Feature Extraction Based Pneumonia Detection

A PROJECT REPORT

Submitted in fulfillment

in

Digital Image Processing

by

SURYA K	16BIT0029
SHWETA VIJAY	16BIT0423

Under the Guidance of

PROF. PRABUKUMAR M

School Of Information Technology And Engineering

November, 2019

ACKNOWLEDGEMENTS

We sincerely thank Dr.G.Viswanathan - Chancellor, VIT University, for creating an opportunity to use the facilities available at VIT. We also thank **Prof.Prabukumar M Department of Information Technology**, VIT University, for giving us the opportunity to do this project. We also thank the Dean and entire department of Information Technology, School of Information Technology and Engineering, for giving us this opportunity.

DECLARATION BY THE CANDIDATE

We hereby declare that the project report entitled **"FEATURE EXTRACTION BASED PNEUMONIA DETECTION"** submitted by us to VIT University, Vellore in partial fulfillment of the requirement for the award of the degree of B.Tech is a record of J component of project work carried out by me under the guidance of **PROF. PRABUKUMAR M**. We further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Vellore

Date:4th November ,2019

Surya K	16BIT0029
Shweta Vijay	16BIT0423

CERTIFICATE

This is to certify that the project work titled **“Feature Extraction Based Pneumonia Detection"** that is being submitted by

SURYA K	16BIT0029
SHWETA VIJAY	16BIT0423

for **Digital Image Processing ()** is a record of bonafide work done under my supervision. The contents of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted for any other CAL course.

Place : Vellore

Date : 30/10/2017

Signature of the students :

Surya K	16BIT0029
Shweta Vijay	16BIT0423

Signature of faculty : (Prabukumar M)

Table of Contents

S.No	Title	Page
1	Abstract	6
2	Introduction	7
3	Literature Review	8
4	Summary	16
5	Proposed method architecture	17
6	Analysis and results	23
7	Conclusion	26
8	Bibliography	27

Abstract

Pneumonia is responsible for over 15% of deaths of children all over the world under the age of 5. Despite how common the disease is, accurately detecting it has proved to be extremely difficult and usually requires the review of a chest radiograph (CXR) by highly trained radiologists and confirmation through clinical trials. Recent trends in high performance computing and parallel and distributed computing have greatly contributed in the use of computational algorithms for pneumonia detection. 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. We found that the essential features required to detect pneumonia are area of opacity, perimeter of visible lung regions, irregularity index, equivalent diameter, mean, sd of unenhanced image and hu moments. We tried classification through several models and recorded the results. We found that the gradient boosting classifier gave the highest roc_auc (Train: 0.86 Test: 0.78).

Introduction

Pneumonia is an infection caused by a variety of possible pathogens including viruses, bacteria, fungi, etc. Once the infection sets in, one or both the air sacs in the lungs get inflamed. This may result in the air sacs being filled with pus and may cause fits of coughing, phlegm, fever and difficulty in breathing.

It has been found that children under the age of 5 and the elderly over the age of 65 are more susceptible to the onset of pneumonia. While the severity of the disease may vary, if left untreated it may eventually lead to respiratory failure or death.

Typically, a doctor may perform a chest x-ray, chest CT, chest ultrasound or needle biopsy to detect and diagnose pneumonia by checking the lungs for infiltrates, enlarged lymph nodes or pleural effusions.

The disease is quite difficult to detect and may differ from person to person depending on their age, immune system, etc, and the symptoms themselves mimic more benign afflictions like the flu.

Many factors play into the processing of CXRs such as the position of the patient, depth of inspiration eyc. This further complicates the interpretation of CXRs.

However, with the advent of modern computer technology and image processing, it has become easier to diagnose pneumonia and provide treatment on time.

Literature review

(Pranav Rajpurkar et al.,2017)[1] develop an algorithm CheXNet which detects pneumonia using the frontal view of chest X-rays. CheXNet is a 121 layer convolutional neural network that was trained on the largest publicly available dataset for chest X-rays, Chest-X-ray14. The performance of the algorithm is compared with practicing academic radiologists using an F1 score and found to be significantly better at detecting pneumonia.

