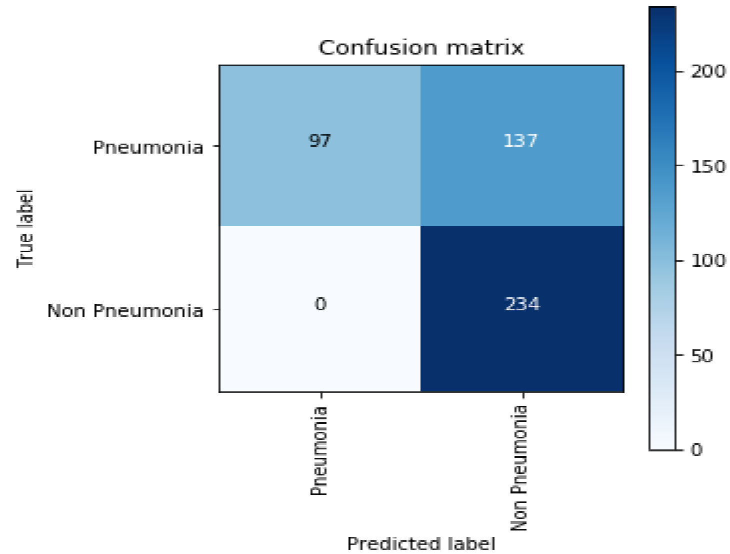
	Precision	Recall	f1-score	Support
Pneumonia	1.00	0.41	0.59	234
Non Pneumonia	0.63	1.00	0.77	234

Accuracy			0.71	468
Macro avg	0.82	0.71	0.68	468
Weighted avg	0.82	0.71	0.68	468

The following chart shows the accuracy of the training model and the validation accuracy for each epoch:



From the above chart, we can not get a clear view to decide if we should increase the number of epochs to get better accuracy. Thus, the only solution in hand would be to maximize the number of normal X-rays. This kind of benchmark is usually available since any kind of chest problem that is diagnosed by X-ray is common.
The confusion matrix of this model is as follows:



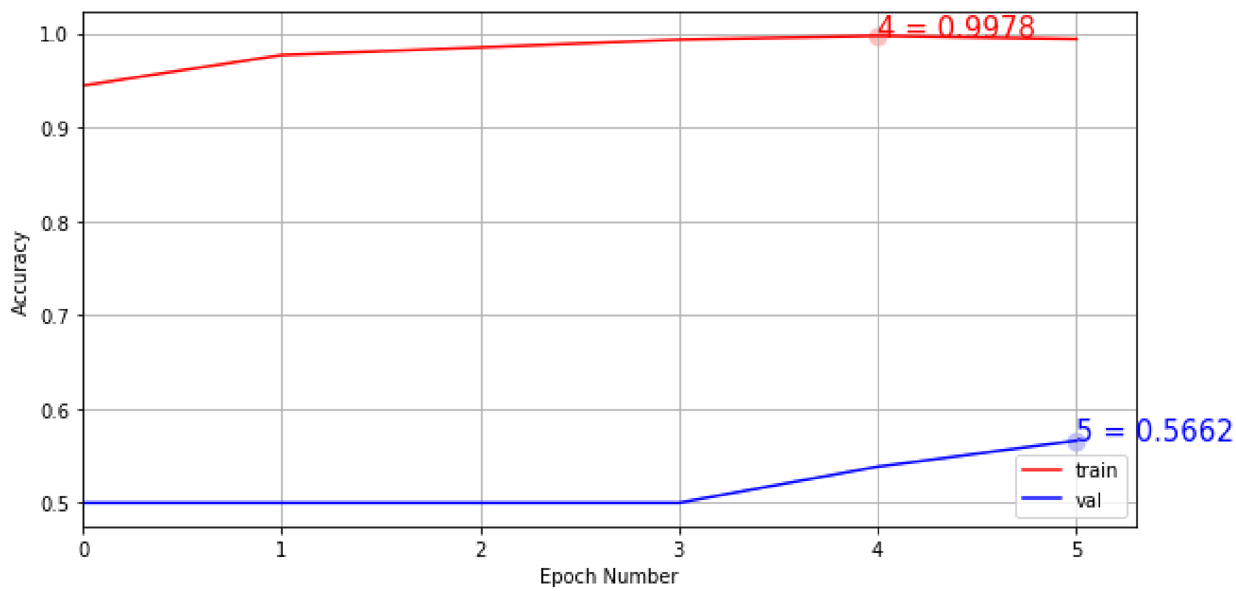
From the confusion matrix, it can be seen that the accuracy is affected by mislabeling pneumonia subjects. The dataset was imbalanced, thus we had to under-sampling normal subjects. We could find normal chest X-rays as these are publicly available to make our data balanced. This increase will give the model a chance to learn more about normal chest x-rays.

