Indian Institute of Technology Bombay

Prediction of Pneumonia from X-ray images

CS725
Foundations of Machine Learning

Project

Chitransh Saxena – 214106009 Aditya Prasad Roy – 184106010 Murali Aadhitya M S – 22D2011

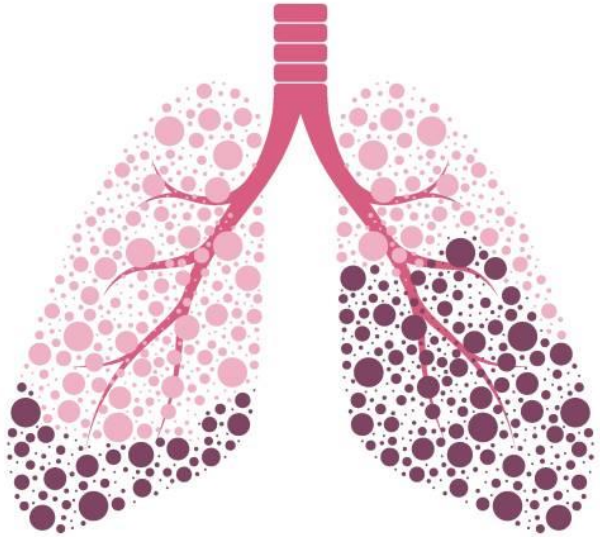
1

Introduction

Pneumonia

- Pneumonia is an infection of the alveoli in the lungs, which makes breathing painful and reduces the oxygen intake.

- The symptoms occur due to pus and fluid accumulation by pathogens such as viruses, fungi or bacteria.



