Feature extraction based

Pneumonia Detection

on chest X-rays

CSE4019 – IMAGE PROCESSING

J–COMPONENT PROJECT DOCUMENT

FALL 2022-2023

Under the Guidance of:

Dr. ANISHA M LAL

Submitted by:

Yashas M 20BCE2607
Sabya Sachi Panda 20BCE0979

INDEX

S No.	Contents	Pg no.
1.	Abstract	3
2.	Keywords	3
3.	Introduction	4
4.	Literature survey	5
5.	Architecture	6
6.	Methodology	7
7.	Image Enhancement Techniques	7
8.	Image Segmentation Techniques	9
9.	Feature Extraction Techniques	12
10.	Performance Evaluation	13
11.	ROC Curves	14
12.	Conclusion	16
13.	References	17

Abstract

Pneumonia is a life-threatening infectious disease affecting one or both lungs in humans commonly caused by bacteria called Streptococcus pneumoniae. Chest X-Rays which are used to diagnose pneumonia need expert radiotherapists for evaluation.

In this project we attempt to detect pneumonia through image enhancement and feature extraction methods. We perform several techniques to derive important features that can help predict pneumonia. Image enhancement is performed using histogram equalization and sharpening filters (high pass filter). Next step is image segmentation done by using otsu thresholding. For edge detection we'll be using sobel filter. We found that the essential features required to detect pneumonia are area of opacity, perimeter of visible lung regions, irregularity index, equivalent diameter, mean, standard deviation of unenhanced image and hu moments.

Keywords

Pneumonia detection, thresholding, segmentation, lung opacity, hu moments

Introduction

Pneumonia detection will determine whether the patient is diagnosed with pneumonia by identifying the region of opacity in the lungs using many image enhancement and segmentation techniques. Further in the project we perform feature extraction and apply machine learning algorithms on the derived data and make our model more feasible to make predictions regarding pneumonia detection.

Pneumonia detection has been proposed using several methods. Most methods revolve around the image processing of chest X-rays using convolutional neural networks (CNNs). Several algorithms have been developed and suggested for this purpose. However, there has also been several novel suggestions such as the analysis of lung ultrasound and ultrasound videos. Images are generally preprocessed using various image processing methods before being classified using CNNs. The dataset used for these experiments have mostly been collected from hospitals, although, a handful of studies were done with public datasets like those from kaggle and the largest available public dataset chestX-ray14. The results of these experiments are promising. However, there is still a lot of scope for better results.

Literature Survey

Author(s)	Method	Dataset	Metrics
Pranav Rajpurkar et al [1]	● Develop a 121 layer CNN and train it on the largest publicly available dataset for CXRs. ● Evaluate the performance of the CNN against traditional radiologists using an F1 score.	ChestX-ray14	F1 score: 0.95
Yoonha Choi et al. [2]	● Develop a classifier using RNA sequencing data. ● This RNA sequencing data is used to identify the UIP pattern for the detection of pneumonia.	Custom dataset from 90 patients	AUC: 0.89
Daniel G. Pankratz et al. [3]	Develop a genomic classifier in tissue obtained by TBB that distinguishes UIP from non-UIP.	Independent test set of specimens from 31 subjects	AUC: 0.92
Mattia Guerra et al. [4]	Testied pneumonia detection techniques against LUS and CXRs and compared their results.	Custom data collected from 190 children	Accuracy: 0.97
Daniel S. Kermany et al. [5]	Used convolution networks to detect retinal diseases and evaluated the AI performance of three models-multiclass comparisons, limited model and binary classifiers	Custom dataset with images captured using OCT techniques	Accuracy: Binary classifiers: 0.98 Limited model: 0.9
Sriram Vijendran & Rahul Dubey [6]	● Develop an algorithm by combining two algorithms based on ELM (MLELM & OSELM), and apply it to the dataset ChestX-ray14. ● The results of this algorithm in detecting pneumonia is then compared to other	ChestX-ray14	Accuracy: Train: 0.96 Test: 0.917

	baseline models like SVM.		
Gilberto de Melo et al. [7]	● A CUDA-based parallel algorithm is developed to detect pneumonia. ● Wavelet features are extracted from high-resolution DICOM images. ● The extracted features are then used to detect pneumonia.	Collected DICOM images for the experiment	Accuracy: 0.876

Architecture

We derive the following features from our above results to build a classifier for pneumonia detection:

- Area of Opacity
- Perimeter of visible lung region
- Irregularity Index
- Equivalent Diameter
- Mean and Standard Deviation of unenhanced image
- Hu Moments

