Analysis of CNN model for Pneumonia Detection

Kyle McLennan s4531857 Dominic Scocchera s4642675
Jinghan Wang s4610584

June 2022

We give consent for this to be used as a teaching resource.

The source code can be found here.

1 Problem and Significance

Striving towards accessible and affordable healthcare for all should be one of the key ventures of the modern world, especially given the events of the past few years. The use of artificial intelligence in the medical field is growing rapidly as a wide range of research is being conducted into how AI can aid clinical decisions and enhance physicians' judgement. Accurate diagnosis is crucial in the context of a functional healthcare system. Each year, within Australia, there are approximately 140 000 diagnostic errors, many of which are preventable [1]. Building tools that can assist practitioners will not only enable the reduction in diagnostic errors, but can also help to reduce burnout in doctors. Furthermore, these tools can allow us to improve diagnoses in regions where access to doctors is limited, providing more accessible healthcare to people all over the world. Although there are many instances in which AI perform healthcare tasks as well or better than humans, the context of diagnosis means that machine learning models will assist practitioners with making the final decision. With vastly growing amounts of data available, its natural to consider deep learning as a means for constructing tools for diagnosis. There exist growing databases with thousands or millions of training examples, allowing researchers to build the appropriate models.

Chest pneumonia is one such infection that can be identified with the assistance of machine learning models. Since pneumonia is the second most misdiagnosed condition leading to readmission after a previous hospitalisation [2], we are incentivised to aid in the diagnosis procedure so that patients can get the treatment they require in adequate time.

Pneuomonia is the leading cause of death from infectious diseases in children worldwide [3]. It accounts for 14% of all deaths of children under 5 years old and 22% of all deaths in children aged 1 to 5 [3]. Pneumonia is a respiratory infection that affects the lungs. The small alveoli within the lungs fill with air when a healthy person breathes, but when an individual has pneumonia, the alveoli are filled with pus and fluid, making breathing painful and limiting oxygen intake [3]. The inflammation of the lungs can be seen in a chest x-ray, as seen in Figure 1.

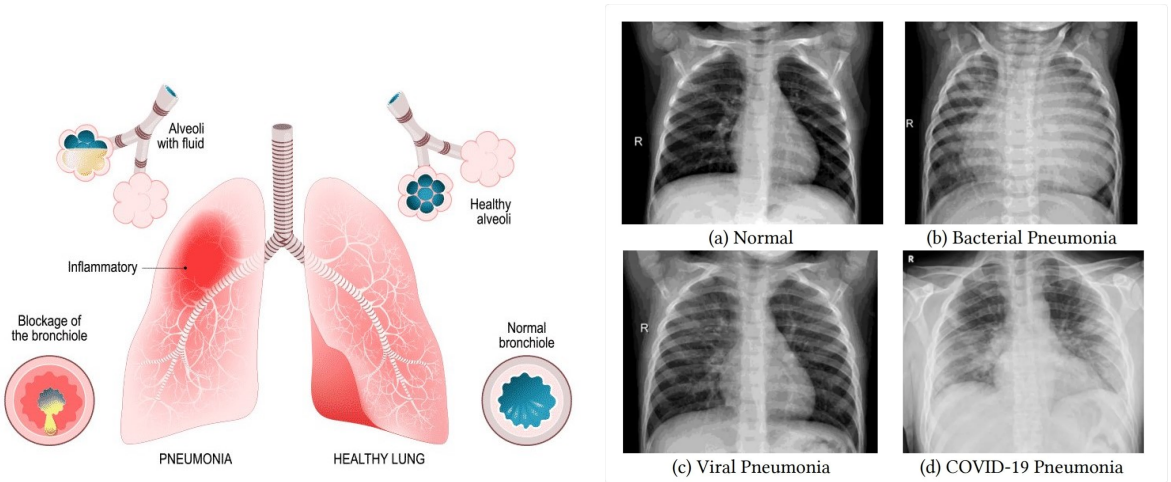

Figure 1: (a) shows a schematic difference between a healthy lung and a lung with pneumonia. Obtained from [4]. (b) shows examples of frontal-view chest X-Ray images showing various types of pneumonia.

The reason that chest pneumonia is so hard to diagnose is that the symptoms are very similar to a wide

range of other illnesses [5]. Since deep learning models are good at identifying patterns in data that humans may not necessarily be able to detect, they are a natural choice to train for the recognition of chest pneumonia. In particular, convolutional neural networks are commonly used to analyse visual images. Thus, the aim of this application project is to implement and investigate a convolutional neural network that can successfully identify pneumonia based on front-view chest x-rays.