b) ResNet50 **Model**

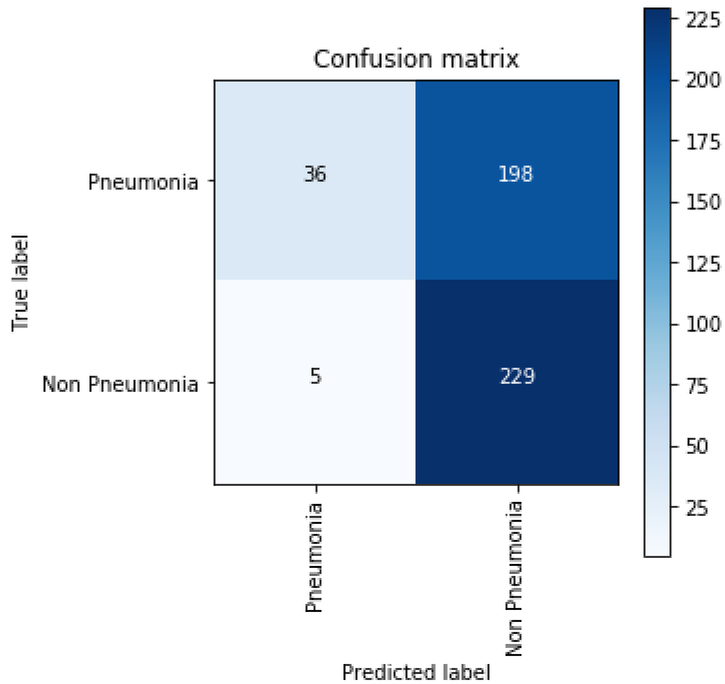
This model is the least performing model among the other two models. This explains that this kind of model could not be a right choice to solve this problem. As mentioned above, this model used a CNN model for training. The number of epochs is 6. The accuracy of the model is ranging from 94% to 99%. However, the validation accuracy is ranging from 50% to 58% with an average accuracy of 56%. The difference between the training accuracy and testing accuracy shows that the model will need more data to be built in order to train the model so that will prevent the overfitting. The model summary result is as follows:

	Precision	Recall	f1-score	Support
Pneumonia	0.88	0.15	0.26	234
Non Pneumonia	0.54	0.98	0.69	234
Accuracy			0.57	468
Macro avg	0.71	0.57	0.48	468
Weighted avg	0.71	0.57	0.48	468

The following chart shows the accuracy of the training model and the validation accuracy for each epoch:



From the above chart, it can give a clue that the number of epochs increased the validation accuracy would be also increased, However, for the time being and lack of resources, the number of epochs can not be increased at this time, and it will be one of the future work if we decide to continue training the model using ResNet50. The other models give better result while the time processing would be less.
The confusion matrix of this model is as follows:



From the confusion matrix, it can be seen that the accuracy is affected by mislabeling pneumonia subjects. As a solution for this would be to increase the dataset especially for pneumonia class. This increase will give the model a chance to learn more about pneumonia subjects.

Conclusions

In this project, we were able to predict/classify whether the X-ray image is of Pneumonia or No-pneumonia. Convolutional neural network technique was used to classify the images. We have used VGG16, InceptionV3, and ResNet50 algorithms for image classification. The results of VGG16 and InceptionV3 were similar. However, they need some improvements such as increasing the dataset. ResNet50 is not performing well compared with the other algorithms. In the future, the models can be improved by providing more data to the dataset. There are several ways to do the increase. One technique is to collect more data which requires time and effort, the other technique is to do data augmentation. The future work should include implementing data augmentation to increase the data and tr-train the model again. We think this would give a better accuracy that we are getting now.

References

1. Pneumonia Fast Facts: https://www.cdc.gov/nchs/fastats/pneumonia.htm
2. Data Source: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
3. G. Labhane, R. Pansare, S. Maheshwari, R. Tiwari and A. Shukla, "Detection of Pediatric Pneumonia from Chest X-Ray Images using CNN and Transfer Learning," 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), Jaipur, India, 2020, pp. 85-92, doi: 10.1109/ICETCE48199.2020.9091755.
4. H. Sharma, J. S. Jain, P. Bansal and S. Gupta, "Feature Extraction and Classification of Chest X-Ray Images Using CNN to Detect Pneumonia," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 227-231, doi: 10.1109/Confluence47617.2020.9057809.
5. Ambita A.A.E., Boquio E.N.V., Naval P.C. (2020) Locally Adaptive Regression Kernels and Support Vector Machines for the Detection of Pneumonia in Chest X-Ray Images. In: Nguyen N., Jearanaitanakij K., Selamat A., Trawiński B., Chittayasothorn S. (eds) Intelligent Information and Database Systems. ACIIDS 2020. Lecture Notes in Computer Science, vol 12034. Springer, Cham