# Assessing the Feasibility of Diagnosis of Pneumonia using Chest X-Ray Images

Krishnakanta Maity[1], Rajat Gaur[2], and Saikat Patra[3]

[1]kkm.stat.edu.1729@gmail.com
[2]rajatgaur625@gmail.com
[3]saikatpatra365@gmail.com
[1,2,3]Computer Science Department , Ramakrishna Mission Vivekananda Educational and Research Institute , M.Sc. Big Data Analytics

June 26, 2022

## Abstract

Pneumonia is one of the widely found diseases among people especially children. Recent studies suggest that globally there are over 1400 cases of this lung disease per 100,000 children, or 1 case per 71 children [6]. This disease is caused by bacterial infection in the lungs. One challenge which is usually faced in the medical domain is early diagnosis which becomes critical in the treatment process. Our study aims to develop a computer aided diagnosis system which will assist to guide the clinicians. Through our study, we would be applying deep learning concepts and will be applying modelling techniques for the detection and diagnosis of Pneumonia.

Initially we applied traditional machine learning algorithms for the classification task which involved Logistic Regression and Decision Tress and then we compared our findings with the ones we encountered after applying Deep Learning Architecture of ResNet-50, VGG-16, VGG-19 and a custmised DL architecture. The best test accuracy achieved in this project came out be 91% via applying VGG-19 Deep Learning Architecture.

# 1 Introduction

Pneumonia is a disease that adversely affects the lungs of an individual and this inflammatory condition primarily affects the small air sacs in the lungs called *alveoli* [7]. Certain symptoms which emerge in this condition include dry cough, fever, chest pain and the person faces difficulty in breathing.

It is usually caused by the infection with viruses and bacteria. Available data suggests that there were 2.5 million deaths from pneumonia in 2019 and pneumonia is the single largest infectious cause of death in children worldwide. Among many other reasons, one of the probable reasons is non diagnosis of pneumonia in the very early phases.

Our study aims to analyze the chest x ray images and to come up with a modelling classification problem to assess the feasibility of detecting the disease so that proper diagnosis can be made available to the patients at an adequate time.

In this study, We achieved our objective with the help of traditional ML algorithms along with applying more advanced Deep Learning techniques. Some studies suggest that at times, it becomes difficult even for the domain experts in detecting pneumonia via x-ray images. One probable reason for it may be the subjectivity bias. Hence, our study is an effort in the medical domain in order to assist doctors and front line healthcare professionals so that a meaningful impact can be made in society with the help of responsible AI.

# 2 Literature review

Main objective of this research [1] is to implement DLS to diagnose and classify the chest radiographs into two classes - Normal and pneumonia. Researchers used pre-trained DLS like AlexNet, VGG16, VGG19 and ResNet50 with the softmax classifier unit. As a result they got 86.97% accuracy with VGG19 which was better than the other accuracies using other systems. For the betterment of VGG19's performance, EFS is implemented, which integrates the traditionally picked handcrafted features with DL features. The VGG19 is equipped with the classifiers such as linear-SVM, KNN, RF, DT for the better result. VGG19 with RF offered better classification accuracy (95.7%).

According to the researchers' work, the DLS based technique offered better accuracy (85.1%) compared to the radiologist prediction (82.3%) during the pneumonia class database analysis. The aim of the proposed work was to develop a custom version of the pre-trained DLS to enlarge the classification accuracy. The VGG19 consists of ReLU, 5 max-pooling layers, 3 fully connected and dropout layers and finally a softmax classifier.

In the medical image evaluation task, the performance of the proposed technique is generally appraised by computing well known performance measures, such as true-positive (TP), true-negative (TN), false-positive (FP), false-negative (FN), true positive-rate (TPR), false-negative-rate (FNR), true-negative-rate (TNR), false-positive-rate (FPR), accuracy (ACC), precision (PRE), sensitivity (SEN), specificity (SPE) and F1 Score (F1S).