World Pneumonia Statistics - 2019

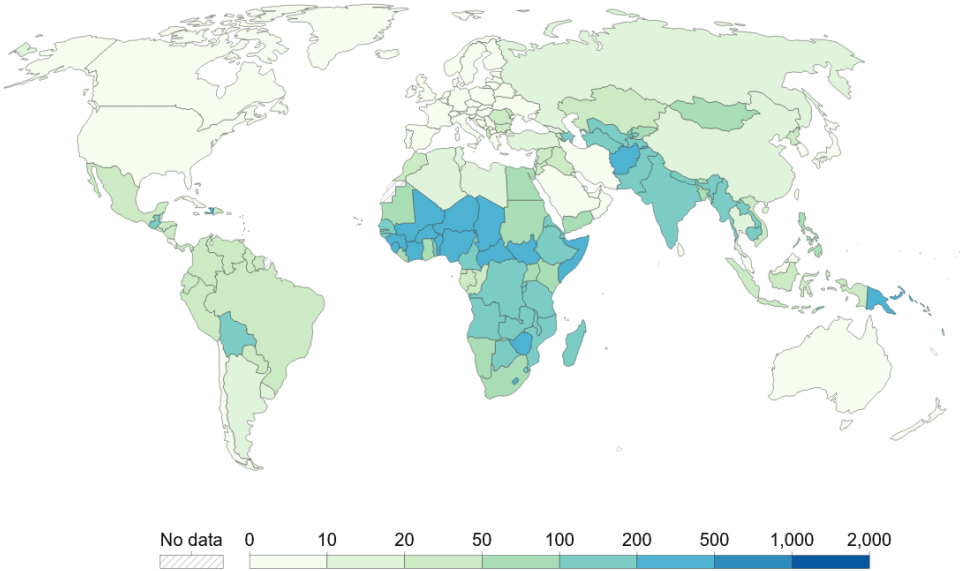


Figure 1. Death rate demographics of pneumonia in children

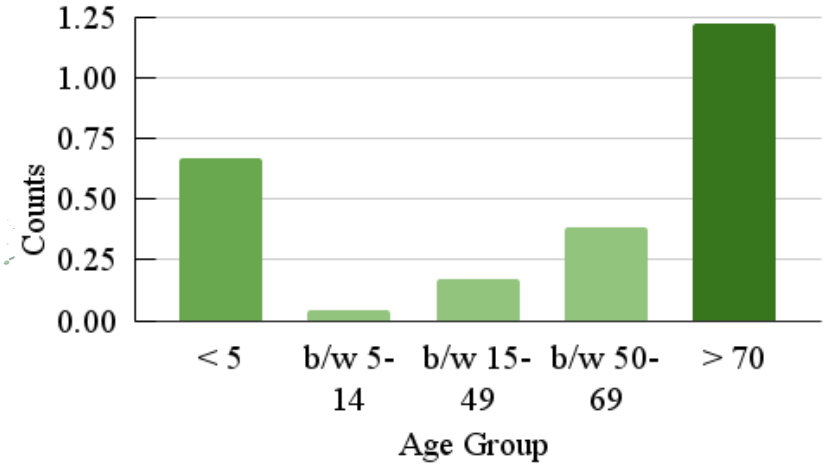


Figure 2. Worldwide death statistics of pneumonia. (Y-axis in millions)

Source: ourworldindata.org/pneumonia

Pneumonia Diagnosis

- Physicians diagnose pneumonia through methods such as radiological imaging and blood tests.

- Blood tests allow identification of the infectious agent, while X-ray images assists in locating the area of infection as well as to describe the severity

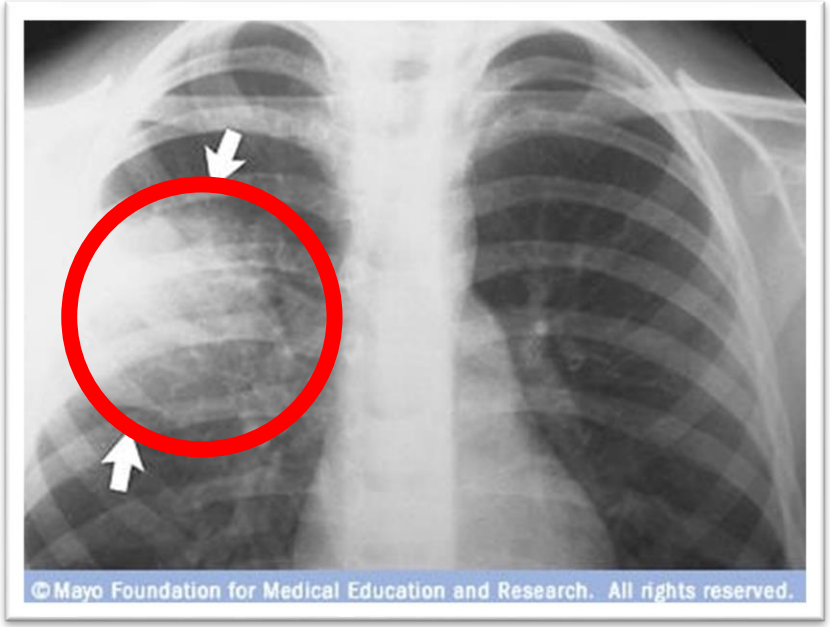


Figure 3. X-Ray with opacification of lungs

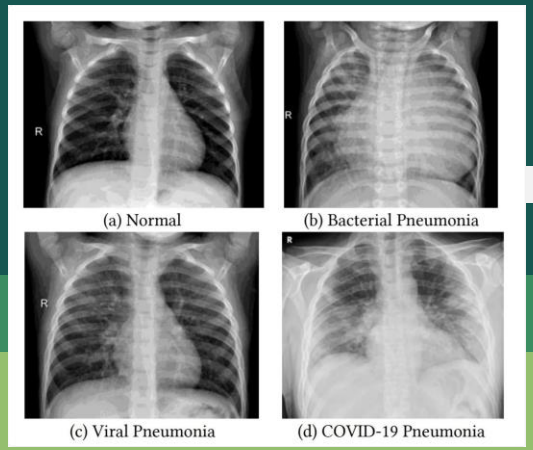
2

Aim, Methods and Data Used

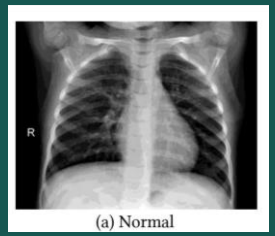
AIM

To attempt classification of x-ray data into normal and infected by using machine learning techniques