(Yoonha Choi et al.,2018)[12] develop a classifier using RNA sequencing data. The RNA sequencing data identifies the usual interstitial pneumonia (UIP) pattern for the detection of pneumonia. The paper addresses several issues like limited sample size, disease heterogeneity, and developed a highly accurate and robust classifier for the classification of UIP.

(Daniel G. Pankratz et al., 2017)[14] develop a genomic classifier in tissue obtained by TBB that distinguishes UIP from non-UIP, trained against central pathology as the reference standard.

(Mattia Guerra et al.,2015)[18] tested the extent to which lung ultrasound (LUS) detection could be used to diagnose pneumonia and other respiratory diseases in febrile children as compared with chest x-rays (CXR). The experiment found that LUS is as competent as CXRs and in some cases outperformed CXRs. The paper concluded that in the hands of trained clinicians, LUS can be a very useful supplement for the diagnosis of pulmonary diseases like pneumonia.

(Daniel S. Kermany et al.,2018)[17] use transfer learning, a method to build AI systems using convolution networks, to use the knowledge gained from training one particular dataset on another. The paper used convolution networks to detect retinal diseases and evaluated the AI performance of three models-multiclass comparisons, limited model and binary classifiers. The authors also performed occlusion testing and successfully identified the areas of importance in the convolution network that allowed an accurate diagnosis.

(Ronald Barrientos et al.,2016)[6] present a method for the detection and diagnosis of pneumonia using ultrasound imaging. In the paper, this diagnosis is done through the analysis of patterns found in rectangular segments from the images. With this approach, the paper has been able to establish high accuracy for the diagnosis of pneumonia.

(Sriram Vijendran & Rahul Dubey,2019)[7] take advantage of deep learning and extreme learning machine (ELM) and combine two algorithms based on ELM, multilayer extreme learning machine (MLELM) and online sequential extreme learning machine (OSELM), and apply it to the dataset ChestX-ray14. The results of this experiment are then compared with other, more popular algorithms like SVM, CNN, etc, and found to be more accurate.

(Gilberto de Melo et al.,2019)[8] propose the automation of pneumonia detection using a CUDA-based parallel algorithm. This algorithm is designed to extract wavelet features from high-resolution DICOM images. The extracted features are then used to detect pneumonia. The proposed algorithm improves the computing speed by 12.75 times. KNN is then used to classify the images.

(Sivaramakrishnan Rajaraman,2018)[9] examine, evaluate, visualize and make use of custom CNNs to detect pneumonia and differentiate between bacterial and viral pneumonia infection. The paper proposes a novel way to localize the region and observes that the customized VGG16 results in the highest accuracy for the detection of pneumonia and distinction between bacterial and viral pneumonia.

(Gino Soldati & Marcello Demi,2017)[29] aim to describe the differences between the sonographic interstitial syndrome related to lung diseases and that related to cardiogenic edema in the light of current knowledge regarding the pleural plane's response to ultrasound waves.

(Shubhangi Khobragade et al.,2016)[11] propose lung segmentation; lung feature extraction and its c1assification using artificial neural network technique for the detection of lung diseases such as TB; lung cancer and pneumonia. From the experiment, it can be said that histogram equalization and image segmentation give good results.

(Kristina Dietert et al.,2018)[28] adopt a pattern recognition-based software to quantify user-defined pathologies from whole slide scans of lungs infected with Streptococcus pneumoniae or influenza A virus and then compared to PBS-challenged lungs

(O.Zenteno et al.,2016)[30] present a pneumonia detection algorithm based on the measurement of the fundamental bandwidth downshift over depth of ultrasound radiofrequency (RF) signals. The results of this algorithm was over 90% accuracy.

(Ewoud Pons, MD et al.,2016)[2] propose a method where NLP is used to process large amounts of data to detect pneumonia.