Researchers [5] used four different pre-trained deep CNNs : AlexNet, ResNet18, DenseNet201 and SqueezeNet for transfer learning. Three schemes of classifications were reported : normal vs pneumonia, bacterial vs viral pneumonia, and normal bacterial vs viral pneumonia. Classification accuracies were 98%, 95% and 93.3% respectively. Being this the highest accuracy reported in the literature, the proposed study can be useful in diagnosing pneumonia more quickly and also in fast screening of pneumonia patients in airports. The motivation of the study was to diagnose pneumonia by analyzing radiographs and to differentiate viral and bacterial pneumonia utilizing the power of machine learning with better accuracy.

We get to learn a deep-CNN-based transfer learning approach for the automatic detection of pneumonia and its classes. As we saw, DenseNet201 outperforms the other three different deep

CNN networks. DenseNet201 exhibits an excellent performance in classifying pneumonia by effectively training itself from a comparatively lower collection of complex data, such as images, with reduced bias and higher generalization. This computer-aided diagnostic tool can significantly help the radiologist to take more clinically useful images and to identify pneumonia with its type immediately after acquisition.

## 3 Proposed methodology

The figure 1 below shows an overall structure which has been followed during the course of this project.

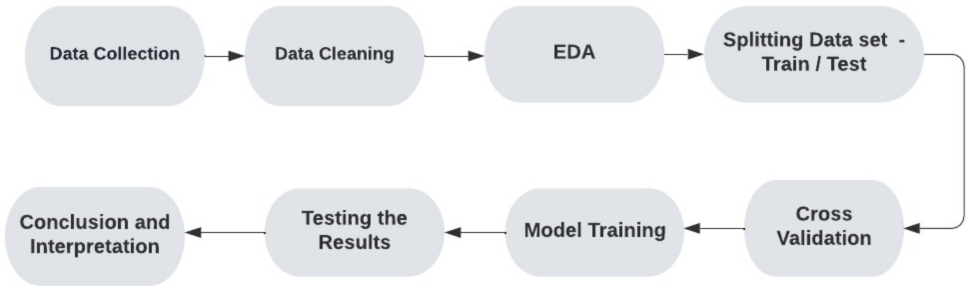



Figure 1: Proposed methodology

- **Methodology for building up Validation Set:** The data-set which has been used in our analysis contains the following two folders in train and test categories:
  NORMAL – This folder contains the chest x-ray images of people not having Pneumonia and PNEUMONIA – This folder contains the chest x-ray images of people who were diagnosed with Pneumonia.

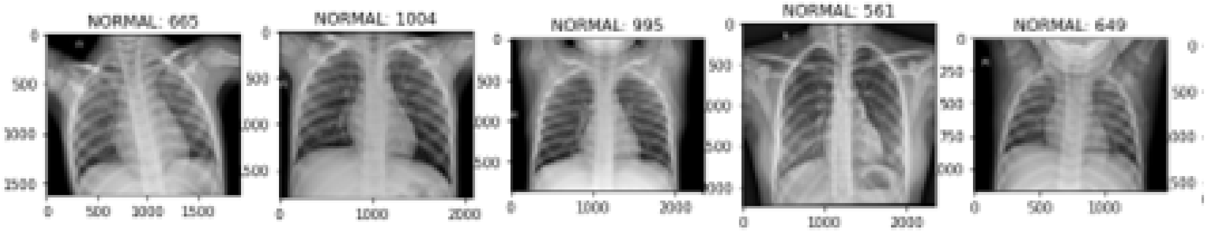
There are a total of 5232 images in train data out of which 1349 belongs to the Normal Class and 3883 belongs to the Pneumonia Class. Also, out of 624 images available in Test Data, we had 234 images belonging to the Normal Class and 390 images belonging to the Pneumonia Class. Calculating the percentage-wise allotment of images we had 89% images for training and 10% images for testing. Since the test data was kept untouched, hence the validation set was made up by randomly taking samples from Training set to keep the Train-Validation-Test Split as 80% – 10% – 10%.

