Pneumonia Detection from X-Ray Images Using CNN

Feyza Ozkefe
Isik University, Institute of Science, Computer Engineering, Istanbul, Turkey
Email: feyzaozkefe@gmail.com

Abstract This paper presents convolutional neural network models to accurately detecting pneumonia from chest X-rays, which can be utilized in the real world by medical practitioners to treat pneumonia. Experimentation was conducted on Chest X-Ray Images (Pneumonia) dataset available on Kaggle. A total of 4 models were tried and the best result was obtained with the 3rd model. It was tried to increase the accuracy by performing various hypertuning in the four models used. Achieved the highest accuracy of 90.54%. Furthermore, precision,recall and F1 scores are calculated for each model.

Keywords Convolutional neural networks (CNNs) · Pneumonia detection · ReLU · Max-pooling · Forward and backward propagation

1. Introduction

Pneumonia is an infection that inflames the air sacs in one or both lungs. The air sacs may fill with fluid or pus (purulent material), causing cough with phlegm or pus, fever, chills, and difficulty breathing. A variety of organisms, including bacteria, viruses and fungi, can cause pneumonia. Pneumonia can range in seriousness from mild to life-threatening. It is most serious for infants and young children, people older than age 65, and people with health problems or weakened immune systems. [1] In 2017, 2.56 million people died from pneumonia. Almost a third of all victims were children younger than 5 years, it is the leading cause of death for children under 5. 15% of all child deaths in 2017 were caused by pneumonia and it is therefore the leading cause of death of children. The number of children dying from pneumonia has decreased significantly over the past three decades. More than two million children died of pneumonia each year in 1990. By 2017, that number had fallen by almost two-thirds. Advances in major risk factors such as the waste of childhood, air pollution and poor sanitation, declining global poverty and better availability of health technologies such as pneumococcal vaccines and antibiotics have all contributed to this decline. While the mortality rate in the elderly decreased slightly, the number of deaths aged 70 and over increased. 1.13 million who died of pneumonia in this age group. This is because, as we showed in our introduction to the changing global age structure, the number of people reaching the age of 70 is increasing very strongly globally. [2] This article presents convolutional neural network models that detect pneumonia from X-ray images with high accuracy to assist medical practitioners.

These models have been trained to classify chest X-ray images as normal and pneumonia in a matter of seconds, therefore, it serves the purpose of early detection of pneumonia. Four classification models were created using CNN. Detect pneumonia from chest X-ray images to help control this deadly infection for all age groups. To obtain the best results, a certain number of combinations of convolution layers, dense layers, dropouts and learning rates have to be trained by evaluating the models after each execution. Three convolution layers were trained first on the data set with batch size 16 and then with batch size 64, thus a %0.27 increase in accuracy was observed. The aim of the article is to develop CNN models from scratch that can classify and thus detect pneumonic patients from chest X-rays with high validation accuracy,recall and F1 scores. Recall is primarily evaluated in medical cases, because unlike other parameters, it also returns false negative values. Number of false negatives is very important in determining the real-world performance of the models [3]. If one model provides high accuracy but low recall values, called poor performance, ineffective or even unsafe as higher false-negative values imply higher numbers the model predicts a patient as normal, but in reality the person is sick. Therefore, it can put the patient's life at risk. To avoid this, focus only be on models with great recall values, reasonable accuracies and F1 scores. [4] The article is organized in 5 parts: Ch. 1 introduces the topic of this research paper addresses its importance and relevance, its purpose and purpose. Chapter 2 explores studies done so far in this field. Chapter 3 explains methodology of the article, flowchart describing the architecture of the models and the dataset used to train and test the four models. Chapter 4 presents the results obtained by various CNN models and compares the performance of each model using accuracy and loss graph. Chapter 5 provides a brief overview ends the article and presents the most suitable model. All references those mentioned in the article are listed at the end.

2. Related Works

Many researchers have tackled the problem of classifying images with high accuracy. Here are one citation related to this paper: Anand Nayar and Rachna Jain [5] developed a CNN model to detect pneumonia disease chest X-ray images on the same dataset (kaggle). The highest success rate of 92.31% was achieved with 4 different models. The best model was trained for three convolutional layers with 32 at the first, 64 at the second and lastly 128 feature maps in third convolutional layer for more detailed feature extraction . The number of epochs for the model was 20. Adam optimizer function used for classifiers. Initially, a simple classifier model with convolutional layer of image size set to 64 * 64, 32 feature maps and employing ReLU activation function was trained. Fully connected dense layer with 256 perceptrons was utilized. Dropout layer was introduced at 0.3, and learning rate of optimizer was 0.0001 to reduce the overfitting.

