FINAL REPORT – GROUP 29

Haobo Wang
Student# 1006669384
hbo.wang@mail.utoronto.ca

Ariana Lin
Student# 1007113053
ariana.lin@mail.utoronto.ca

Rosalind Wang
Student# 1006930519
rosalind.wang@mail.utoronto.ca

Muchen Liu
Student# 1006732145
muchen.liu@mail.utoronto.ca

ABSTRACT

In this report, group 29 would demonstrate their progress and results about the Pneumonia X-ray image classification of the machine learning project. —-Total Pages: 7

1 INTRODUCTION

Pneumonia, as it has been discovered by human beings, has been affecting millions of people every year, and brought about four million people to death. In many developing countries, pneumonia is still the major cause of death among old and young people. The goal that the team wants to achieve is to develop a machine learning model to diagnose pneumonia is that by training a model to detect pneumonia through X-ray images could improve the efficiency and the accuracy of diagnosis compared to doctors checking it themselves. Currently, the model is forged with a convolutional neural network (CNN) to extract the features from the image imputed, after the feature extraction and flattening, the output would be connected to an artificial neural network (ANN) which contains two fully connected layers for classification.

2 ILLUSTRATION / FIGURE

The purpose of the model is to classify chest X-rays into two classes, which are X-rays with pneumonia and without pneumonia. The team plans to use a convolutional neural network (CNN) to extract certain features from X-ray images and connected an artificial neural network to it for classification. The figure below briefly shows the overall idea of our project. A more detailed demonstration of the model can be found in a later part.

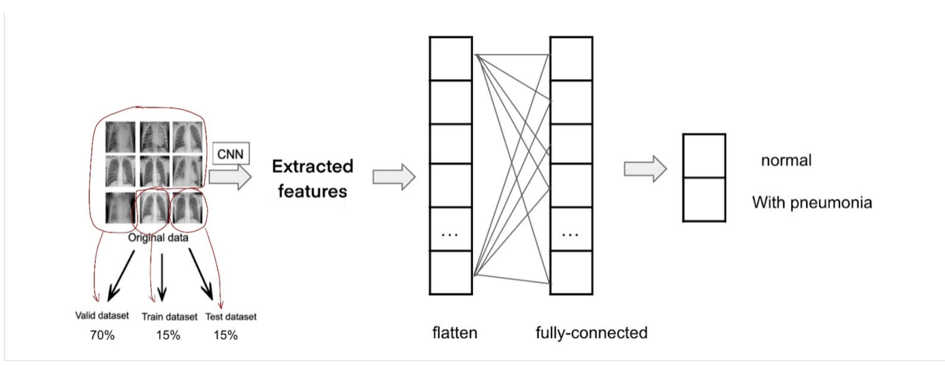

Figure 1: Brief structure of the model(CNN + ANN)

3 Background and Related Work

Pneumonia detection has been a problem for years. In a traditional convolutional neural network, the structure of the model is convolutional layers, activation functions, pooling layers, flattening and fully connected layers, and algorithms (Kaushik et al., 2020). Amongst the steps, algorithms of CNN classifiers as the core of the program attract various types of deep learning methodologies. Four approaches are used as transfer learning to detect pneumonia of COVID-19, which are AlexNet, ResNet18, DenseNet20, and SqueezeNet (Tawsifur Rahman et al., 2020). The comparison and visualization of each approach help the team have a better understanding of the architectures of each of the approaches while providing the team with some potential options for the structures of the model. Besides, GoogLeNet applies inception blocks with dimension reduction are introduced in order to control the computational complexity. The improvement of the performance can be evidenced by the result of the final accuracy (Rohit Kundu et al., 2021). In addition, the VGG19 network can be used to develop the proposed Deep CNN by improving its accuracy to as high as 0.98 (Rajasenbagam et al., 2021). This paper introduces Convolutional Generative Adversarial Network (DCGAN) to create augmented images, which is especially inspiring and useful for the team (Rajasenbagam et al., 2021). Last but not least, the detailed explanations in the research help the team explore the usages of the confusion matrix. The calculation and analysis of accuracy, precision, recall, F1 score, ROC (receiver operating characteristic) and AUC (area under the curve), etc are all useful tools to translate numeric data into helpful information for improvement (Kareem et al., 2022).