(Jufriadif Na`am et al.,2017)[15] use edge detection to obtain information that is more recognizable for the diagnosis of pneumonia and other lung diseases. These image processing methods were applied to a dataset collected from infant patients treated at Central Public Hospital (RSUP).

(Yu. Gordienko et al.,2018)[16] display the efficiency of lung segmentation and bone shadow exclusion techniques in the diagnosis of pneumonia and other pulmonary diseases such as lung cancer. The results show the high efficiency of the demonstrated pre-processing techniques. The preprocessed dataset without bones proved to show a higher level of accuracy.

(Timothy L Wiemken et al.,2017)[3] aim to assess several statistical and machine learning models for their ability to predict 30-day mortality in hospitalized patients with CAP.

(Enes Ayan & Halil Murat Ünver,2019)[5] adopt two different convolution networks vgg16 and xception to diagnose pneumonia and compare their results. It was found in the study that vgg16 achieved a more successful result.

(Deniz Yagmur Urey et al.,2019)[4] propose a deep learning architecture for classification by training models with modified images, through multiple steps of preprocessing. Classification is performed using residual network architecture and convolutional neural networks.

(Ying Sha & May D. Wang,2017)[20] develop a gated recurrent unit-based recurrent neural network with hierarchical attention for mortality prediction, and then, using the diagnostic codes from the Medical Information Mart for Intensive Care, the model was evaluated. The results show that the prediction accuracy of the proposed model outperforms baseline models and has high interoperability.

(Marios Anthimopoulos et al.,2016)[21] propose a 5 layers CNN for the detection of interstitial lung diseases (ILD) like pneumonia. The paper successfully shows effectiveness of CNNs in the detection of ILDs by classifying lung CT patches into 7 classifications. These include 6 different ILD patterns and healthy tissue. The paper also proposes the continuation of this project with three dimensional CT scans and the integration of CNNs with CAD systems.

(Louis Rosenberg et al.,2018)[22] applies artificial swarm intelligence to explore if a small group of radiologists can improve their accuracy in diagnosing pneumonia from chest X-rays. Performance data was collected for individual radiologists generating diagnoses alone, as well as for small groups of radiologists working together to generate diagnoses as a real-time ASI system. It was noted that the ASI reduced diagnostic errors by 33%. The ASI also performed better than the state of the art CheXNet that was developed at Stanford (22% more accurate).

(James A. Nichols et al.,2018)[23] review the fundamentals and algorithms behind machine learning and highlight specific approaches to learning and optimisation. They then summarise the applications of ML to medicine. In particular, they showcase recent diagnostic performances, and caveats, in the fields of dermatology, radiology, pathology and general microscopy.

(Pedro Cisneros-Velarde et al.,2016)[24] proposes the analysis of ultrasound video for the detection of pneumonia. This method is applied by analysing small chunks of the video using an image processing algorithm to derive the ultrasound statistics. The algorithm can be used directly as a classifier for pneumonia and was tested on videos of children between 3 to 5 years of age. The paper also suggests the algorithm as an extremely efficient supplement to other traditional pneumonia detection processes.

(Senthil Kumar Veeramani1 & Ezhilarasi Muthusamy,2015)[25] suggest adaptive mean filtering for the preprocessing stage. These preprocessed images are then used to extract features using Harlick feature extraction. The extracted features are selected by applying particle swarm optimization and differential evolution feature selection. In the final stage classifiers are used to perform the classification for the lung diseases. The experiment exhibits higher accuracy, sensitivity and specificity than other baseline models.

(Franklin Barrientos et al.,2016)[26] propose a method by using lung ultrasound echography to detect pneumonia so as to compensate for the lack of professional doctors in rural regions. This is done using pattern recognition and eliminating noise i.e the image portion of the skin.

(Ganesh Kumar T et al.,2018)[27] use image processing methods to enhance the images of pneumonia bacteria. This is done based on two domains. Single scale retinex, Multiscale retinex, wiener filter and median filter were used to enhance the images. The paper concluded that the median filter performed best to identify pneumonia bacteria for grey-scale images and multiscale retinex worked best on coloured images.