3. Methodology

Dataset: The dataset that is used for this Project, *the Chest X-Ray Images (Pneumonia)* from Kaggle. The dataset consists of training data, validation data, and testing data. The training data consists of 5,216 chest x-ray images with 3,875 images shown to have pneumonia and 1,341 images shown to be normal. The validation data is relatively small with only 16 images with 8 cases of pneumonia and 8 normal cases. The testing data consists of 624 images split between 390 pneumonia cases and 234 normal cases.

Keras neural network library with TensorFlow backend has been used to implement the models. Data augmentation has been applied to achieve better results from the dataset. The two models have been trained on the training dataset, each with different number of convolutional layers. Each model was trained for 30 epochs, with training and testing batch sizes of 64. The following sub-headings further explain the above stages in depth.

CNN Architecture: CNN image classifications take an input image, process it and classify it under certain categories. CNN models are feed-forward networks with convolutional layers, pooling layers, flattening layers and fully connected layers employing suitable activation functions.

Convolutional layer: A convolution is the simple application of a filter to an input that results in an activation. Repeated application of the same filter to an input results in a map of activations called a feature map, indicating the locations and strength of a detected feature in an input, such as an image. [6] . In the CNN models, the input image is first converted into matrix form. Convolution filter is applied to the input matrix which slides over it, performing element-wise multiplication and storing the sum. This creates a feature map. 3 × 3 filter is generally employed to create 2D feature maps when images are black and white. Convolutions are performed in 3D when the input image is represented as a 3D matrix where the RGB color represents the third dimension. Several feature detectors are operated with the input matrix to generate a layer of feature maps which thus forms the convolutional layer.

Activation functions: All four models presented in this paper use two different activation functions, namely ReLU activation function and sigmoid activation function. The ReLU activation function stands for rectified linear function [7]. It is a nonlinear function that outputs zero when the input is negative and outputs one when the input is positive. Advantages of ReLU over other activation functions are computational simplicity and representational sparsity. Sigmoid activation function is used in all four models presented in this paper. This commonly used activation function is applied in the last dense layer of all the four models.

Pooling layer: Convolutional networks may include local and/or global pooling layers along with traditional convolutional layers. Pooling layers reduce the dimensions of data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Local pooling combines small clusters, tiling sizes such as 2 x 2 are commonly used. Global pooling acts on all the neurons of the feature map.There are two common types of pooling in popular use: max and average. Max pooling uses the maximum value of each local cluster of neurons in the feature map, while average pooling takes the average value. [8]

Flattening layer and fully connected layers: Fully connected layers connect every neuron in one layer to every neuron in another layer. It is the same as a traditional multi-layer perceptron neural network (MLP). The flattened matrix goes through a fully connected layer to classify the images.

Reducing overfitting: Dropout technique helps to reduce overfitting and tackles the problem of vanishing gradients. Dropout technique encourages each neuron to form its own individual representation of the input data. This technique on a random basis cuts connections between neurons in successive layers during the training process [9]. Learning rate of models was also modified, to reduce overfitting. Data augmentation technique can also be employed to reduce overfitting.

Algorithms of models: The models used are shown in fig1, fig2 and fig3. The flowchart is shown with details in fig4. The epoch number was used as 20 at the first model and 30 for the rest as a result of the trials. Rmsprop optimizer was used in all except the fourth model. Initially, the first model with 16 batch size,convolutional layer of input image size set to 180 * 180 and trained for three convolutional layers with 32, 32, 64 feature maps and employing ReLU activation function was trained. Fully connected dense layer with 64 perceptrons was utilized. To improve the result, batch size increased to 64, epochs number to 30 and keep the other parameters unchanged in the second model. In the third model, convolutional layers changed with 64, 64, 128 feature maps and the number of dense layer has been increased to 256. Lastly, in the fourth model, used the first models convolutional layers and the model was trained with one more convolutional layer of 128 feature maps for better feature extraction. The number of perceptrons in dense layer was also increased to 256, and extra two layer with 128 and 64 perceptrons, so that better learning could be achieved. Dropout layer was introduced at 0.5 for all models. The results are summarized in the next section of this article.