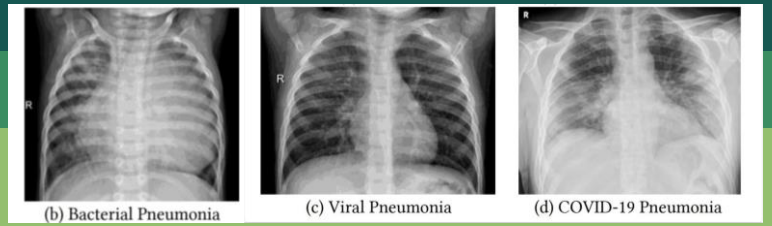
All Dataset



Normal



Pneumonia



Dataset used

- We use dataset from Ref. **[1]** to train our machine learning model.
- Our dataset comprised of 5,232 labelled X-ray images, of which 1,349 are normal and 3,883 marked pneumonia (inclusive of all types of infection – bacterial and viral).
- We use 80 % of the data for training our model and rest for validation and testing purpose.

[1] - ChestXRay2017.zip - Mendeley Data; https://data.mendeley.com/datasets/rscbjbr9sj/2/files/f12eaf6d-6023-432f-acc9-80c9d7393433

Methods

- In this project, we have investigated 2 methods to generate a model that can be used to classify a given X-ray image into 2 classes, i.e, normal and pneumonia:
  - Convolutional Neural Networks (CNN)
  - Support Vector Machines (SVM)

3

Convolutional Neural Network

Model - CNN

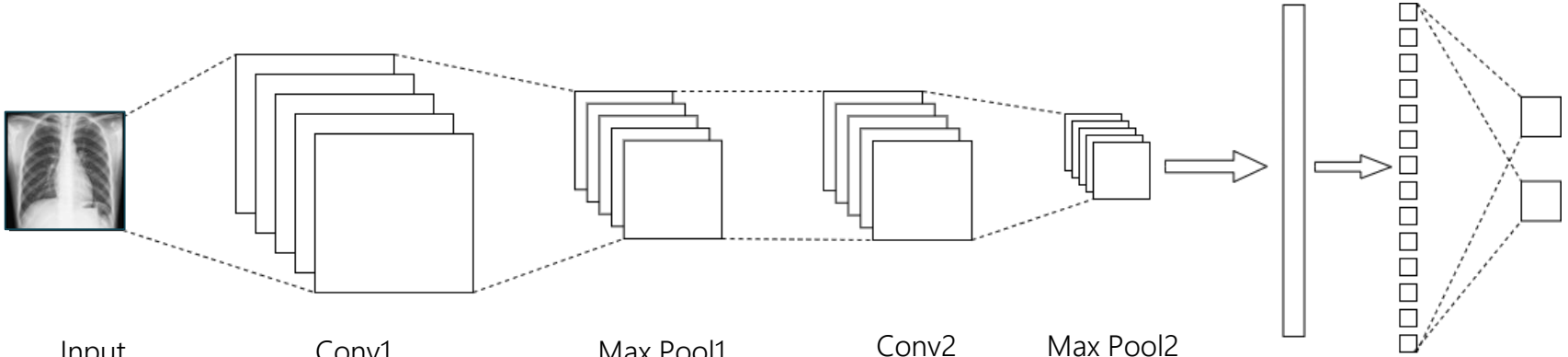


Figure 4. Diagram of CNN model structure.

Model Hyperparameters - CNN

- We performed a binary classification of the X-ray data with the following hyperparameters:

Epochs	=	50
Batch Size	=	32
Input Channels	=	1 (Greyscale)
Image size	=	256 x 256 (Rescaled and resized)
Filters	=	32 (1 - Convolution layer)
Dense layer 1	=	32 Neurons (ReLU)
Dense layer 2	=	2 Neurons (Softmax)

Model Results - CNN

- Our CNN model achieved an accuracy of 98.44 %, with the training loss attaining 1.65 x 10^{-4}.