(Mark Cicero, MD, BESc,2017)[10] use CNN to classify frontal chest radiographs to detect pneumonia. The purpose of this paper is to show the significance of using CNN to provide an initial interpretation of CXRs to detect pneumonia.

(Justin Ker et al.,2018)[19] cover key research areas and applications of medical image classification, localization, detection, segmentation, and registration. They conclude by discussing research obstacles, emerging trends, and possible future directions.

(John R. Zech et al.,2018)[13] aim to show the extent to which CNNs generalize to new data across three hospital systems. The paper shows that contrary to what might be expected, despite the increasing interest in using CNNs to analyze medical images for pneumonia detection, CNNs might fail to generalize to new data. Pneumonia-screening CNNs achieved better internal than external performance in 3 out of 5 natural comparisons.

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. [12]	● 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. [14]	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. [18]	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. [17]	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
Ronald Barrientos et al. [6]	Pneumonia diagnosis is done through the analysis of patterns found in rectangular segments from the images and specificity and sensitivity measured.	Collected from children under 5yo from hospital del Nino in Lima	Sensitivity: 0.915 Specificity: 1.00

Sriram Vijendran & Rahul Dubey [7]	• 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 baseline models like SVM.	ChestX-ray14	Accuracy: Train: 0.96 Test: 0.917
Gilberto de Melo et al. [8]	• 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
Sivaramakrishnan Rajaraman et al. [9]	• Uses the CNN VGG16 to localize and detect pneumonia. • This information is then used to differentiate between bacterial and viral pneumonia.	CXRs from children under 5 from Guangzhou women and children's medical centre	Detection: Accuracy: 0.962 F1 score: 0.918 Distinguish: Accuracy: 0.936 F1 score: 0.951
Gino Soldati & Marcello Demi [29]	Distinguish between different B lines (discrete laser-like vertical hyperechoic lines arising from the pleural plane) that are an indication of f sonographic interstitial syndrome (SIS)	Custom dataset	Not Mentioned
Shubhangi Khobragade et al. [11]	• Lung feature extraction is performed. • The features are classified using artificial neural network techniques. • Histogram equalization and image segmentation are performed..	Chest X-rays of 80 patients from Sasoon hospital, Pune	Accuracy: 0.92
Kristina Dietert et al. [28]	A pattern recognition-based software is developed to quantify user-defined pathologies from whole slide scans of lungs infected with influenza A virus and then compared to PBS-challenged lungs	Stained slides were automatically digitized using the Aperio CS2 scanner	Not Mentioned
O.Zenteno et al. [30]	• RF-data was obtained from lung ultrasound samples of children aged between six months and five years. • Sampling was performed using a 6.6 MHz linear transducer. • Finally, a descriptor function was build	Data was collected from children between the ages 6 months to 5 years.	Specificity: 0.927 Sensitivity: 0.804 Accuracy: 0.891

	concatenating all fitted values from the RF-lines for each frame respectively		
Ewoud Pons, MD et al. [2]	NLP is used to process large amounts of data to detect pneumonia.	Custom dataset	F1 score: 0.98
Jufriadif Na`am et al. [15]	• Edge detection is performed to obtain information that is more recognizable for the diagnosis of pneumonia and other lung diseases. • These image processing methods are then applied to a dataset collected from infant patients treated at Central Public Hospital (RSUP).	CXRs of infant patients treated at Central Public Hospital (RSUP)	Accuracy: 0.602
Yu. Gordienko et al. [16]	Lung segmentation and bone shadow exclusion techniques are performed to diagnose pneumonia and other pulmonary diseases.	JSRT	Accuracy: Train: 0.97 Test: 0.71
Timothy L Wiemken et al. [3]	Six different statistical and/or machine learning methods were used to develop patient level prediction models for hospitalized patients with CAP. For each model, nine different statistics were calculated to provide measures of the overall performance of the models.	University of Louisville (UofL) Pneumonia Study database	Specificity: 83.5% Sensitivity: 69.8%
Enes Ayan & Halil Murat Ünver [5]	Performed the diagnosis of pneumonia using two CNNs vgg16 and Xception and compared their results.	Dataset consisting of 5856 frontal chest X-ray images provided by Kermany et al	Accuracy: Vgg16: 0.87 Xception: 0.82
Deniz Yagmur Urey et al. [4]	Perform classification using residual network architecture and convolutional neural networks and try to achieve an accuracy higher than the previously achieved accuracy by CheXNet.	Kaggle	Accuracy: 0.7873
Ying Sha & May D. Wang [20]	• Develop a gated recurrent unit-based recurrent neural network with hierarchical attention for mortality prediction. • Using the diagnostic codes from the Medical Information Mart for Intensive Care, the model was then evaluated.	MIMIC-III	F1 score: 0.57
Marios Anthimopoulos et al. [21]	• A 5 layer CNN is used for the detection of ILDs like pneumonia. • The lung CT patches are then classified	14696 image patches, derived by 120 CT scans	Accuracy: 0.855