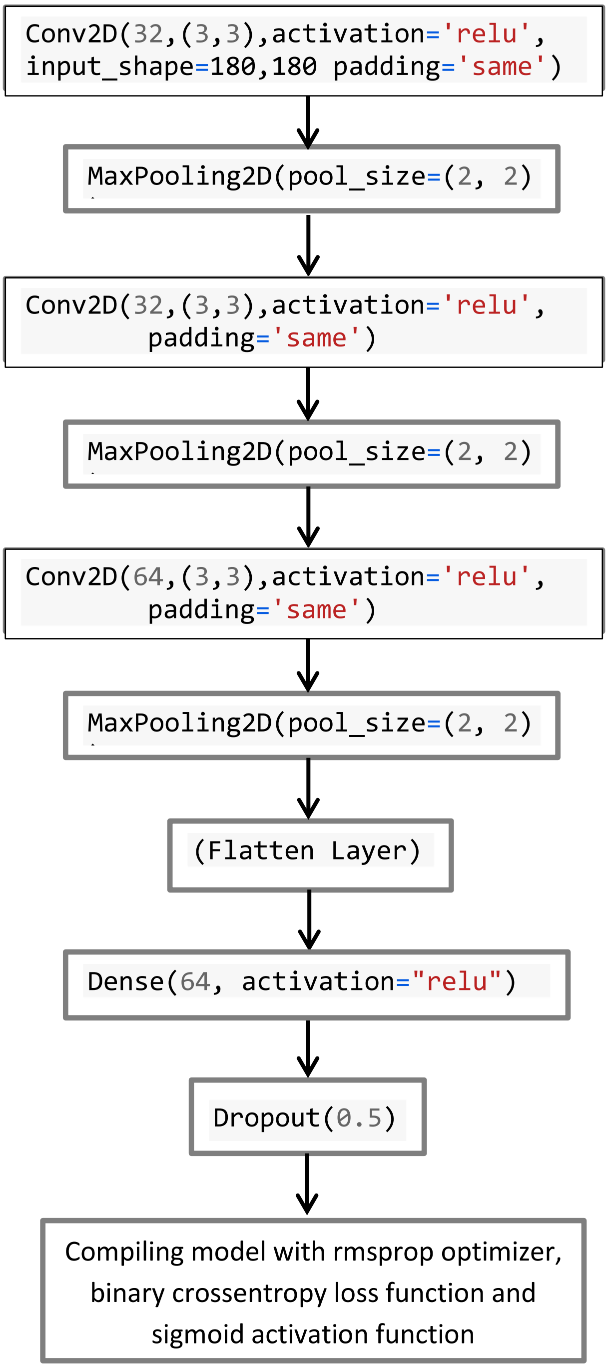


Fig. 1 Algorithms of CNN classifier model 1 & model 2

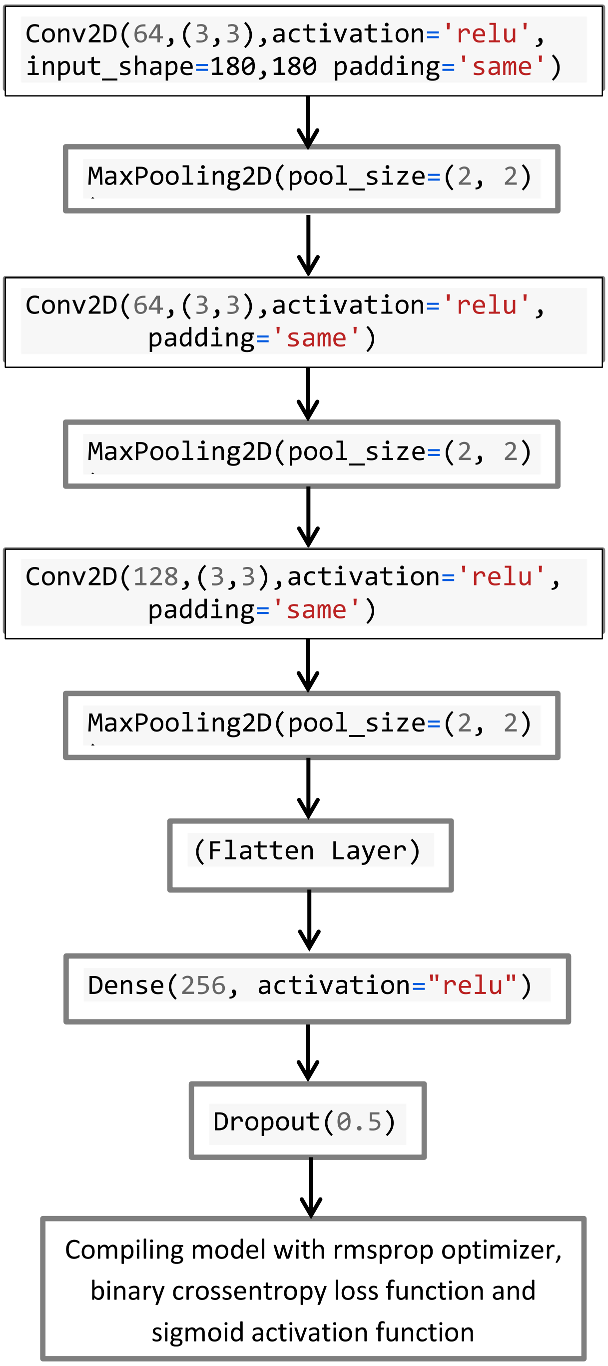


Fig. 2 Algorithms of CNN classifier model 3

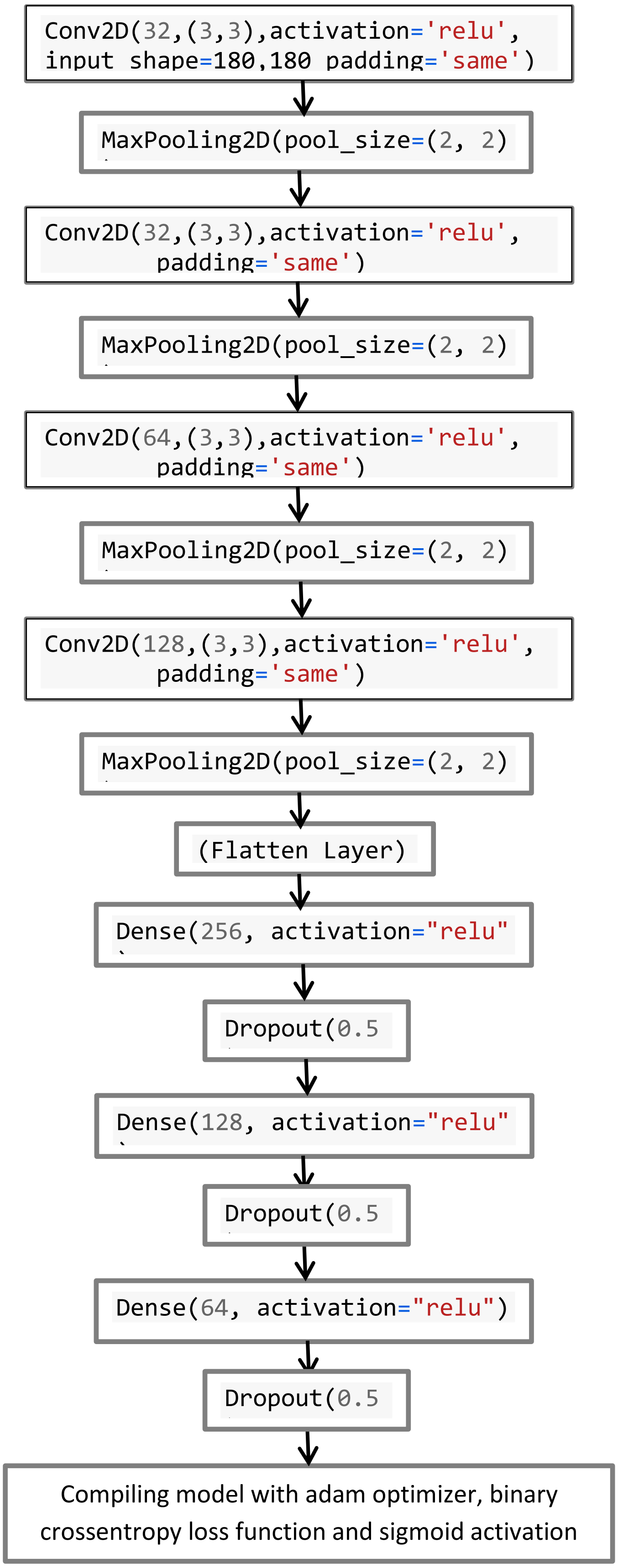


Fig. 3 Algorithms of CNN classifier model 4

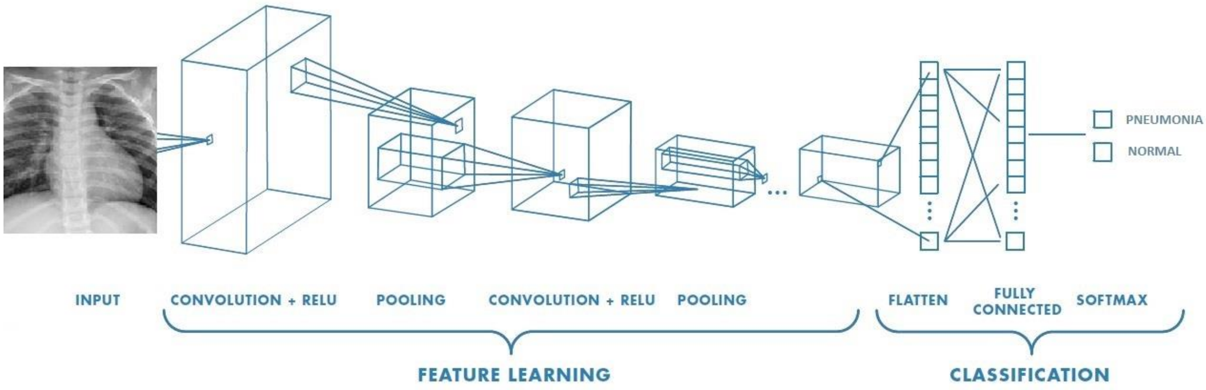


Fig. 4 CNN model flow

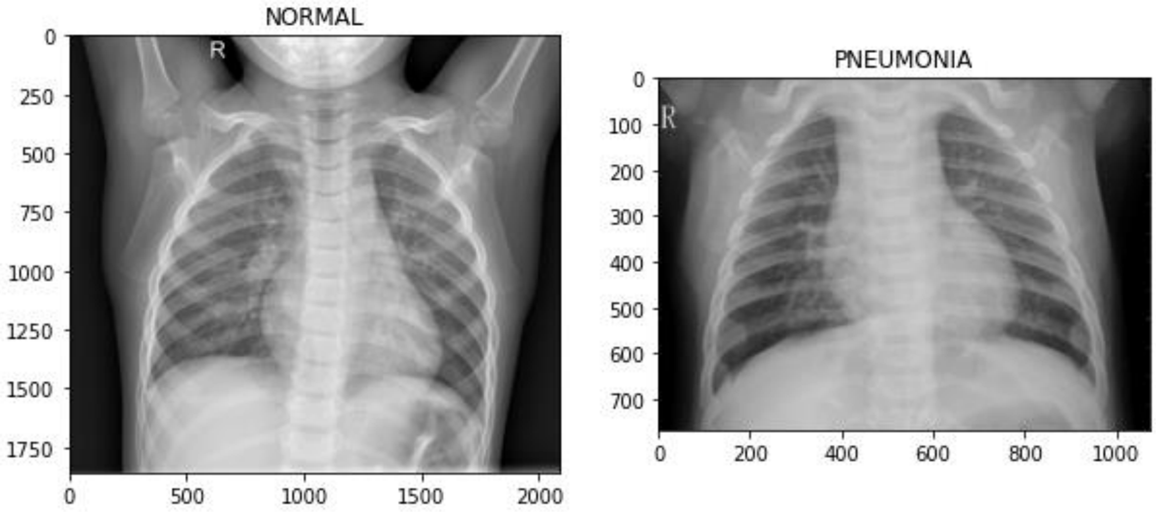


Fig. 5 Sample images from the dataset

4. Experimental Results

To study the performance of each CNN classifier model, validation accuracy, recall and F1 score were evaluated as the performance measures. Accuracy and loss graphs were also shown for all models. Figures 6,7,8,9 and Table 1 show the, accuracy graphs and loss graphs of all CNN classifier models. The recall values are very close in each model, but the accuracy and f1 scores vary. Therefore, when we choose our most successful model by comparing F1 and accuracy scores, it is seen that the most successful model is the 3rd model.

Table 1 Performance comparison of different CNN models

Model	Model Accuracy	Model Loss	Precision	Recall	F1 Score
Model 1	0.8621	1.0447	0.8261	0.9871	0.8994
Model 2	0.8894	0.4198	0.8590	0.9846	0.9175
Model 3	0.9054	0.5346	0.8787	0.9846	0.9286
Model 4	0.8958	1.0425	0.8701	0.9879	0.9252