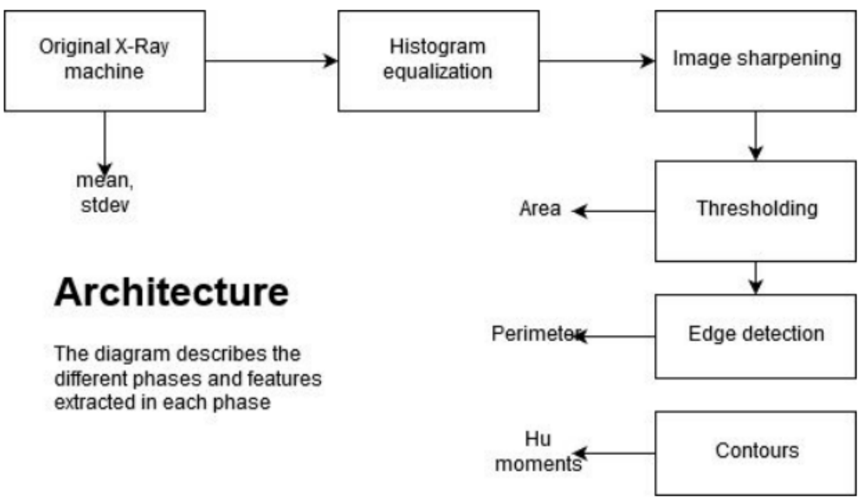


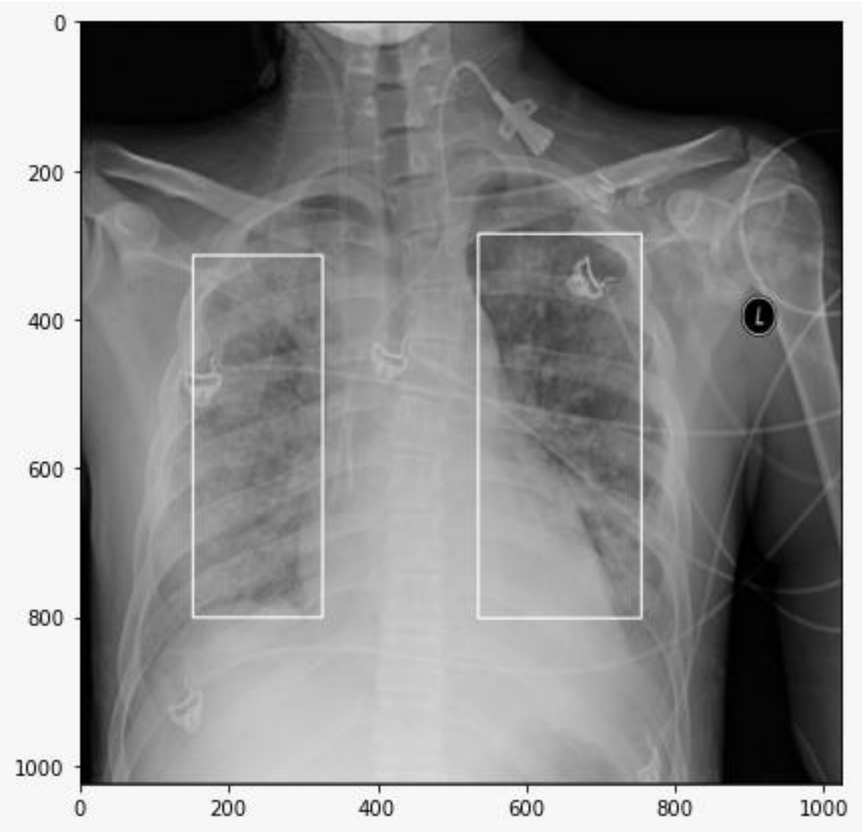
Fig 1. Architecture Diagram

Methodology

- Image Enhancement
- Image Segmentation
- Feature Extraction
- Model building using machine learning applications

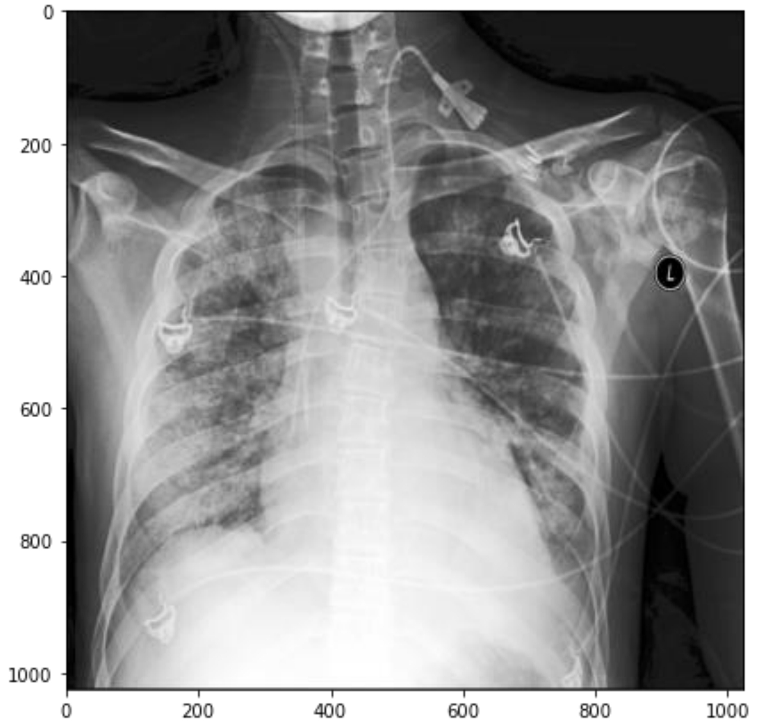
Image Enhancement Techniques

Input image



1. Histogram Equalization

We performed histogram equalization to enhance the image. Histogram equalization is used to improve contrast of image and it does so by spreading out the intensity values that are most frequent or in other words extending intensities. We found that equalisation present a good contrast of the lungs and further accents the presence of opacity.