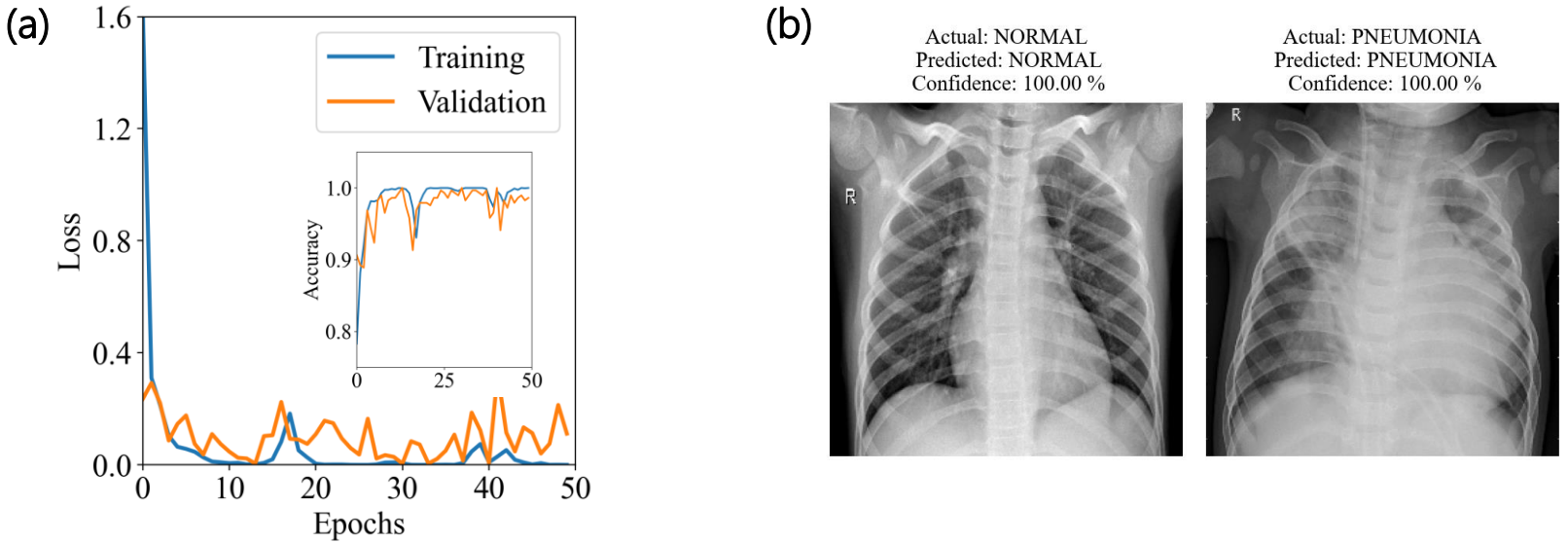


Figure 5. (a) Loss and accuracy (inset) as a function of epochs (b) Predictions of CNN model trained with 2944 datasets. Datasets are equally divided for both normal and pneumonia classes