Fig.6 Performance of classifier model 1

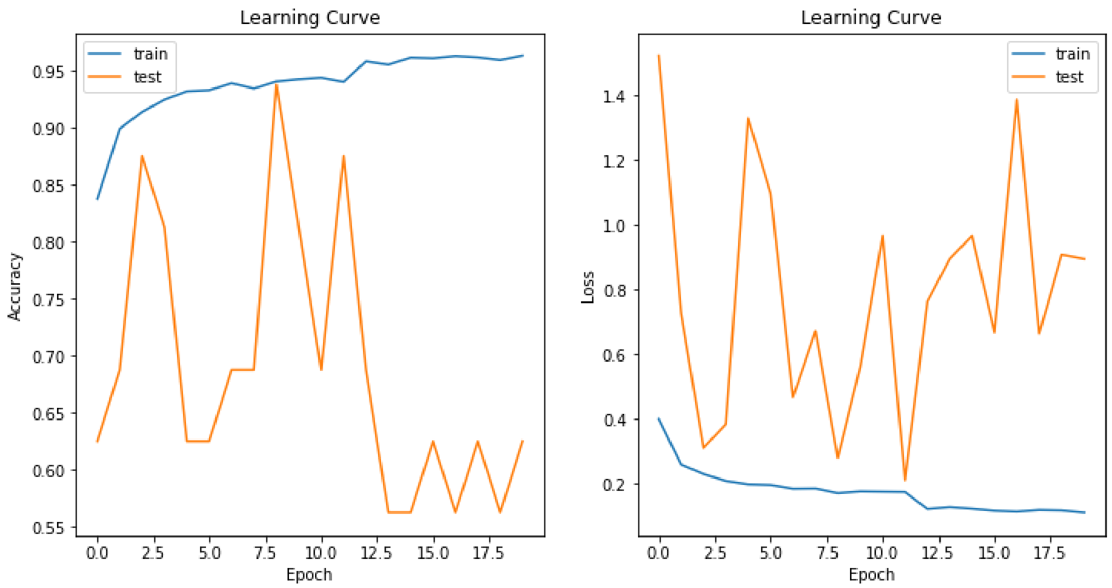


Fig.7 Performance of classifier model 2