2 Solution

Aim

The aim of the network design is to minimise the number of layers whilst maintaining a high test set accuracy (>95%). To achieve this a convolutional neural network will be implemented. In order to maximise the test set accuracy of a smaller network some preprocessing of data will also be necessary.

The Dataset

The dataset of frontal-view chest x-ray images was found on kaggle. The set contains 5,863 x-ray images labelled in two categories - those with pneumonia and those without. The x-rays were originally taken of pediatric (ages one to five) patients at the Guangzhou Women and Children's Medical Center, Guangzhou. Images were screened for quality control by removing low quality or unreadable scans. The diagnoses were graded by two expert physicians before being cleared for training AI systems.

Preprocessing of Data

Each x-ray image in the data set is a 150×150 image with pixel values from $0 - 256$. The image set from Kaggle already came with a train-test split performed, however to improve the test set accuracy the images were regrouped together and a new random train-test split was formed. The pixel values were also normalised into the range $0 - 1$ and the labels one hot encoded, $[1, 0]$ for normal and $[0, 1]$ for Pneumonia.

Network Architecture

The network architecture was designed to be a series of "layers" where each "layer" has a convolutional layer followed by a max pool layer. The final model had three "layers" followed by three linear layers. To improve the test set accuracy after every "layer" a batch normalisation layer was added and after all three "layers" a dropout layer was added with $p = 0.65$. The network was then trained using stochastic gradient descent along with the cross entropy loss function.

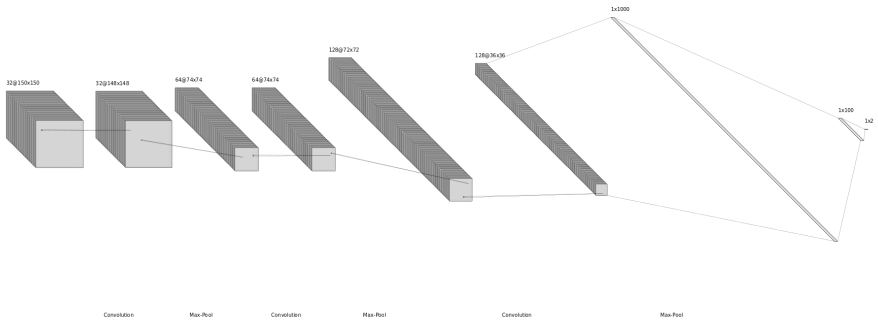

Figure 2: Overview of Main Network Components

Dropout Layer

The dropout layer was implemented to prevent the network from over fitting to the training data. Dropout layers were shown to be an effective method for achieving this in the paper "Improving neural networks by preventing co-adaptation of feature detectors"[**?**]. Dropout layers achieve this by randomly zeroing elements of an input tensor where each element has a probability p of being zeroed according to a Bernoulli distribution.

Hyperparameters

The activation function used throughout the network was ReLU. For stochastic gradient descent the learning rate was set at 0.001, the momentum at 0.01 and the dampening at 0.01. For training the batch size was set to 240.