4 Data Processing

The data of chest X-Ray images of pneumonia is from patients aged 1 to 5 years from Guangzhou Women's and Children's Medical Center by accessing the public data platform, Kaggle, published by PAUL MOONEY in 2018(Mooney, 2018) . Although the data is already divided into testing, training, and validation including normal chest X-Ray images and pneumonia chest X-Ray images, the team reclassified the images for testing, training, and validation in a more proper proportion.

The chest X-Ray images from Kaggle published by PAUL MOONEY are in JPEG format, which basically has no effect on the project compared with JPG format. There are 5856 images in the dataset, with 1583 normal and 4273 pneumonia. Since the inputs of the neural networks should be of the same size, the team resized images into the same size to easily get these images loaded and processed. As exceeded shrinking will lead to deformation of features and patterns inside the image, the team resized the original images with dimensions of 1500 *1400 to 100*100.

After cleaning and resizing all the images to a considerably smaller size (100*100), the team decided to use 70% of each class to be in the training set, 15% to be in the validation set, and 15% to be in the testing set. Therefore, there would be 1110 normal and 2990 pneumonia in the training dataset; 237 normal and 642 pneumonia in the validation dataset; and 236 normal and 641 pneumonia in the test dataset.

To ensure randomness in the splitting process, the team randomly selected the images that appear first from the datasets. For example, the team used the first 1110 normal images from the datasets as the training dataset.

Since machine learning algorithms cannot operate on label data directly, the team converted all input variables and output variables to be numeric. The one-hot encoding would be applied to change the label 'normal' to 0, and 'pneumonia' to 1.

5 ARCHITECTURE

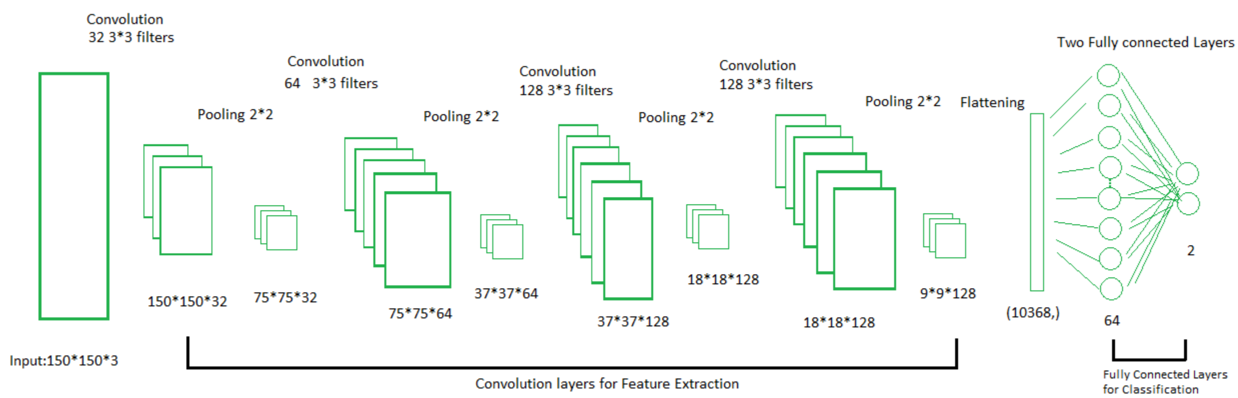

Figure 2: Architecture of model

In this project, there are two categories (pneumonia/normal) of images of chest X-rays in the database with different features. In order to extract the correct image features related to the chest state from the input images, the team first uses the convolutional neural network (CNN) that applies kernels through a total of four convolutional layers to extract features from chest X-ray images, and learn the arrangement, and the relationship between these features. Each pooling layer between convolutional layers helps the model to keep the parts of the image that have valid information, and also simplifies the pixel metrics of the image that sends to the next layer.