	into 7 classes including 6 different ILD patterns and healthy tissue.	from different scanners and hospitals	
Louis Rosenberg et al. [22]	• Performance data was collected for individual radiologists generating diagnoses alone as well as for small groups of radiologists working together to generate diagnoses as a real-time ASI system. • The results of both systems were then compared.	Set of 50 chest X-rays by working together as a real-time system	Accuracy: Binary classification: 0.82 MAE: 0.086
James A. Nichols et al. [23]	• Review the fundamentals and algorithms behind machine learning and highlight specific approaches to learning and optimisation. • Summarise the applications of ML to medicine.	CAMELYON 16 dataset	Not Mentioned
Pedro Cisneros-Velarde et al. [24]	• Analysis of ultrasound video is conducted for the detection of pneumonia. • This method is applied by analysing small chunks of the video using an image processing algorithm to derive the ultrasound statistics. • The algorithm is then used directly as a classifier for pneumonia and was tested on videos of children between 3 to 5 years of age.	Videos were taken from children between 3 to 5 and classified by professional doctors.	AUC: 0.7851-0.9177
Senthil Kumar Veeramani1 & Ezhilarasi Muthusamy [25]	• Use adaptive mean filtering for the preprocessing stage. • These preprocessed images are then used to extract features using Harlick feature extraction. • The extracted features are selected by applying particle swarm optimization and differential evolution feature selection. • In the final stage, classifiers are used to perform the classification for the lung diseases.	ChestX-ray14	Accuracy: 0.9
Franklin Barrientos et al. [26]	• Data was collected from 23 23 children under 5 from Hospital Nino in Lima. • LUS were obtained using a SONIXTOUCH system. • The surrounding skin (noise) was	Custom dataset from 23 children under 5 from Hospital Nino in Lima	MQE: 11.17 pixels

	removed from the images. ● The filtered images were used to locate the pleural zone.		
Ganesh Kumar T et al. [27]	● The images were filtered using four filters. Namely, single scale retinex, multiscale retinex, wiener filter and median filter. ● The results of these filters were then compared for both grey-scale and coloured images.	Custom dataset	PSNR value: Grey-scale: Median filter: 49.5db Coloured: SSR: 49.43db
Mark Cicero, MD, BESc et al. [10]	Use CNN to classify frontal chest radiographs to detect pneumonia.	35,038 adult posterior-anterior chest radiographs from 2005-2015	AUC: 0.964 Sensitivity: 0.91 Specificity: 0.91
Justin Ker et al. [19]	Cover key research areas and applications of medical image classification, localization, detection, segmentation, and registration.	Kaggle	Accuracy: 0.88-0.98
John R. Zech et al. [13]	● Use CNNs to generalize data (for the detection of pneumonia) across three hospitals. ● These results are then compared to see if CNNs can be generalized over different datasets.	Data was collected from 3 hospitals:Indiana University Network for Patient Care, Mount Sinai Hospital, NIH	CNN trained to detect at MSH: AUC (internal): 0.802 AUC (external): 0.717 CNN trained to detect at MSH-NIH: AUC (internal): 0.931 AUC (external): 0.815

Summary

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.