4

Support Vector Machines

Model - SVM

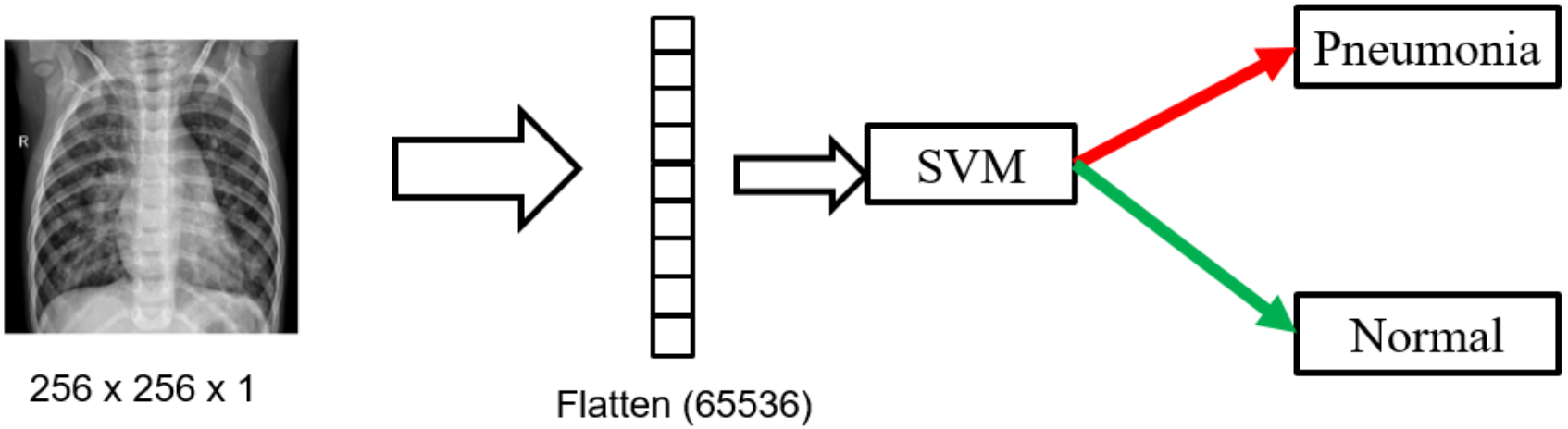


Figure 6. Diagram of SVM model structure.

Model – SVM (Feature Space)