The validation set after the split contains a total of 546 images out of which 173 belong to the Normal Class and 373 belong to the Pneumonia Class. All the further analysis and model fitting has been done on the above-mentioned Train-Val-Test data-set.

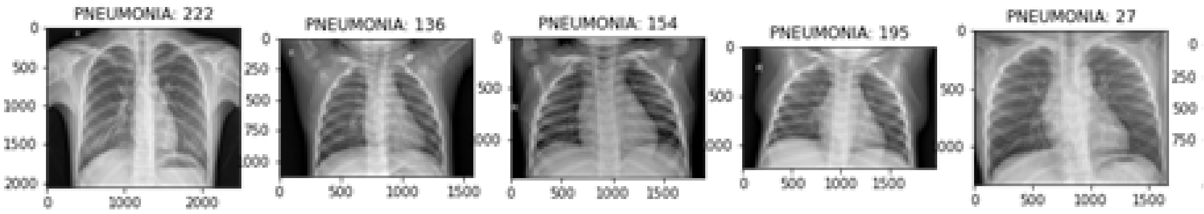
# 4 Experimental result

- **Dataset for the project**

  The dataset which we have considered for our analysis is the ZhangLabData [3] which consists of the OCT and chest x-ray images. However, for our analysis, we have made use of the chest x-ray images and have used them in order to create the train-test-validation split on which various ML algorithms have been applied. A few of the images have been shown in figure 2.




(a) NORMAL




(b) PNEUMONIA

Figure 2: Example of images

- **Experimental settings**

  We have applied classical machine learning algorithms - Logistic Regression and Decision Trees along with a customized architecture and a few Deep Learning Architectures - ResNet50, VGG-16 and VGG-19.

  In the dataset, there was no uniformity in the size of the images and the dimensions of images were not uniformly followed. Therefore in order to feed the images to any of the two methods - Classical or Deep Learning, a uniform size was supposed to be maintained, hence a uniform dimension has been used.

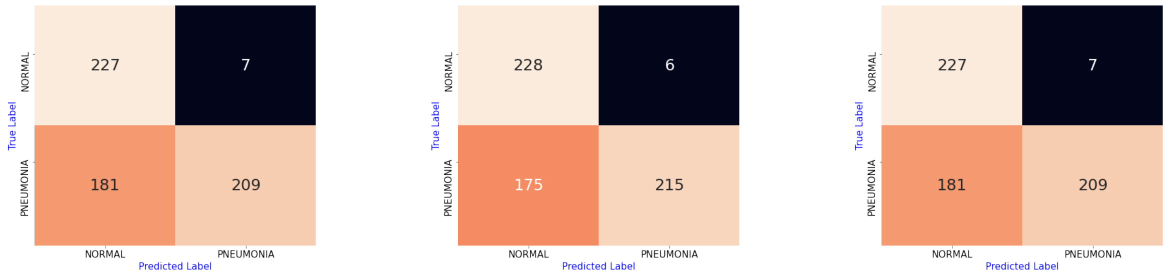
  For each of the two classical methods, the pre-processing of the data was followed by resizing the images to 32x32 post which we normalized the values and this was given as an input to the model.

  In case of Deep Learning Architectures, data augmentation was carried out by random rotation, resizing, random horizontal flip and normalization. In each case of pre-trained models or customised model, we resized the images to the dimension 224x224 and these were given to input layer for feature extraction and training of the model.

- **Experimental results and comparison with the state-of-the-art methods**

  We will be discussing the results of the following ML and DL algorithms which we have applied on the proposed dataset.