2. Image Sharpening

To further sharpen the image and isolate opacities, we make use of high pass filter. A high pass filter allows signals above a frequency cut off to pass through and eliminates lower frequency signals essentially sharpening the image and emphasizing the finer details.

We also tried to sharpen the image with unsharp mask filter. It unsharpens the image and use the difference with the original image to sharpen the image.

We found that high pass filter showed more details while unsharp mask filter was still a little blurred and not ideal for feature extraction.

Image sharpening

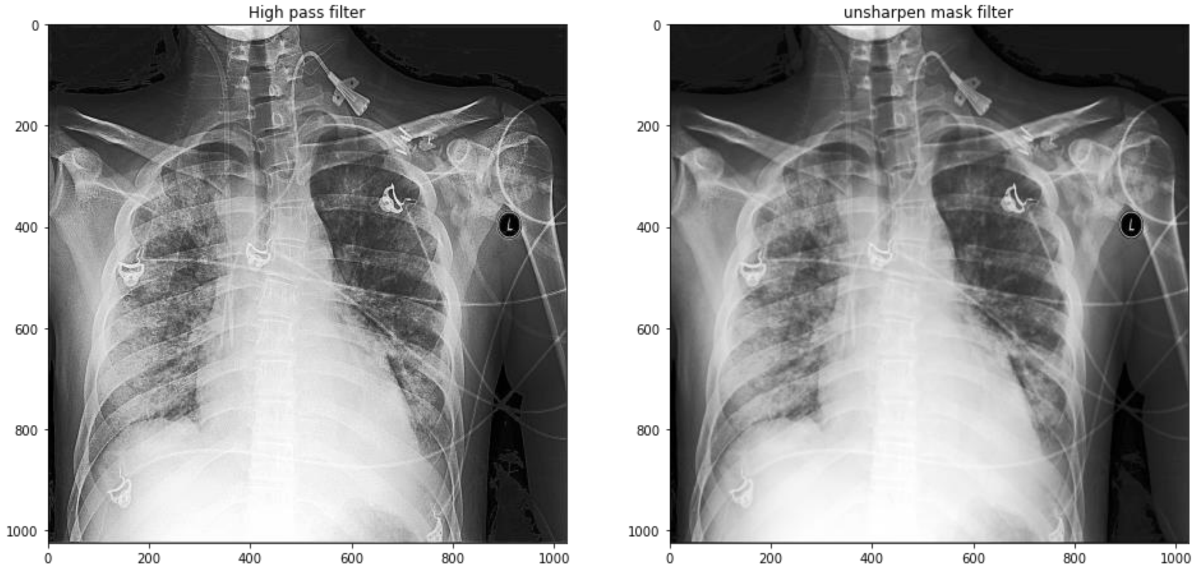


Image Segmentation

1. Thresholding

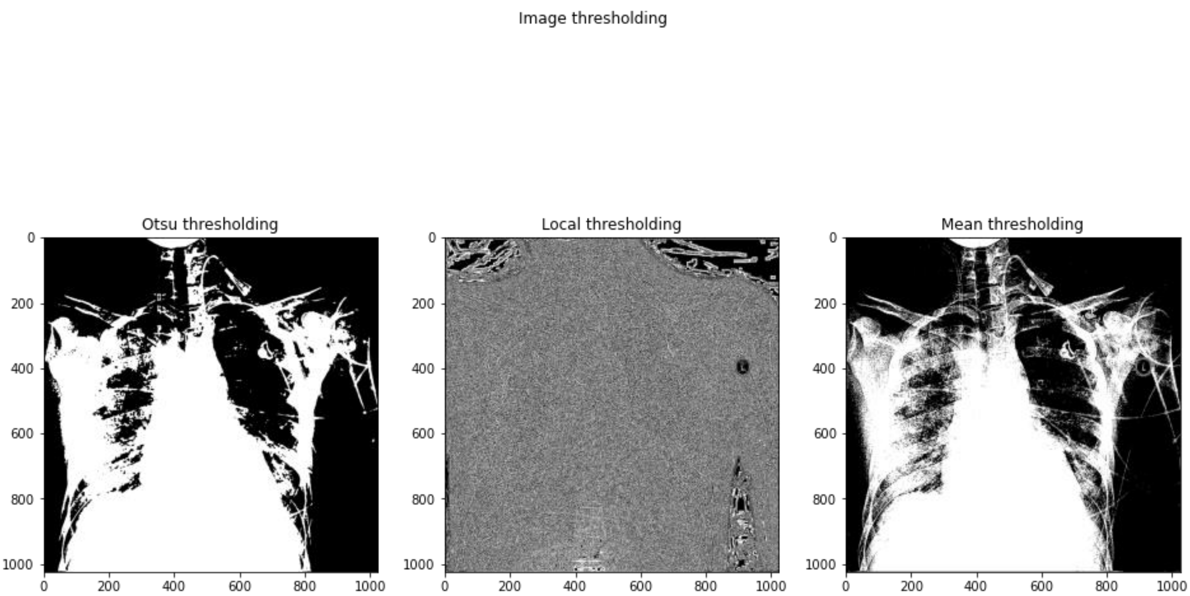
To separate the lungs for analysis we attempted to use three kinds of thresholding methods.