Proposed method and architecture

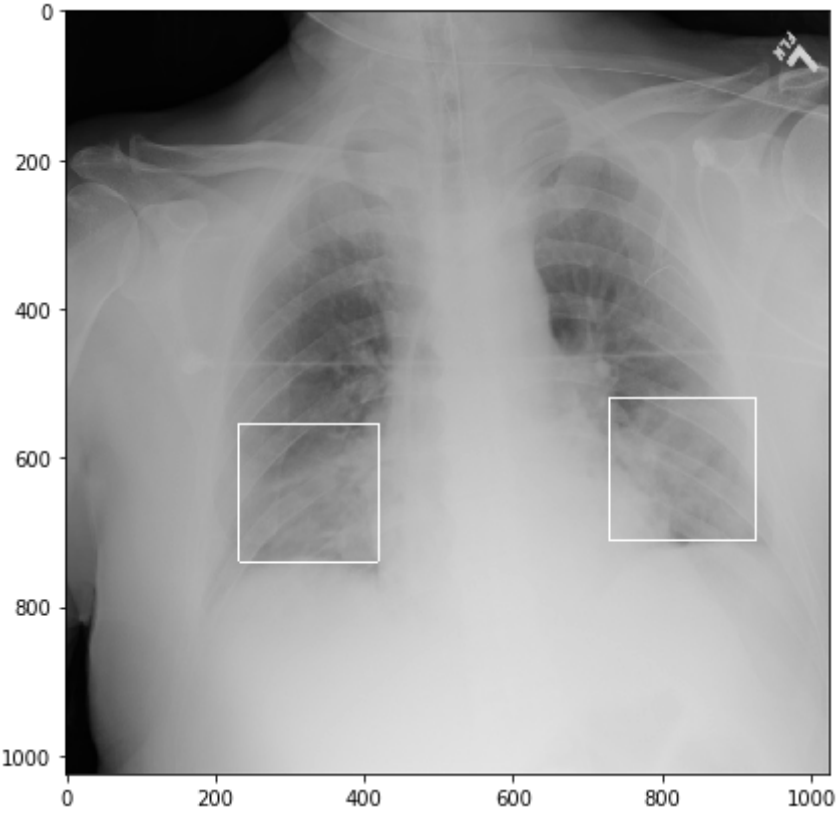


Image enhancement

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.

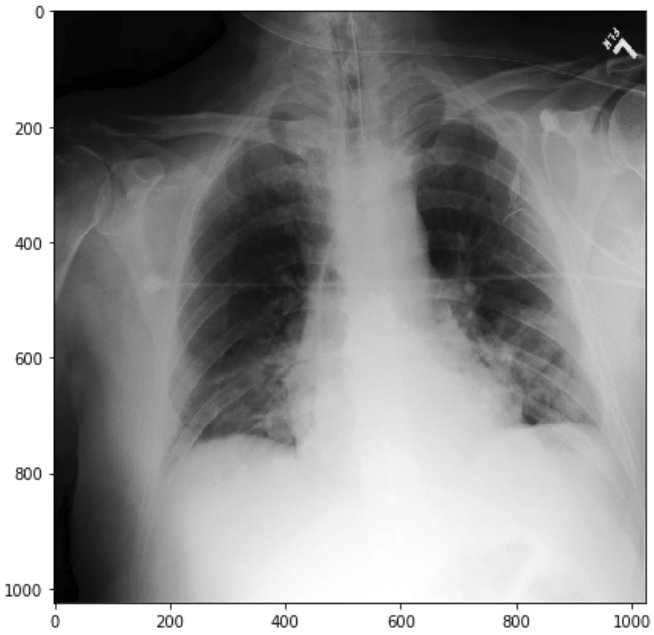


Image sharpening

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 **unsharpen 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 unsharpen mask filter was still a little blurred and not ideal for feature extraction.