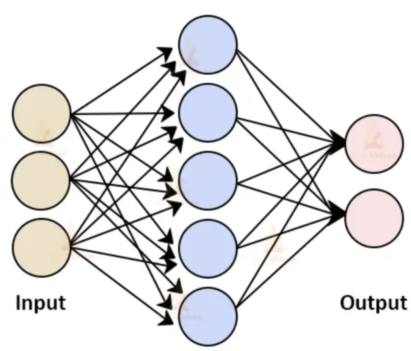

Figure 3: Architecture of an Artificial neural network (Agrawal, 2021)

After the CNN, the team connects an artificial neural network (ANN) that consists of two fully connected layers for further classification. As the figure shows above, an ANN would process the input image to classify the inputs into the desired set of outputs. The images will be flattened from two-dimensional images to one-dimensional images to avoid making ANN too complex. For the classification result, after processing through the last layer in the ANN, the model would give an output from 1 which represents the image is identified as pneumonia or 0 which demonstrates that the image belongs to a normal person.

6 BASELINE MODEL

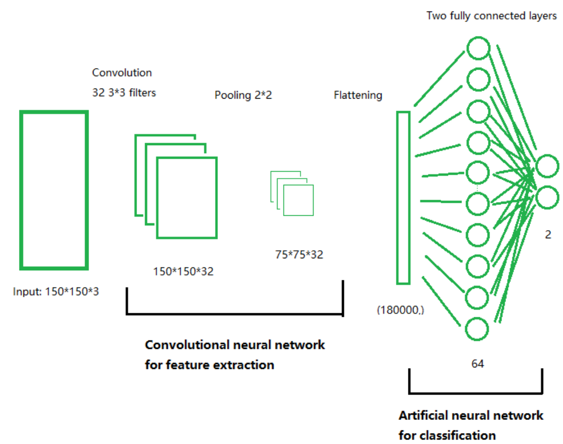

Figure 4: Demonstration of baseline model

The baseline model is inspired and modified from the models in the tutorials and practical, which consists of one convolutional layer, with a conv2D and a max-pooling operation. ReLu activation function is used in the baseline model. Hyperparameters in the baseline model include kernel size, number of epochs, number of convolutional layers, steps per epoch, validation steps, etc. In this model, the result of the accuracy turns out to be approximately 0.3, which can be evidenced by the figure below.

```
test accuracy: 0.2690992057323456
test loss: 9.802160263061523
```

Figure 5: Accuracy and loss report of the baseline model()

7 QUANTITATIVE AND QUALITATIVE RESULTS

For the final model, we improved the initial model with 3 more convolutional layers and more epochs. We could see the result we get from this model, the Test Accuracy is 0.734 and the loss is 5.405, which is way better than the original one. But from the Validation accuracy and loss, it can be interpreted that the models have not converged yet with the 30 epochs. We tried to change different Hyperparameters including increasing the number of epochs to 60, steps per epoch to 10, and the validations steps to 5. But it may be too overwhelming for the computer and the code could not be operated successfully. Although the model may overfit, the result still showed a 73.4 percent of accuracy, as well as in the testing set.