Local thresholding: It converts a grayscale image to black and white by choosing a different threshold value for each pixel based on the analysis of neighbouring pixels.

Mean thresholding: Finds the mean of surrounding pixels to convert gray scale image into binary.

Otsu thresholding: It works directly on the grey level histogram by dividing the grey levels into two classes (a background and a foreground) and then finding the within class variance.

The minimum within class variance is used as the threshold. We chose otsu threshold because it extracted smoother edges and segmented the lungs better.

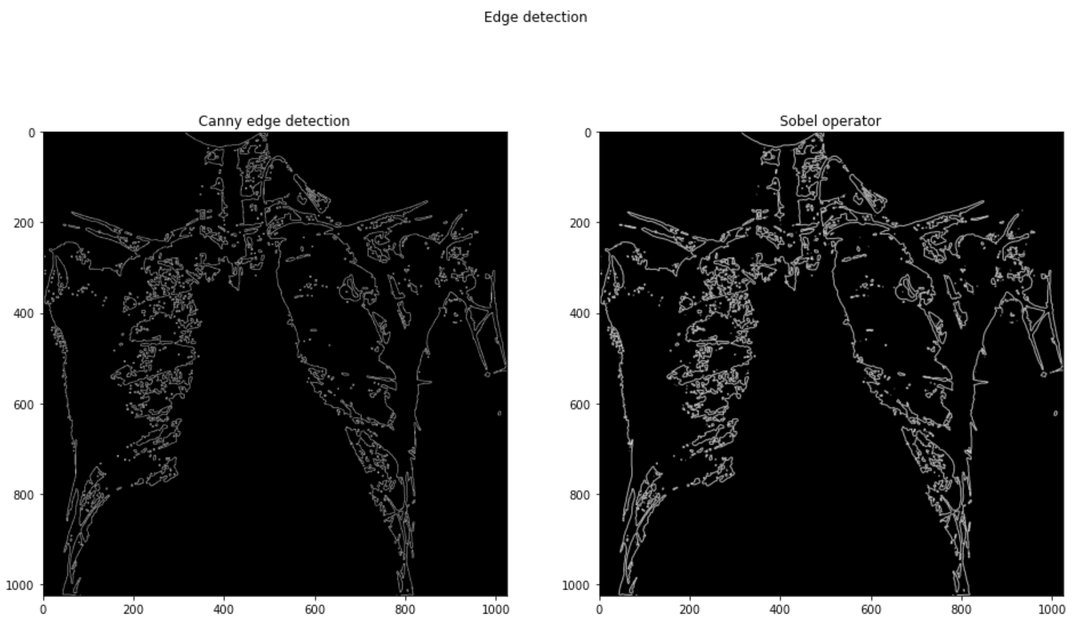


2. Sobel Operator

We made use of Sobel edge detection and also attempted to use canny edge detection.

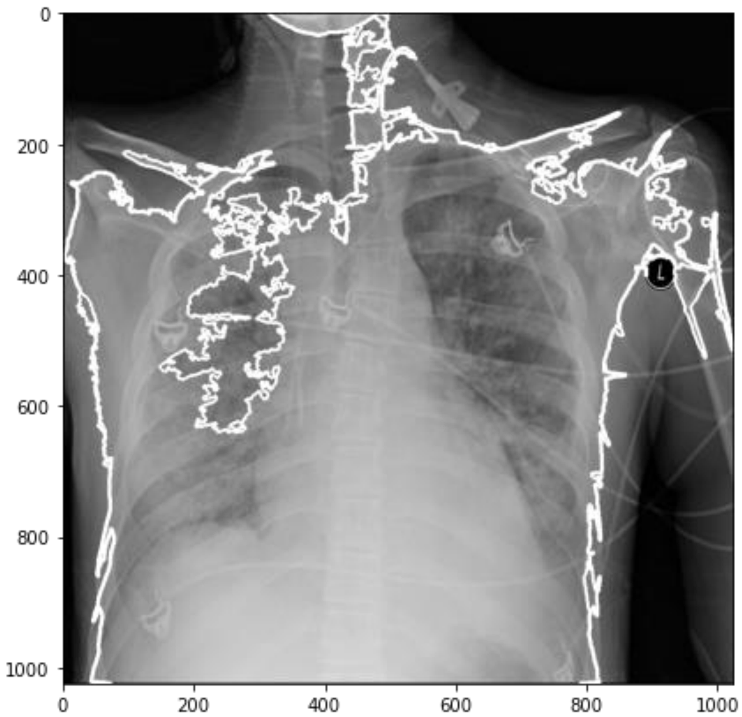
Sobel edge detection: It works by calculating the gradient of intensity at each pixel. It finds the direction of the largest increase from light to dark and the rate of change in that direction.