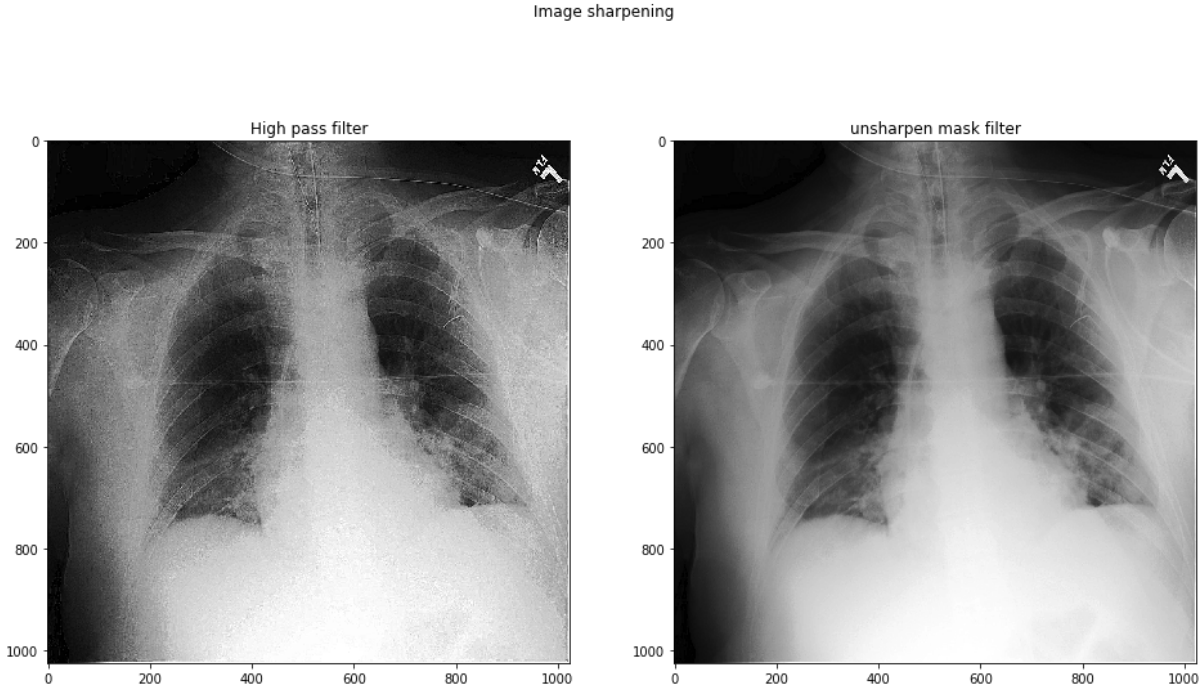


Image segmentation

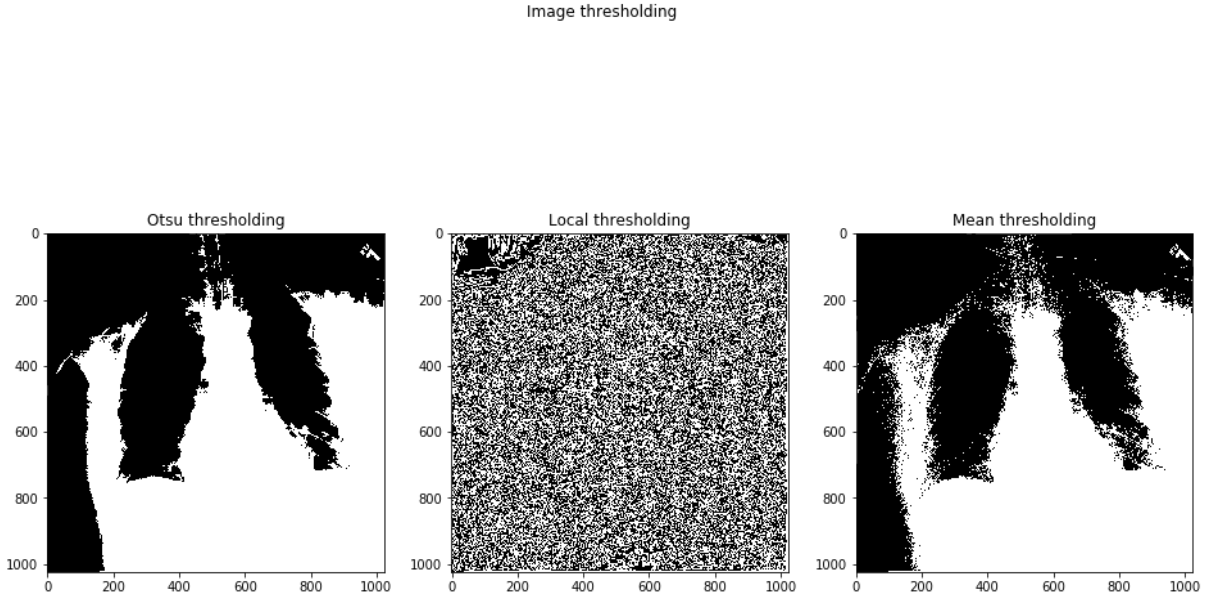
To separate the lungs for analysis we attempted to use three kinds of thresholding methods.