  - **Logistic Regression:** The ML algorithm which has been applied in our data is the Logistic Regression for binary classification. We have reported the confusion matrices for different solvers such as ‘Newton-CG’, ‘liblinear’ and ‘LBFGS’. Maximum no. of iterations used are 1000. The confusion matrices for each of the solvers have been shown in figure 3:




(a) Newton-CG (b) liblinear (c) lbfgs

Figure 3: Confusion matrices for different solvers for Logistic Regression

The values of precision, recall , F1 score and accuracy for the above confusion matrix has been shown in the table 1 below.

| | Accuracy (%) | | | | |
|---|---|---|---|---|---|
| Solver | Training | Testing | Precision | Recall | F1 Score ($) |
| Newton-CG | 99.98 | 69.87 | 0.967 | 0.535 | 0.689 |
| liblinear | 99.98 | 70.99 | 0.972 | 0.551 | 0.703 |
| LBFGS | 99.98 | 69.87 | 0.967 | 0.535 | 0.689 |

Table 1: Model performance summary

As we are getting more training accuracies and lesser testing accuracies, we can conclude that the model is being over-fitted here. Among all the solvers, we are getting a relatively better test accuracy of 70.99% when we are using Liblinear.

- **Decision Tree:** The results of the decision trees ML algorithm using Gini Index in terms of the classification report have been shown in figure 4.
  Here the model accuracy which we have achieved using decision trees is 62.5%. For better interpretability, a visual representation of the model has been shown in figure 4. Summary of the best model performances of various traditional Machine Learning algorithms used to evaluate the model.

| Model | Accuracy | Remarks |
|---|---|---|
| Logistic Regression | 70.99 | Solver - Liblinear |
| Decision Trees | 62.5 | Max depth -3 |

Table 2: Summary of traditional ML models

- **ResNet-50:** After the application of traditional algorithms, next we have used a variant of the ResNet Model, a Deep Neural Network architecture which consists of 48 convolution layers and 1 average pool layer. We have used the technique of Transfer learning for the application of various Deep Neural Network Learning architectures wherein we have used the pretrained models and its parameters to be used in training our model At the end of the fully connected layers, we have used softmax classifier for predicting the class label into Normal Vs Pneumonia.. The model architecture has been shown in figure