Canny edge detection: It makes use of gaussian filter to smoothen the images and suppress noise and finds the intensity of the gradient. It then finalises the edges by suppressing all the weak edges.

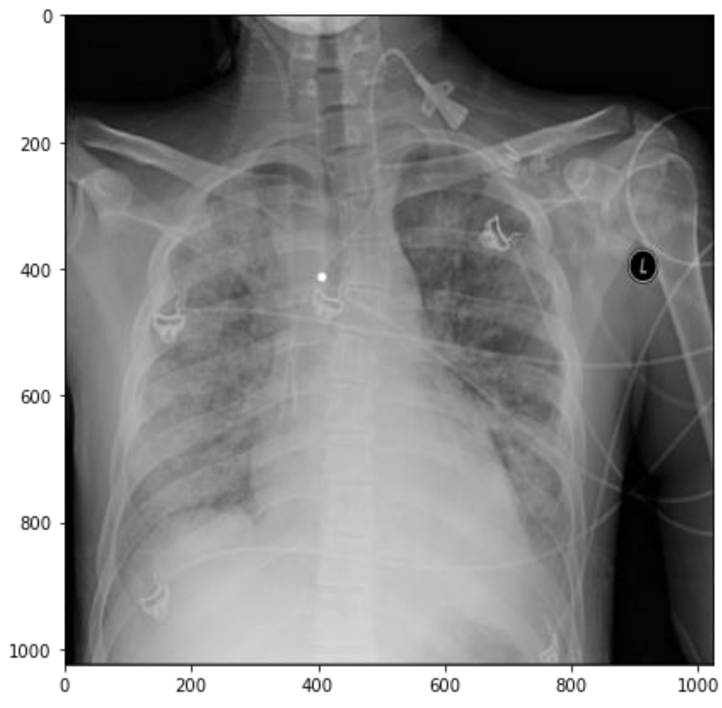


3. Lung Segmentation

After identifying the lung segment, we can extract the center of moment of this segment. Since, all the X-ray images are from the same dimension, this can be a valid feature for prediction.



Center of moments



Feature Extraction Methods

1. Mean and Standard Deviation:

The mean and standard deviation of the raw pixels are calculated before any image enhancement method.

2. Area:

The area of the image is computed by counting the number of white pixels in the binarized image after otsu thresholding.

3. Perimeter:

The perimeter of the image is computed by counting the number of white pixels in the edges of the image detected by sobel operator.

4. Irregularity Index:

Also called compactness is defined as: $(4 * \pi * \text{area})/(\text{perimeter} \wedge 2)$

5. Equivalent diameter:

The formula for Equivalent diameter is: $\sqrt{4} * \text{area}/\pi$

6. Hu Moments:

Hu moments are invariant moments. The values are generally invariant to translation, rotation and scaling. We will log moments to make it easy to compare and drop the 3rd moment as it depends on the other values and 7th moment as it distinguishes mirror images and there are no flipped images in the dataset.

Performance evaluation

On testing with different machine learning models and classifiers the following results are obtained:

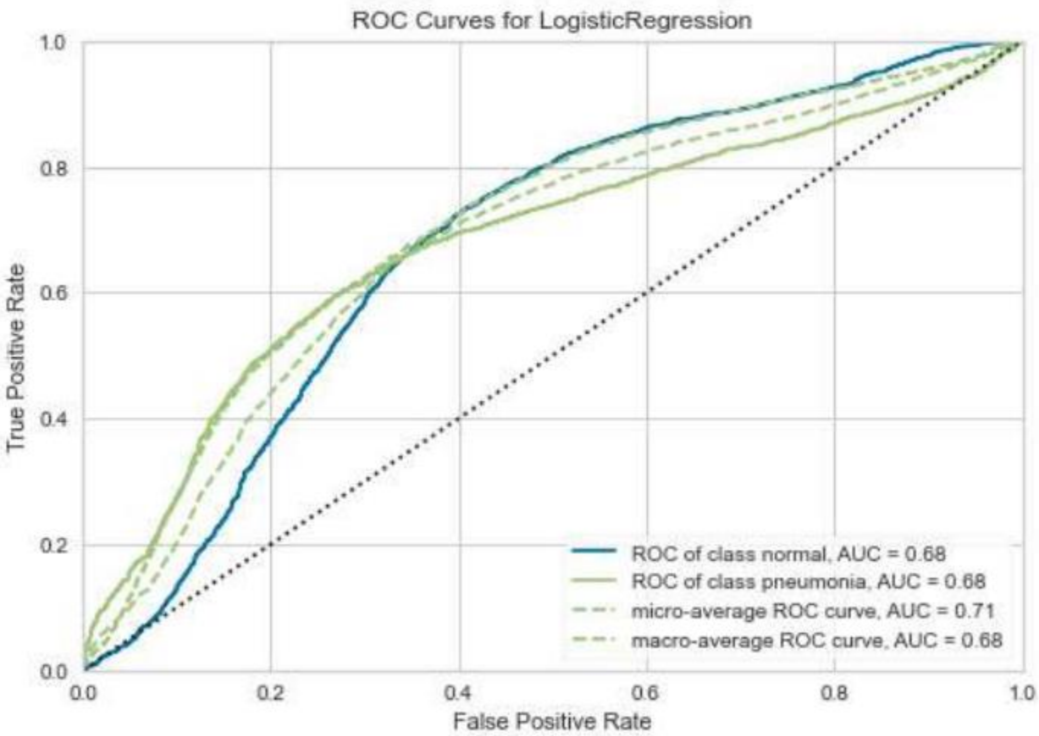
		Logistic Regression	Random forest Classifier	Gradient Boosting Classifier	SVC	KNeighbors Classifier
Train data metrics	Accuracy	0.6598631826847594	0.8253897050577549	0.8897611304250308	0.6562745317932039	0.7262532241785353
	Precision Score	0.6411012782694199	0.8227329974811083	0.8890865954922894	0.7178456591639871	0.7094109195402298
	Recall	0.3615192680898253	0.724424729692265	0.8311616301635708	0.24757416135292487	0.5475464374826726
	F1 Score	0.4623293742244282	0.7704555506413091	0.8591488751970197	0.3681715110286539	0.6180566421530277
	ROC AUC score	0.6120214042897338	0.8091991821719329	0.8803642425770773	0.5907362332188354	0.697596194259227
Test data metrics	Accuracy	0.6626303397241843	0.7803565422132526	0.7954927682475614	0.6505213588967373	0.7058526740665994
	Precision Score	0.659217877094972	0.762291169451074	0.7693701857725419	0.7188755020080321	0.6814159292035398
	Recall	0.34345114345114347	0.664033264033264	0.7060291060291061	0.2232848232848233	0.5122661122661123
	F1 Score	0.4516129032258065	0.7097777777777778	0.7363399826539462	0.3407360406091371	0.5848563968668408
	ROC AUC score	0.6114318693816011	0.761697513123664	0.7811422005717402	0.5819897711453769	0.6748000993411895

Based on the above results, the Gradient Boosting Classifier algorithm is found to provide the highest accuracy **(0.89)** among other machine learning models.

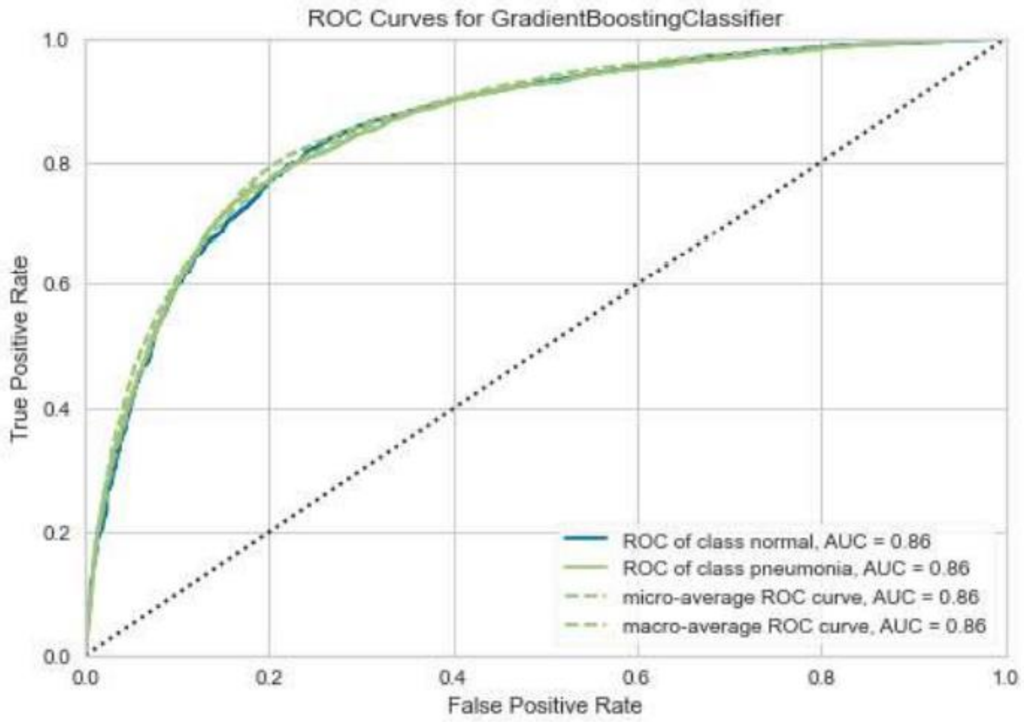
CODE: [Daaliyash/Pneumonia-Detection: A model to identify the presence of pneumonia in a patient's chest X-ray (github.com)]