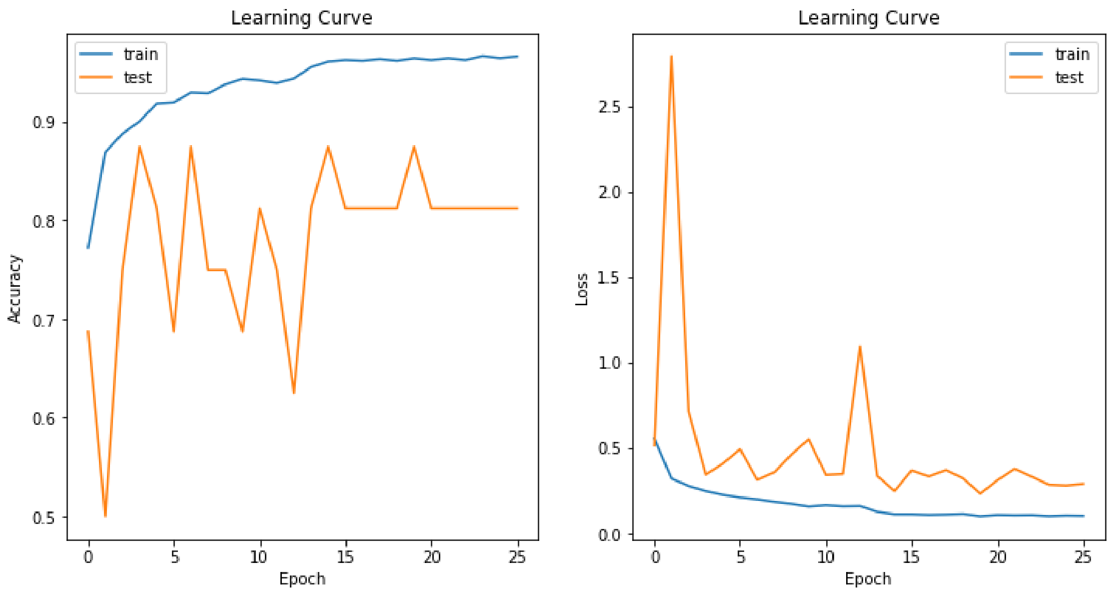


Fig.8 Performance of classifier model 3

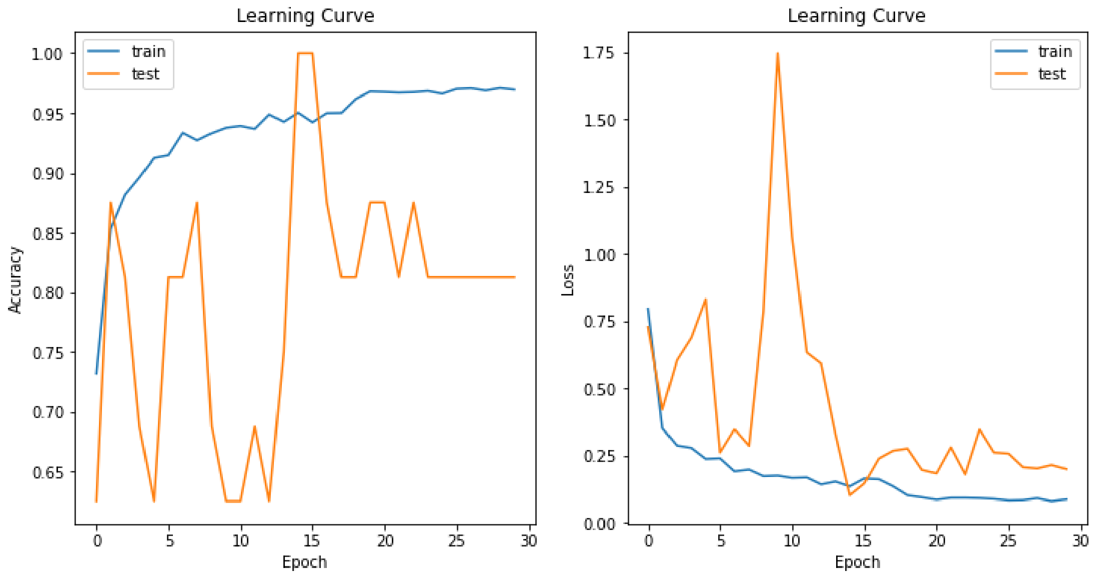
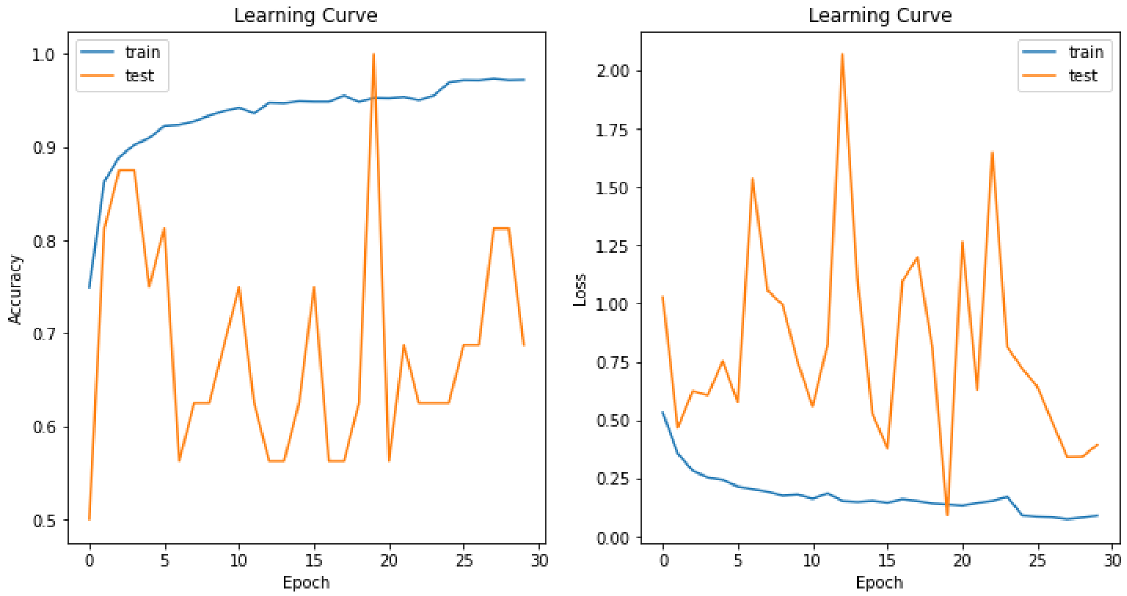