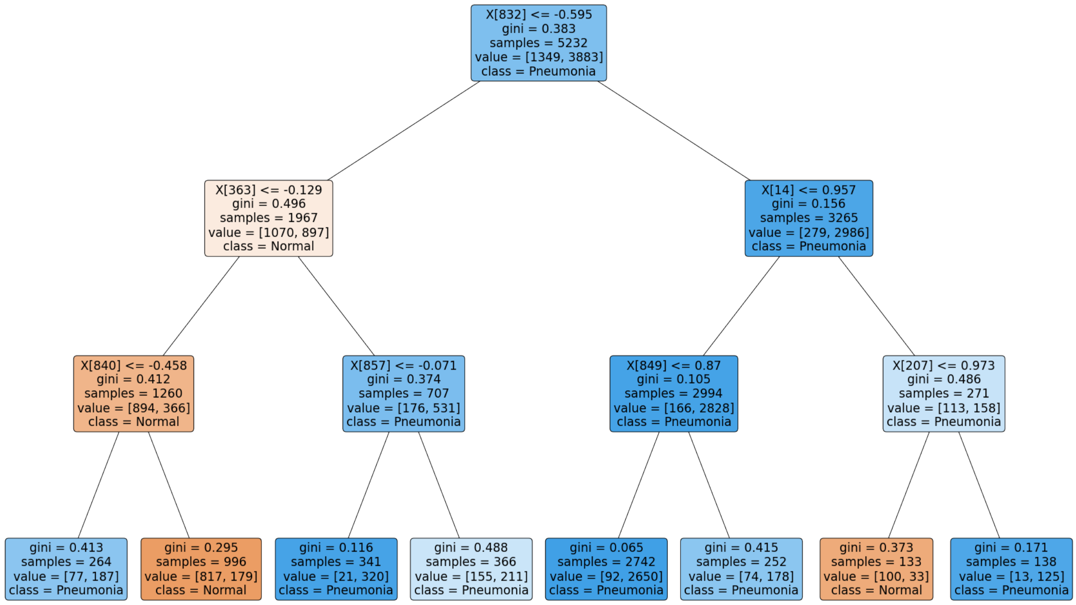



Figure 4: Visual representation of the decision tree

The graphs for losses and accuracies have been shown below,

- **VGG 16:** We have explored the model performance with the help of another DL architecture VGG-16 that is 16 layers deep. It is also a pretrained model which we have used in our classification problem. The architecture of VGG 16 has been shown in figure 5. We have received an overall accuracy of 91% on test dataset (see Figure 8).
- **VGG 19:** A more advanced version of VGG is VGG 19 that is 19 layers deep, which also allows us to use the pretrained model to train our classification dataset. The architecture for VGG 19 has been shown in figure 7. For loss, accuracy and prediction see figure 8.
- **Custom Model:** We also trained the model with the help of manually building a deep neural network having 5 convolution layers and 5 layers for batch normalization. Before feeding the inputs to fully connected layers we apply max pooling with stride =2 and at the last we have 3 fully connected layers for the output. The optimizer used for training the model is Adam/SGD with Nesterov acceleration.

- **Time complexity**

In case of logistic and Decision Trees, the input dimension of images is comparatively lower as compared to the input dimensions of the images in Deep Learning Architecture. Hence the computational complexity of the traditional machine learning algorithms is not as complex as the Deep Learning Architectures wherein non linearity is being introduced by the activation functions and the loss functions become more complex to solve.
Comparison of computational complexity and testing time per image are shown in figure 11. Together both the figure implies that testing time per image is proportional to the number of parameter in the model.

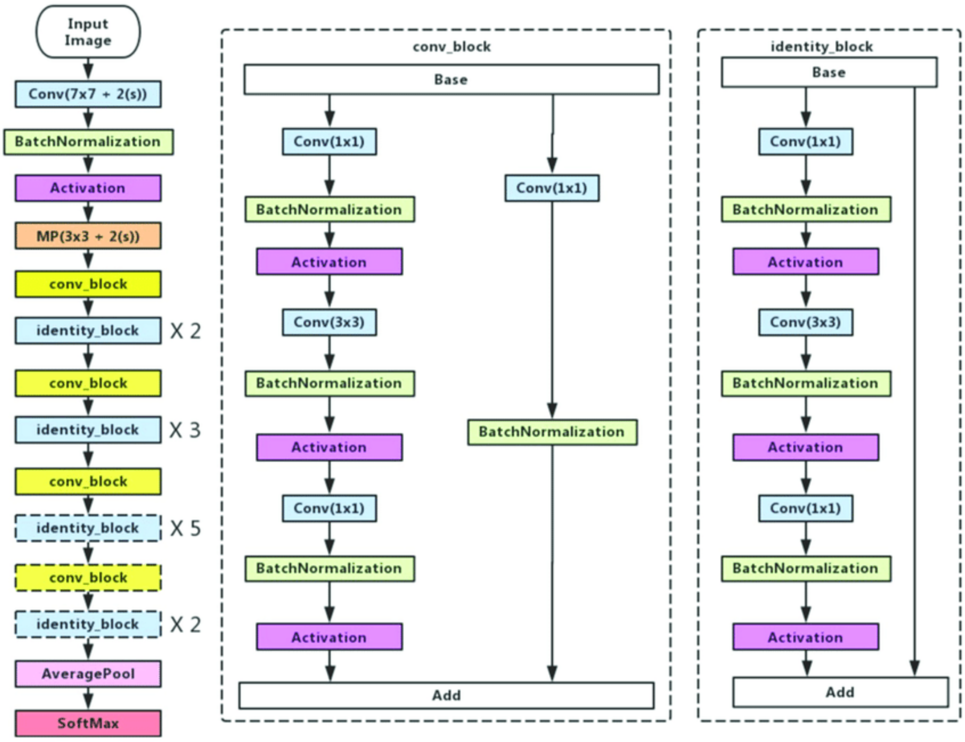
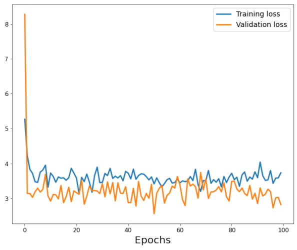
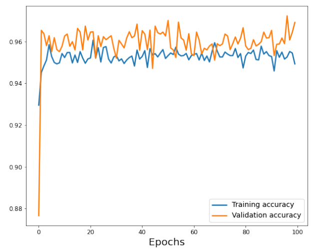