ROC Curves

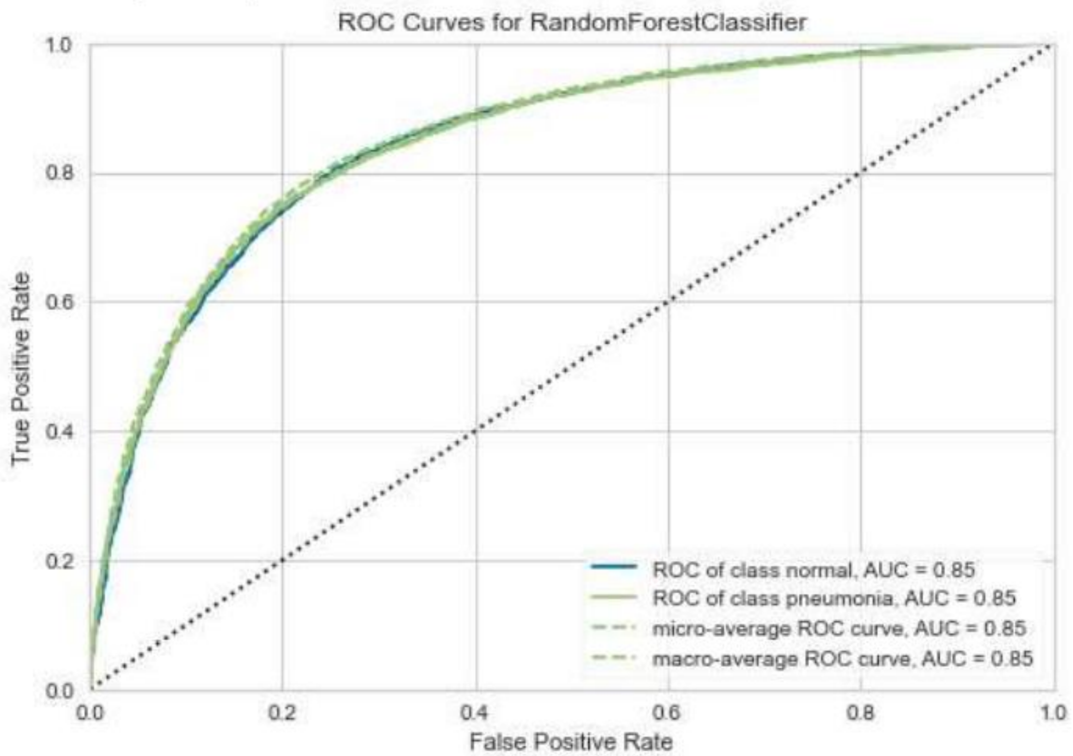
Logistic Regression



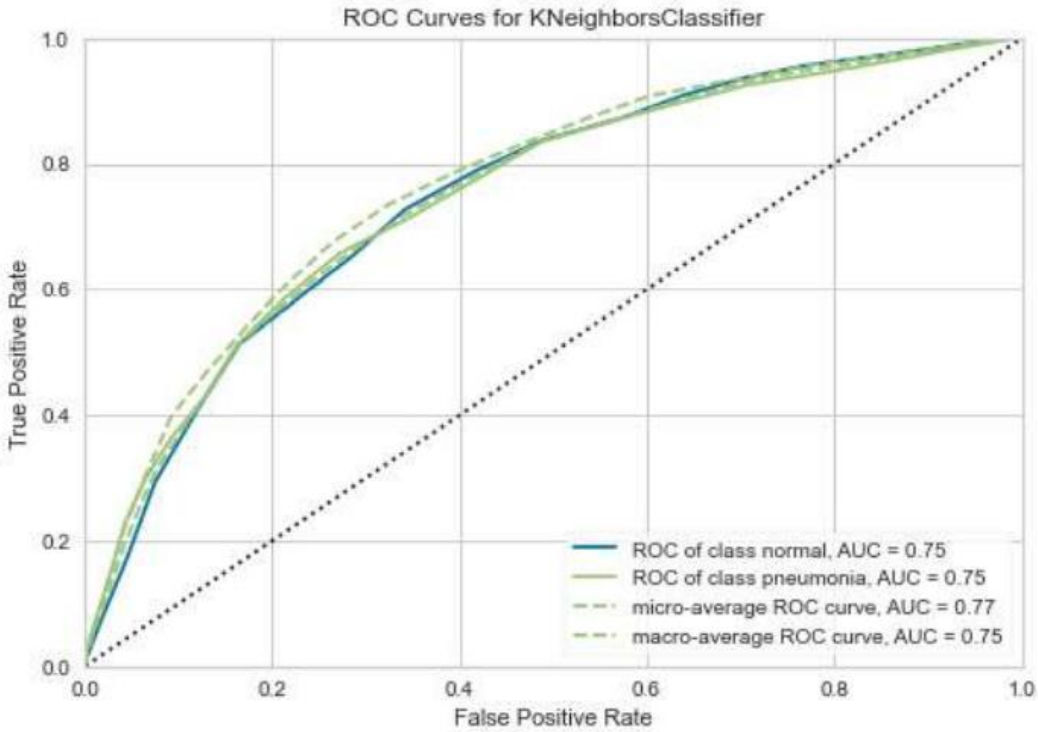
Gradient Boosting Classifier



Random Forest Classifier



K Neighbors Classifier



Conclusion

We were able to successfully detect pneumonia with impressive results. Our model was able to achieve an roc_auc of 0.86. All models were extremely sensitive to noise in the image. Interesting insights regarding what type of features influence the prediction of pneumonia was identified. In the future, we want to explore more advanced features like zernike moments from the region of interest. We would like to develop deep learning algorithms that can predict with higher accuracy and try to interpret the features detected by the network. There is a lot of scope for future research in this area as there are still challenges in the detection and diagnosis of pneumonia. One major challenge being the similarities in appearance of infiltrates between pneumonia and other pulmonary diseases. Further research can also include images other than chest X-rays like ultrasound or videos.

References

[1] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.

[2] Choi, Y., Liu, T. T., Pankratz, D. G., Colby, T. V., Barth, N. M., Lynch, D. A., ... & Huang, J. (2018). Identification of usual interstitial pneumonia pattern using RNA-Seq and machine learning: challenges and solutions. BMC genomics, 19(2), 147-159.

[3] Pankratz, D. G., Choi, Y., Imtiaz, U., Fedorowicz, G. M., Anderson, J. D., Colby, T. V., ... & Martinez, F. J. (2017). Usual interstitial pneumonia can be detected in transbronchial biopsies using machine learning. Annals of the American Thoracic Society, 14(11), 1646-1654.

[4] Guerra, M., Crichiutti, G., Pecile, P., Romanello, C., Busolini, E., Valent, F., & Rosolen, A. (2016). Ultrasound detection of pneumonia in febrile children with respiratory distress: a prospective study. European journal of pediatrics, 175(2), 163-170.

[5] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5), 1122-1131.