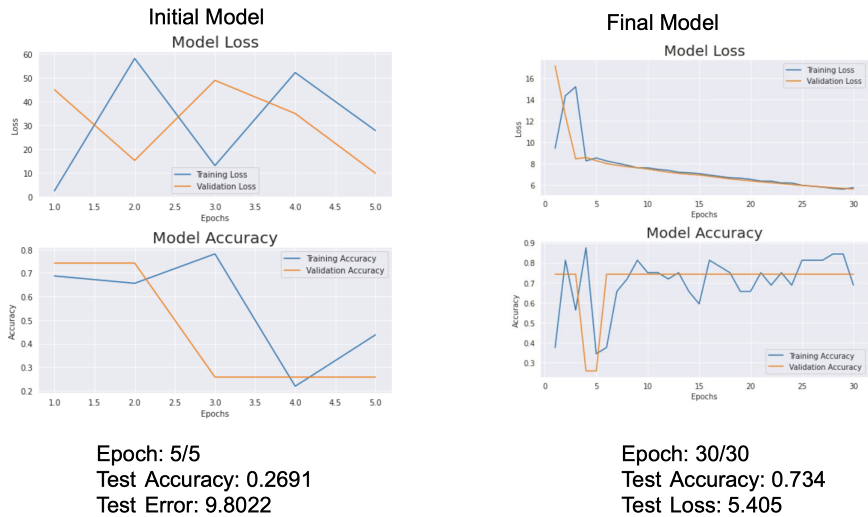

Figure 6: Learning curves of the initial model and the final model

The confusion matrix is a table used to determine the true performance of a model, which is the desired tool to demonstrate qualitative results. It identifies true positives (TP), true negatives(TN), false positives(FP) ,and false negatives(FN) by comparing and reclassifying the results predicted by the model with the true results. In the case of the artificial intelligence model, FN stands for the number of patients with pneumonia that the model missed and judged to be healthy. In this medical project, there are two types of misdiagnosis by doctors: telling patients with pneumonia that they are healthy and telling healthy people that they have pneumonia. The former is clearly more serious in terms of the potential consequences of both types of misdiagnosis, especially in the case of infectious diseases where precision is not as important as recall. In a total of 877 testing data, the results presented in this model show that FN and TP are 49 and 592 respectively and that the model gives an acceptable result in terms of overall proportions.

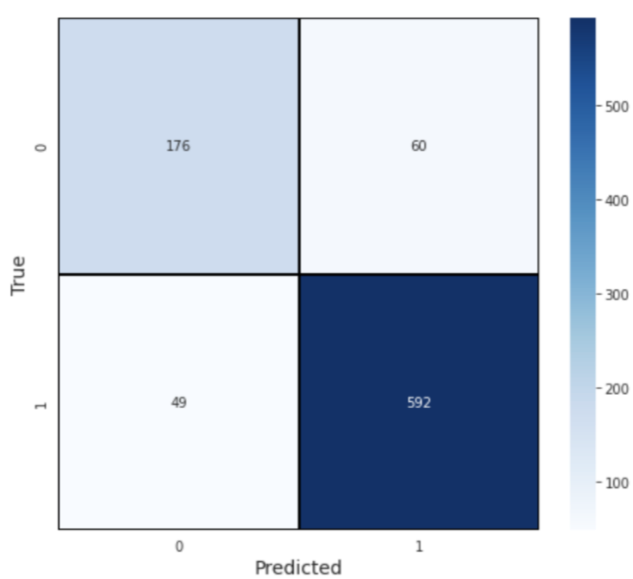

Figure 7: The confusion matrix of the final model

8 Evaluate model on new data

In order to evaluate the project, the team decided to find new data to test the model on. The team ended up collecting new chest x-ray data from Mendeley data, published by Ali Mohammad Alqudah in 2020, which consisted of chest x-rays of healthy people and x-rays of patients with COVID-19. This database was chosen for two reasons; on the one hand, the radiographs of patients that have the symptoms of pneumonia that are caused by COVID-19 have highly compatible features with those of patients with common pneumonia. "Patients with COVID-19 have typical radiological findings on chest imaging, including multifocal and bilateral ground-glass opacities, as well as predominantly peripheral and basal consolidation" (Rousan et al., 2020). On the other hand, many of the other databases the team found had data sources referring to the team's original database, while the team was unable to confirm the percentage of these databases had images that had previously been used by this model. The team, therefore, chose a data source that gave a true picture of the validity of the model to some extent. The new database differs from the previous database in that it includes images of the chest that are rotated at different angles or have a border of some width. The team removed images that were not sufficiently distinctive (e.g., pictures that are too blurred, or pictures with a black border that is too wide) and ended up using 332 normal images and 483 images of pneumonia for the evaluation. These data were not put to use in the test model until the final stage and had never been used in the previous model building section.