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1         [-1, 16, 148, 148]             160
         MaxPool2d-2           [-1, 16, 74, 74]               0
       BatchNorm2d-3           [-1, 16, 74, 74]              32
            Conv2d-4           [-1, 32, 72, 72]           4,640
         MaxPool2d-5           [-1, 32, 36, 36]               0
       BatchNorm2d-6           [-1, 32, 36, 36]              64
            Conv2d-7           [-1, 64, 34, 34]          18,496
         MaxPool2d-8           [-1, 64, 17, 17]               0
       BatchNorm2d-9           [-1, 64, 17, 17]             128
          Dropout-10                [-1, 18496]               0
           Linear-11                 [-1, 1000]      18,497,000
           Linear-12                  [-1, 100]         100,100
           Linear-13                    [-1, 2]             202
================================================================
Total params: 18,620,822
Trainable params: 18,620,822
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.09
Forward/backward pass size (MB): 6.91
Params size (MB): 71.03
Estimated Total Size (MB): 78.02
----------------------------------------------------------------
```

Figure 3: Full Details of Network

3 Findings

The performance measure used was the networks classification accuracy on the images from the test set, i.e. the test set accuracy. After playing around with the layers and hyperparameters of the network it was found that the architecture and hyperparameters specified in the solution gave the best test set accuracy whilst keeping the network architecture relatively small. The final network architecture contained only 13 layers. This simpler architecture was also able to consistently converge to test set accuracy's of around 90% (the best test set accuracy achieved was 91.38%).

In order to test the networks convergence capabilities the network was trained 20 times for 20 epochs.

In figure 3 we see that many of the maximum test set accuracy's are achieved in the first 16 epochs. This means that the network tends to converge to its maximum quickly however the networks maximum is not as as would be ideal (An ideal test set accuracy would be >95%).

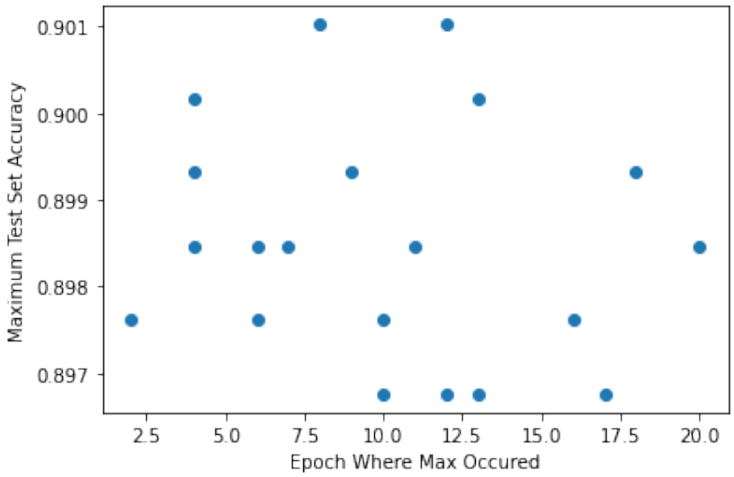

Figure 4: Maximum test set accuracy achieved in a run vs epoch where max occurred

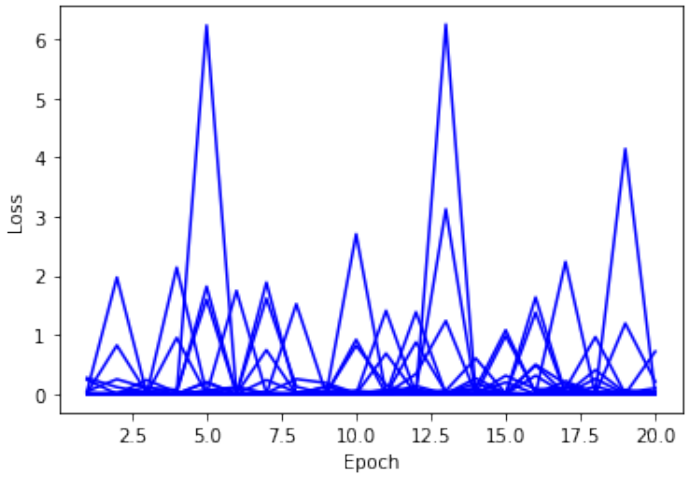

Figure 5: Loss vs epoch for the 20 training sessions

We see in figure 4 that stochastic gradient descent very quickly minimises the training set accuracy and on occasion moves away from a minimum as evidenced by the peaks in the graph. In general it is able to maintain a value below 2 thus meaning the network converges quickly to minimums of the training set.

We see in figure 5 that the test set accuracy moves around quite a bit between 0.91 and 0.88 during the 20 epochs of training. This is good in that it is consistent but is bad in that it doesn't get much better. This is likely due to the fact that the network only has 13 layers and so a more complicated network would be required to increase the accuracy.

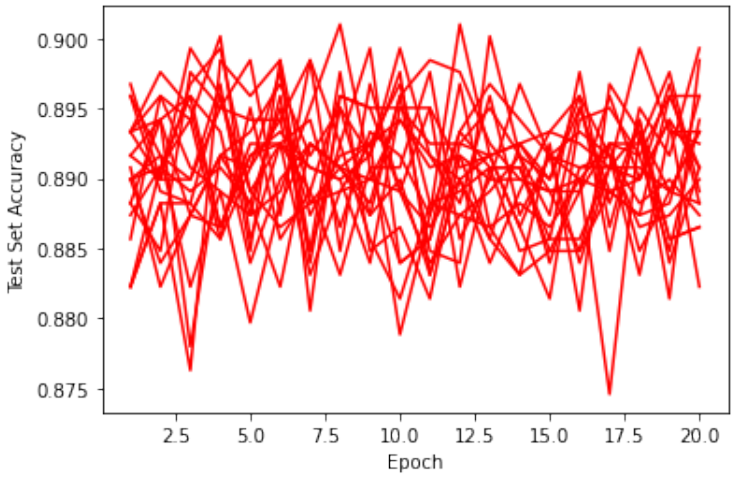

Figure 6: Test set accuracy vs epoch for the 20 training sessions

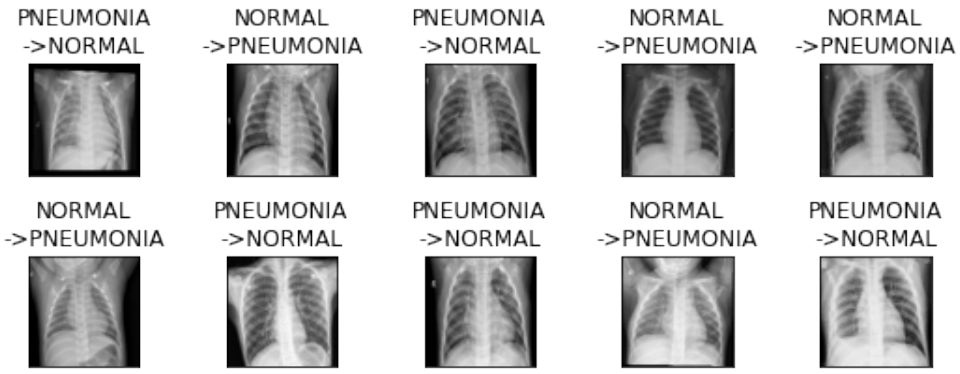

Figure 7: Some of the mistakes the network made in classification (top text is the true diagnosis and the bottom text is the prediction)