[6] Vijendran, S., & Dubey, R. (2019, March). Deep online sequential extreme learning machines and its application in pneumonia detection. In 2019 8th International Conference on Industrial Technology and Management (ICITM) (pp. 311-316). IEEE.

[7] de Melo, G., Macedo, S. O., Vieira, S. L., & Oliveira, L. L. (2018, October). Classification of images and enhancement of performance using parallel algorithm to detection of pneumonia. In 2018 IEEE International Conference on Automation/XXIII Congress of the Chilean Association of Automatic Control (ICA-ACCA) (pp. 1-5). IEEE.

[8] Barrientos, R., Roman-Gonzalez, A., Barrientos, F., Solis, L., Correa, M., Pajuelo, M., ... & Zimic, M. (2016, November). Automatic detection of pneumonia analyzing ultrasound digital images. In 2016 IEEE 36th Central American and Panama Convention (CONCAPAN XXXVI) (pp. 1-4). IEEE.

[9] Rajaraman, S., Candemir, S., Kim, I., Thoma, G., & Antani, S. (2018). Visualization and interpretation of convolutional neural network

predictions in detecting pneumonia in pediatric chest radiographs. Applied Sciences, 8(10), 1715.

[10] Pons, E., Braun, L. M., Hunink, M. M., & Kors, J. A. (2016). Natural language processing in radiology: a systematic review. Radiology, 279(2), 329-343.

[11] Wiemken, T. L., Furmanek, S. P., Mattingly, W. A., Guinn, B. E., Cavallazzi, R., Fernandez-Botran, R., ... & Ramirez, J. A. (2017). Predicting 30-day mortality in hospitalized patients with community-acquired pneumonia using statistical and machine learning approaches. The University of Louisville Journal of Respiratory Infections, 1(3), 10.

[12] Saul, C. J., Urey, D. Y., & Taktakoglu, C. D. (2019). Early Diagnosis of Pneumonia with Deep Learning. arXiv preprint arXiv:1904.00937.

[13] Melo, G. D., Macedo, S. O., Vieira, S. L., & Oliveira, L. G. L. (2018). Classification of images and enhancement of performance using parallel algorithm to detection of pneumonia. 2018 IEEE International Conference on Automation/XXIII Congress of the Chilean Association of Automatic Control (ICA-ACCA). doi: 10.1109/ica-acca.2018.8609734.