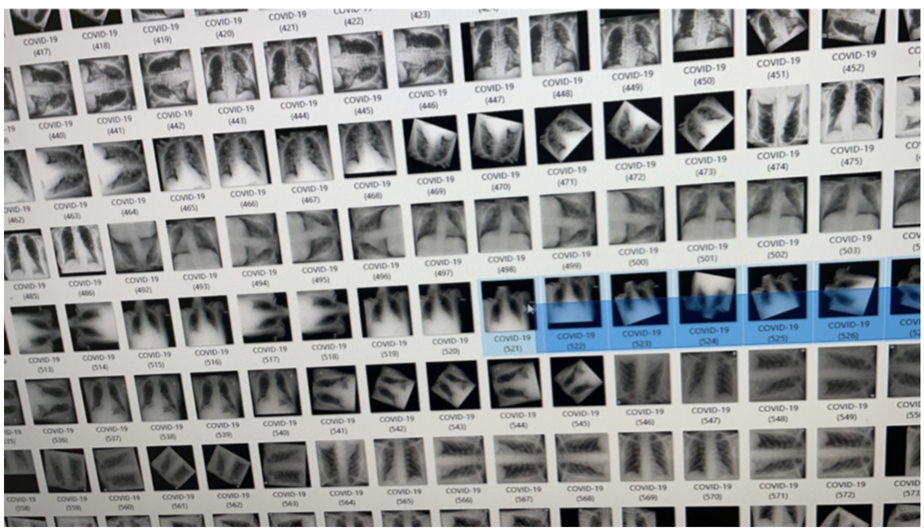

Figure 8: The demonstration of the new database

During the testing process, the team did not adjust any hyperparameters of the final model that could affect the test results to obtain a better testing result, such as the amount of the convolutional layers or the epoch number. The accuracy of the results after testing the model with the new dataset was about 73.09 percent, which is in line with the previous training state and previous test results.

```
test accuracy: 0.7309008240699768
test loss: 5.725829601287842
```

Figure 9: The accuracy and loss of the final model tested by the new database

9 Discussion

Overall, switching to the final results that the model generated with the never used dataset. The final accuracy from the new dataset is 0.7309 with a training epoch of 30. The reason that the team kept 30 epochs is that this shows the best performance overall. Obtaining a result of 73% accuracy is optimal for the team's expectations. During the model training process, the team has been achieving the best accuracy of around 73.4% with the test set that split in the original dataset. With a never used dataset, the team considered that the model has come up with consistent and well-performed results because, with a different dataset, some features of the chest x-ray images slightly differ from the original datasets. For instance, the scale of the main body of the chest may be different from the original data which may influence the feature extraction since the models are mainly built with convolutional neural networks. Also, the color scale of the new dataset may be different from the original database that was used to train the model, which the team thinks may impact the final performance of new data. Therefore, considering these possible effects that new data may bring to the final model, the team considers holding a 73% accuracy still proves that the model is performing well. The interesting part of this result is that the accuracy is surprisingly close to the results obtained from the original test sets. In the first place, we thought the influence of changing to a new test dataset would alter the results much more. Throughout the whole project experience, we learned that to build a model that helps us to reach the goal, we have to understand what are the important functions or features that the model should obtain and follow it to choose the appropriate neural network that fits the most.

10 Ethical Considerations

In the process of data collection, the chest X-ray images that the team uses may contain personal information that some patients do not want to be public. In terms of this ethical consideration, the team tries to minimize the ethical concerns in this research paper by accessing the database through a subsidiary of Google LLC, Kaggle, where the dataset is licensed under a Creative Commons Attribution 4.0 International license (Mooney, 2018). Another ethical consideration is the occurrence of false negative cases. Even though from the confusion matrix the team produces, a false negative diagnosis has the least possibilities. Before applying the model, the hospital should inform the patients' families of the possibility of the model incorrectly classifying a patient with pneumonia as healthy but the person passes away due to lack of treatment and ask for signatures.