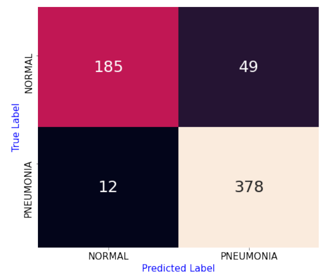

Figure 5: Architecture of Resnet50 [2]




(a) Loss




(b) Accuracy




(c) Confusion matrix

Figure 6: Loss, Accuracy and Confusion matrix for Resnet50 model

## 5 Summary

As we have tried various machine learning and deep learning techniques in our analysis and the classification of Normal vs Pneumonia images, our results have shown that we have achieved a maximum accuracy of around 70% with Logistic Regression when we used solver as liblinear and in case of Decision Trees, we could achieve an accuracy of 62.5% with a Decision Tree of maximum depth 3. A few probable reasons for achieving a comparatively lower accuracy in classical Machine Learning algorithms include firstly, limiting image size of 32x32 and secondly, the classical algorithms are not adequate to extract more meaningful and detailed features from the images . In the Deep Learning Architectures of ResNet50, VGG16 and VGG19, our results have shown that among all three, VGG-19 Deep Learning Architecture has come out to give a better performance in terms of accuracy which is around 91%.

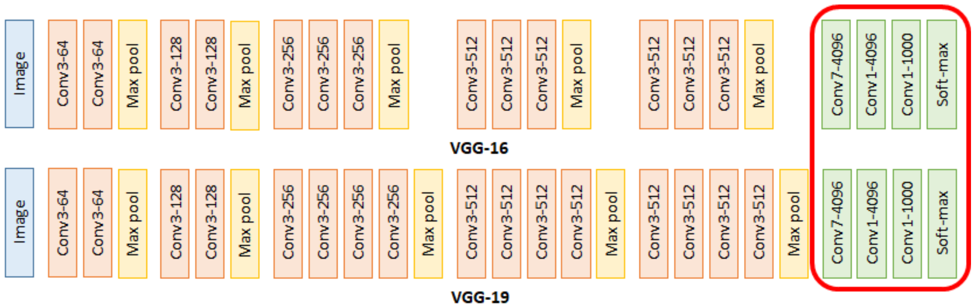
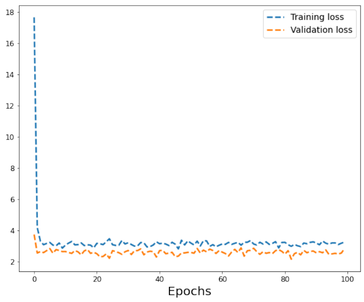
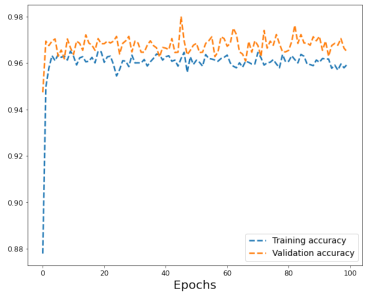


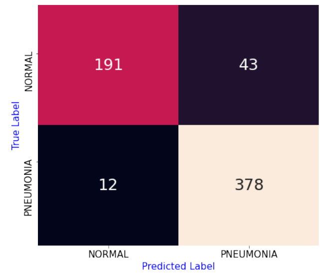

Figure 7: Architecture of VGG16 and VGG19 [4]