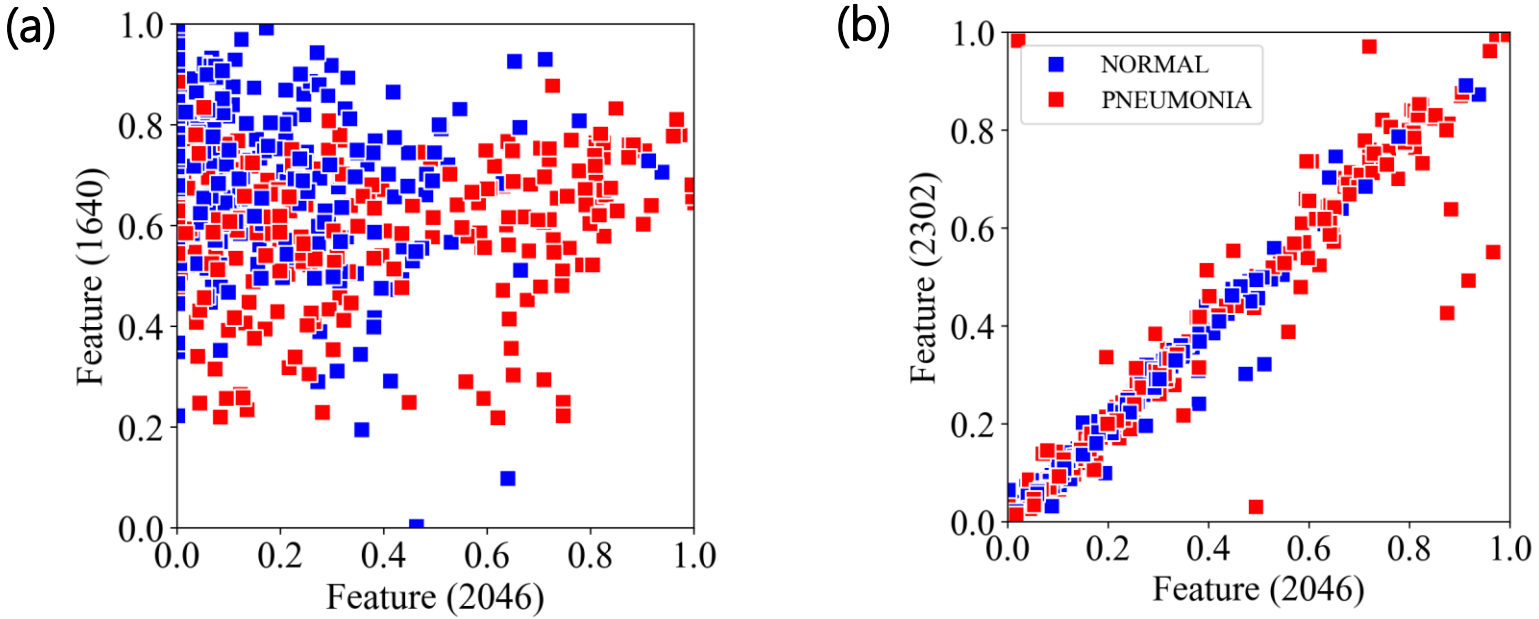


Figure 7. Plots of feature #2046 as a function of (a) #1640 and (b) #2302

Model Hyperparameters - SVM

- We performed a binary classification of the X-ray data with the following hyperparameters:

  - Input Channels = 1 (Greyscale)
  - Image size = 256 x 256 (Rescaled and resized)
  - Kernel = Linear
  - Reg. Parameter = 2×10^{-4} (L2)

Model Results - SVM

- Our SVM model with kernel of linear polynomial kernel achieved a accuracy of 80 %.

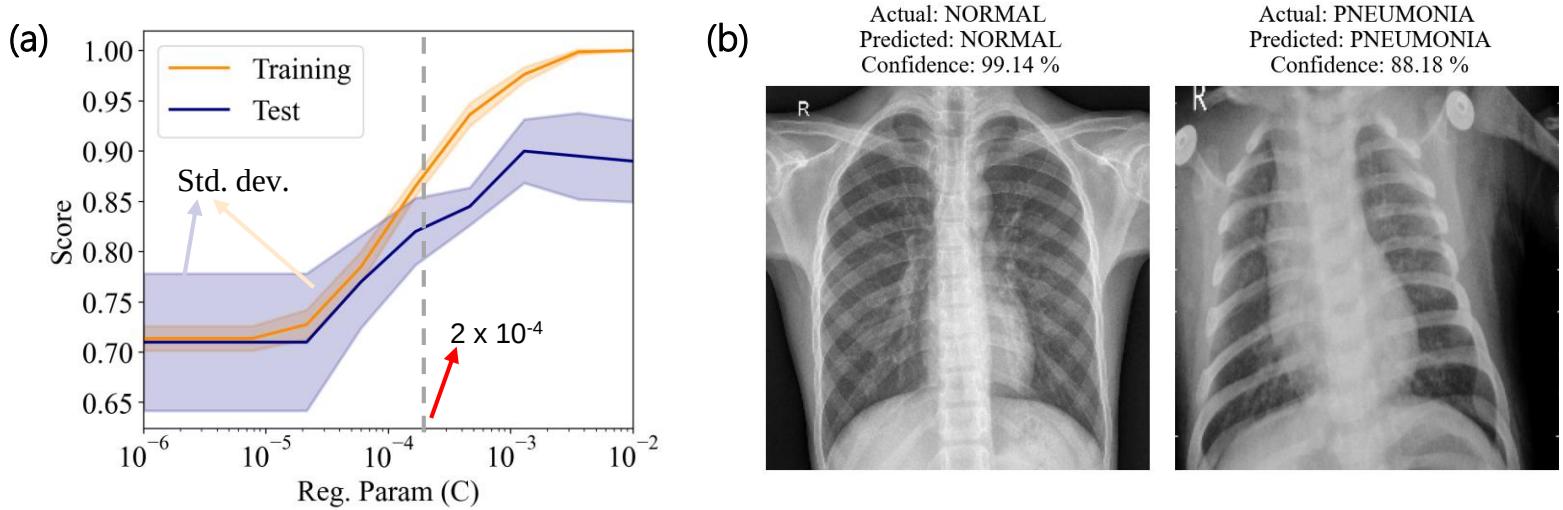


Figure 8. (a) Convergence study with 200 datasets (b) Predictions of SVM model trained with 2944 datasets. Datasets are equally divided for both normal and pneumonia classes

5

Conclusion

CNN vs SVM

- Comparison between CNN and SVM for best models of CNN and SVM

	CNN	SVM
Computational Time	43.39 minutes	4.18 mins
Accuracy	98.44 %	80%

6

Project Details

Code Sources

- CNN and SVM models constructed for other projects were studied from GitHub. Videos from YouTube were also used as reference material.