11 Project Difficulty / Quality

The team found the project challenging and needed to use machine learning to make a medical determination of whether a patient had pneumonia. This required the model to perform its own feature analysis of different categories of X-rays by different means and summarize them to achieve an accuracy rate of at least 65 percent (it was the target the team set, the actual correct rate was around 73 percent).

To meet the machine learning requirements in this project, a large number of chest X-ray images, both of normal people and of patients with pneumonia were required. The team encountered difficulties in searching for a qualified database on the internet by browsing through a large number of websites and academic papers because images from different sources on the internet have different sizes and formats. The team converted all the images in the database to the same size and format and applied data enhancement techniques by setting the parameters to shear_range=0.2, zoom_range=0.2, horizontal_flip=true and rescaling the images to 256 colors.

Combined with what the team learned, the team started modeling the problem by complicating it, envisaging an overly cumbersome model structure, and trying to deal with several different structures at the same time, which had a significant negative impact on the team's work and articulation at the same time. The team later decided to start with the simplest structure, adding layers and hyperparameters as the team progressed, which worked well to address the issues of efficiency and collaboration.

The team did not pay too much attention to the data distribution when they first chose the database, so when improving the model the team found that even though the model was very weakly structured and even when some of the programs were not working the team was still able to get some fairly accurate results. This confused the team and prevented them from accurately determining the viability of the model. The team found that using confusion matrices and analyzing data such as FN, TP, etc., it could help to determine the credibility of the model after training more accurately.

References

Shubham Kumar Agrawal. Understanding the basics of artificial neural network, July 2021. URL https://www.analyticsvidhya.com/blog/2021/07/understanding-the-basics-of-artificial-neural-network-ann/.

Amer Kareem, Haiming Liu, and Paul Sant. Review on pneumonia image detection: A machine learning approach. *Hum-Cent Intell Syst*, 2:31–43, 2022. URL https://link.springer.com/article/10.1007/s44230-022-00002-2.

V. Kaushik, Anand Nayyar, Gaurav Kataria, and Rachna Jain. Pneumonia detection using convolutional neural networks (cnns). *Lecture Notes in Networks and Systems*, pp. 471–483, 04 2020. doi: 10.1007/978-981-15-3369-3_36.

Paul Mooney. Chest x-ray images (pneumonia). 2018. URL https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.

T. Rajasenbagam, S. Jeyanthi, and J. Arun Pandian. Detection of pneumonia infection in lungs from chest x-ray images using deep convolutional neural network and content-based image retrieval techniques. *J Ambient Intell Human Comput*, 2021. URL https://link.springer.com/article/10.1007/s12652-021-03075-2.

Ritacheta Das Rohit Kundu, Gi-Tae Han Zong Woo Geem, and Ram Sarkar. Pneumonia detection in chest x-ray images using an ensemble of deep learning models. 2021. URL https://doi.org/10.1371/journal.pone.0256630.

Liqa A. Rousan, Eyhab Elobeid, Musaab Karrar, and Yousef Khader. Chest x-ray findings and temporal lung changes in patients with covid-19 pneumonia, 2020. URL https://bmcpulmmed.biomedcentral.com/articles/10.1186/s12890-020-01286-5.

Muhammad E. H. Chowdhury Tawsifur Rahman, Khandaker R. Islam Amith Khandakar, Zaid B. Mahbub Khandaker F. Islam, and Saad Kashem Muhammad A. Kadir. Transfer learning with deep convolutional neural network (cnn) for pneumonia detection using chest x-ray. 2020. URL https://mdpi-res.com/applsci/applsci-10-03233/article_deploy/applsci-10-03233-v2.pdf?version=1589350649.