A lot of the images it miss-classifies as pneumonia tend to have a lot more white pixels and the images it miss-classifies as normal tend to have a lot of black space around the image. To solve this in future data manipulation methods could be looked into to in order to edit this sort of data in so that it may not be a problem for the network to classify these images.

Through experimentation we have been able to improve the performance of the network significantly. In particular it was found that one of the best methods for improving accuracy was to randomly assign the images to test and training sets instead of just using the training and test sets the data came in. This improved the performance from around 85% to the final best performance of 91.38%.

4 Limitations

We believe there are two main limitations in our project.

1.The function of our model is to classify the input images as pneumonia or normal. But in a real-world scenario, it's not enough to know whether an x-ray result is an image of pneumonia or not. This is due to the fact that there are many different causes of pneumonia, and each type of pneumonia has different treatments. This leads to all the different types of pneumonia needing to be identified. Ideally, outputting a specific pneumonia category would be more helpful based on the input picture. In order to achieve this we would likely have to consult a radiologist to gain an understanding of the impact characteristics of different kinds of pneumonia so that a model can be developed that identifies different types of pneumonia.

2.The x-ray images are all of children aged 1-5 thus biasing the data towards children. This may result in the model not generalizing well to adults. If a more comprehensive lung x-ray image set can be obtained, the model trained on the this data will generalise to people of all ages.

5 Conclusion

To conclude we did not achieve our aim of >95% test set accuracy but were able to come close with a score of 91.38%. Through the use of data processing techniques such as normalisation and one hot encoding this accuracy was able to be achieved with only 13 layers. The technique that gave the best increase in performance was by doing a random train test split instead of using the train test split that the data came with. Through the mistakes it is evident that some further data manipulation techniques could be used to further increase the accuracy, such as removing the dark areas around some of the chest x-rays which seemed to cause the network to classify it as normal even if the actual classification was pneumonia. In future it would also be desirable to make the model generalise so that it is capable of classifying the specific causes of pneumonia and trained on a test set with adult x-ray images as well as children's. All in all the network was able to consistently converge to around 90% within the first 16 epochs so a simple convolutional network was designed in which a descent but not optimal accuracy was achieved.

References

[1] I. A. Scott and C. Crock, "Diagnostic error: Incidence, impacts, causes and preventive strategies," *Medical Journal of Australia*, vol. 213, no. 7, pp. 302–305.e2, 2020.

[2] Passen & Powell, "Misdiagnosis of Pneumonia." https://www.passenpowell.com/misdiagnosis-of-pneumonia/, Mar. 2016.

[3] World Health Organization, "Pneumonia." https://www.who.int/news-room/fact-sheets/detail/pneumonia, Nov. 2021.

[4] "Risk factors related to Pneumonia | HHC." https://www.h-h-c.com/pneumonia-transmission-and-its-risk-factors/, Aug. 2021.

[5] Kaire & Heffernan, "Common Misdiagnoses of Pneumonia." https://www.kairelaw.com/medical-malpractice/misdiagnose-pneumonia/, Jan. 2016.