Fig.9 Performance of classifier model 4



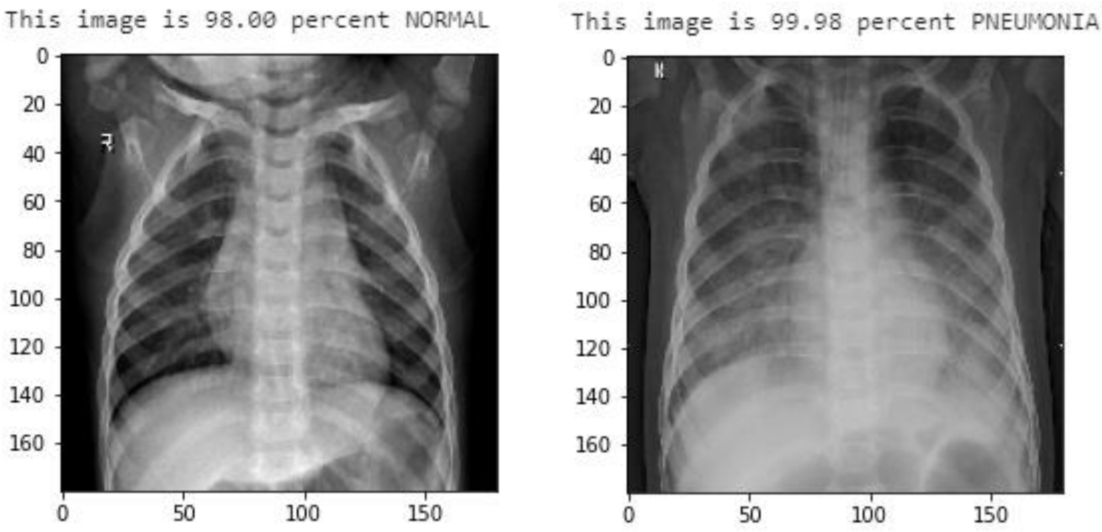


Fig. 10 Prediction for a single image

5. Conclusion

Model	Model Accuracy	Model Loss	Recall	F1 Score
Model 3	0.9054	0.5346	0.98	0.92
RW	0.9231	0.2523	0.98	0.94

Table 2 Comparison of the best models of the two studies (RW:Related Work)

A large number of X-ray images can be processed very quickly to provide highly precise diagnostic results, thus helping healthcare systems provide efficient patient care services and reduce mortality rates. These convolutional neural networks' models were successfully achieved by employing various methods of parameter tuning like adding dropout, changing learning rates, changing the batch size, number of epochs, adding more complex fully connected layers and changing various stochastic gradient optimizers [10].

When the two best models are compared, very close results are observed. The paper by Anand Nayar and Rachna Jain achieved the overall accuracy of 92.31%, %98 recall and %94 F1 score trained on the same dataset. The models presented by me at best could achieve 90.54% accuracy, %92 F1 score which is lower but the same recall with %98. Although the same number of convolutional layers and dense layers are used in the model, a low difference is observed, the reason for this change may be due to data preprocessing or differences in optimizer and activation function used.

In the 3rd model that gives the best results, the convulutional layer filter and the number of dense layers have been increased. However, not much difference was observed when compared with the second model. After the changes made in the parameters, it has been seen that there is no direct relationship with the accuracy with the extra convulitional or dense layer. The accuracy of the model is directly related to the size of the dataset, using large dataset helps to increase the accuracy of the model. It is recommended to use more different datasets to obtain higher accuracy with lower bias. If the model is based on a data collected from the same hospital when it comes to test more widely in the future, it is possible that there will be errors. In addition, when we look at the values in the dataset, it is seen that pneumonia cases are more than normal, but in real life it is the opposite. This shows that the dataset is imbalanced, for more accurate results the data needs to be diverse and ideally international. [11]

References

1. Mayo Clinic, Pneumonia, https://www.mayoclinic.org/diseases-conditions/pneumonia/symptoms-causes/...
2. Pneumonia, by Bernadeta Dadonaite and Max Roser, https://ourworldindata.org/pneumonia
3. Bernal, J., Kushibar, K., Asfaw, D.S., Valverde, S., Oliver, A., Martí, R., Lladó, X.: Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Artif. Intell. Med. 95, 64–81 (2019) https://pubmed.ncbi.nlm.nih.gov/30195984/
4. Arthur, F., Hossein, K.R.: Deep learning in medical image analysis: a third eye for doctors. J. Stomatology Oral Maxillofac. Surg https://www.semanticscholar.org/paper/Deep-learning-in-medical-image...
5. Anand Nayar and Rachna Jain, Pneumonia Detection Using Convolutional Neural Networks (CNNs) Conference Paper · April 2020 https://www.researchgate.net/publication/340961287_Pneumonia_Detection_Using_Convolutional_Neural_Networks_CNNs
6. How Do Convolutional Layers Work in Deep Learning Neural Networks?, by Jason Brownlee on April 17, 2019 https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/
7. Eckle, K., Schmidt-Hieber, J.: A comparison of deep networks with ReLU activation function and linear spline-type methods. Neural Netw. 110, 232–242 (2019), https://www.sciencedirect.com/science/article/...
8. CNN, Wikipedia https://en.wikipedia.org/wiki/Convolutional_neural_network
9. Khan, S.H., Hayat,M., Porikli, F.: Regularization of deep neural networks with spectral dropout. Neural Netw. 110, 82–90 (2019), https://pubmed.ncbi.nlm.nih.gov/30504041/
10. Du, S. S., Zhai, X., Poczos, B., Singh, A.: Gradient Descent Provably Optimizes OverParameterized Neural Networks (2018). arXiv preprint arXiv:1810.02054
11. Machine learning for Covid-19 diagnosis: promising, but still too flawed, University of Cambridge (CC BY 4.0), 17.03.2021 https://healthcare-in-europe.com/en/news/machine-learning-for-covid-19-diagnosis-promising-but-still-too-flawed.html