(a) Loss

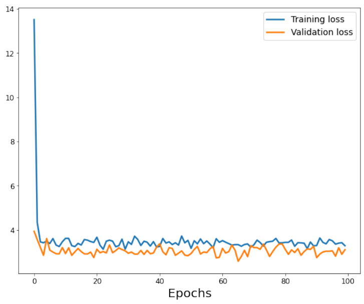



(b) Accuracy

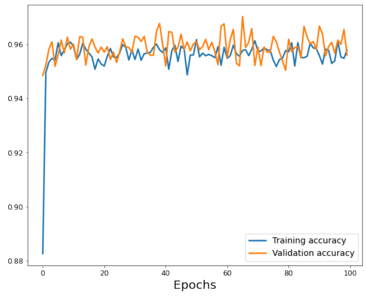



(c) Confusion matrix

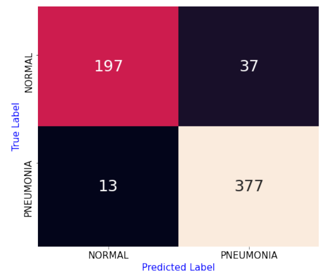
Figure 8: Loss, Accuracy and Confusion matrix for VGG16




(a) Loss

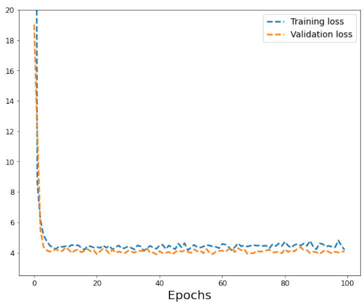



(b) Accuracy

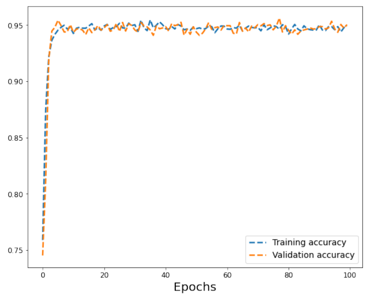



(c) Confusion matrix

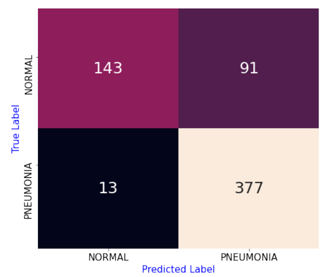
Figure 9: Loss, Accuracy and Confusion matrix for VGG19