- Using them as reference, we wrote the code ourselves for the classification task undertaken in this project.

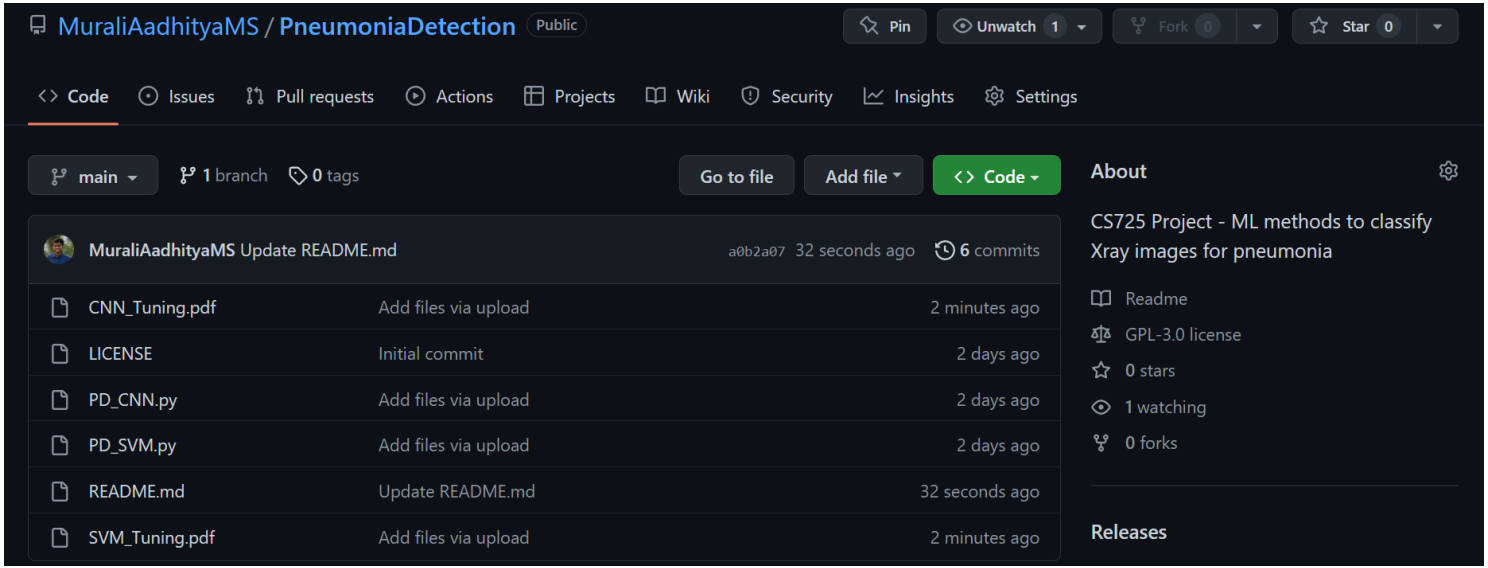
  - https://github.com/codebasics/potato-disease-classification/blob/main/training/potato-disease-classification-model.ipynb
  - https://www.kaggle.com/code/ashutoshvarma/image-classification-using-svm-92-accuracy/notebook

Contributions

- Chitransh Saxena –
  - Code development – CNN
  - Discussion on Presentation and hyperparameter tuning
- Aditya Prasad Roy –
  - Code development – SVM
  - Discussion on Presentation and hyperparameter tuning
- Murali Aadhitya M S –
  - Hyperparameter tuning – CNN, SVM
  - Presentation preparation

Links

- https://github.com/MuraliAadhityaMS/PneumoniaDetection



Thank you.