(a) Loss

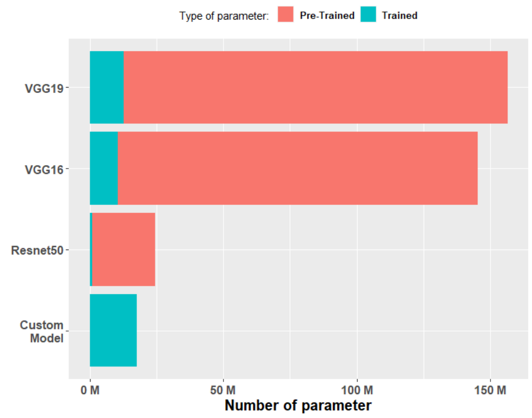



(b) Accuracy

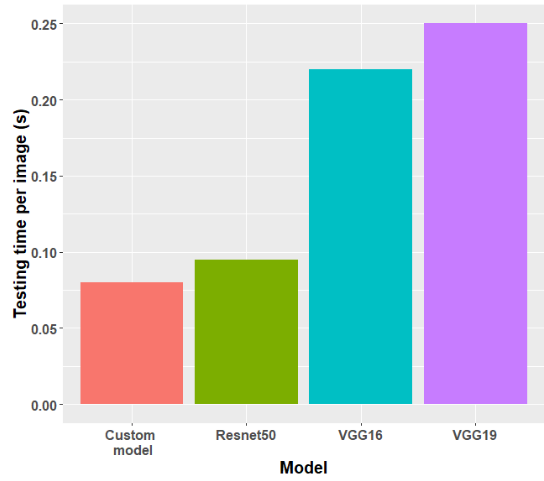



(c) Confusion matrix

Figure 10: Loss, Accuracy and Confusion matrix for custom model




(a) Computational complexity




(b) Testing time per image

Figure 11: Time complexity

| Model | Test Loss | Test Accuracy (%) | | | Precision | Recall | F1 Score |
|---|---|---|---|---|---|---|---|
| | | Normal | Pneumonia | Overall | | | |
| Custom | 0.445 | 61 | 96 | 83 | 0.8056 | 0.9667 | 0.8788 |
| Resnet50 | 0.291 | 77 | 96 | 89 | 0.9444 | 1 | 0.9714 |
| VGG16 | 0.2131 | 81 | 96 | 91 | 0.8979 | 0.9692 | 0.9322 |
| VGG19 | 0.1962 | 84 | 96 | 91 | 0.9106 | 0.9667 | 0.9378 |

Table 3: Model performance summary

| Optimizer | **Adam** | **SGD with Nesterov Accelaration** |
|---|---|---|
| Minimum training loss | 5.06 | 3.99 |
| Minimum validation loss | 3.88 | 3.81 |
| Test loss | 0.50 | 0.44 |
| Maximum training accuracy | 94% | 95% |
| Maximum validation accuracy | 75% | 93% |
| Test accuracy for NORMAL class | 51% | 50% |
| Test accuracy for PNEUMONIA class | 96% | 98% |
| Overall test accuracy | 79% | 80% |

Table 4: Custom model performance summary

## 6 Limitations

- The dataset which has been considered for the analysis was not provided in a uniform format in terms of image quality. We need to also check how the results differ when the images are not uniformly taken or image quality is not adequate.
- The validation dataset had very few images. Hence, we had to manually increase the sample size by dividing the train-test-validation split which might have affected the model accuracies.
- Taking inputs from a subject matter expert becomes valuable for models in the medical domain, expert advice could not be included in our study; if done that could have brought a new perspective in our analysis.

- The demographics of the patients may also play a critical role in the diagnosis of the disease. However, the information about age, sex, etc. has not been provided which might further assist in the interpretation of results.
- As the analysis has been done by a team of 4 only, there could have been individual bias in our models.

## 7 Future Scope

- Getting recent data from hospitals of chest x-ray images and evaluating the model performance on unseen images.
- Applying more advanced deeper neural network architectures to improve model accuracy.
- Extending this study in the diagnosis of other diseases which can be found via chest x-ray images.
- Collaborating with the stakeholders in the medical domain to check the feasibility of using the proposed model to assist doctors and practitioners.
- Taking in account different perspectives of people/patients/doctors or other stakeholders to keep the bias as low as possible.

## References


[1] Nilanjan Dey, Yu-Dong Zhang, V Rajinikanth, R Pugalenthi, and N Sri Madhava Raja. Customized vgg19 architecture for pneumonia detection in chest x-rays. *Pattern Recognition Letters*, 143:67–74, 2021.

[2] Qingge Ji, Jie Huang, Wenjie He, and Yankui Sun. Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images. *Algorithms*, 12:51, 02 2019.

[3] Daniel Kermany, Kang Zhang, Michael Goldbaum, et al. Labeled optical coherence tomography (oct) and chest x-ray images for classification. *Mendeley data*, 2(2), 2018.

[4] Medium. Vgg16 and vgg19 architecture, 2022. [Online; accessed 24-June-2022].

[5] 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.

[6] UNICEF. Pneumonia. `https://data.unicef.org/topic/child-health/pneumonia/`, April, 2021.

[7] Wikipedia contributors. Pulmonary alveolus — Wikipedia, the free encyclopedia, 2022. [Online; accessed 24-June-2022].