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.

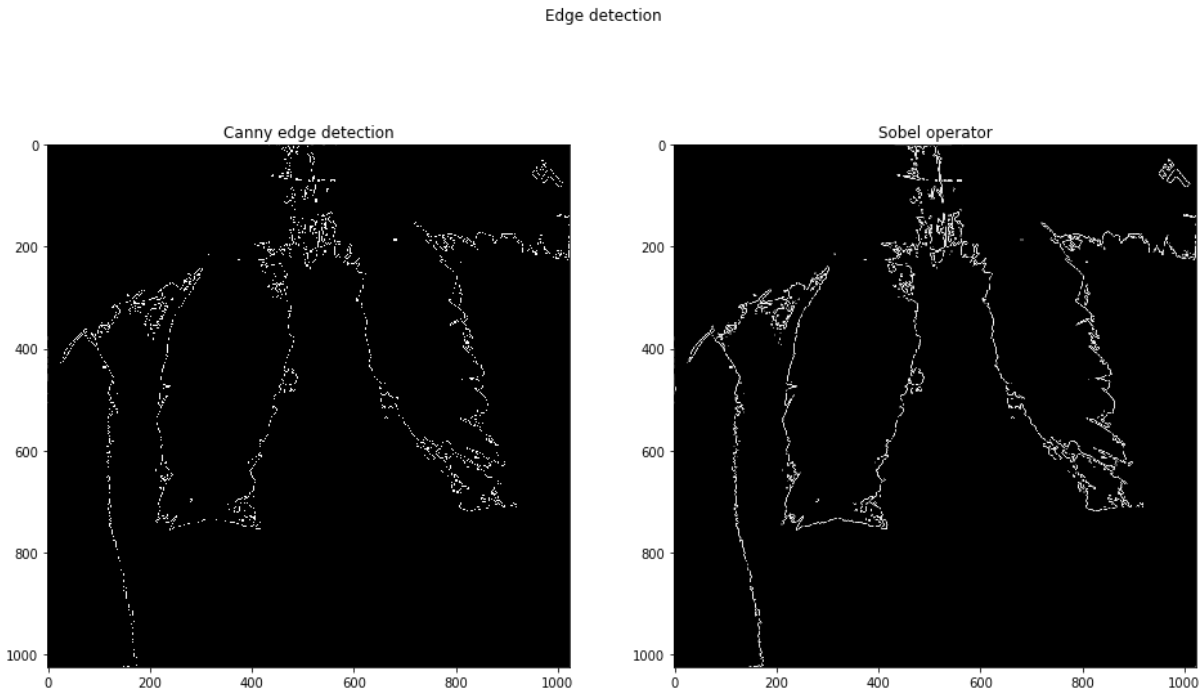
We chose otsu threshold because it extracted smoother edges and segmented the lungs better.



Edge detection

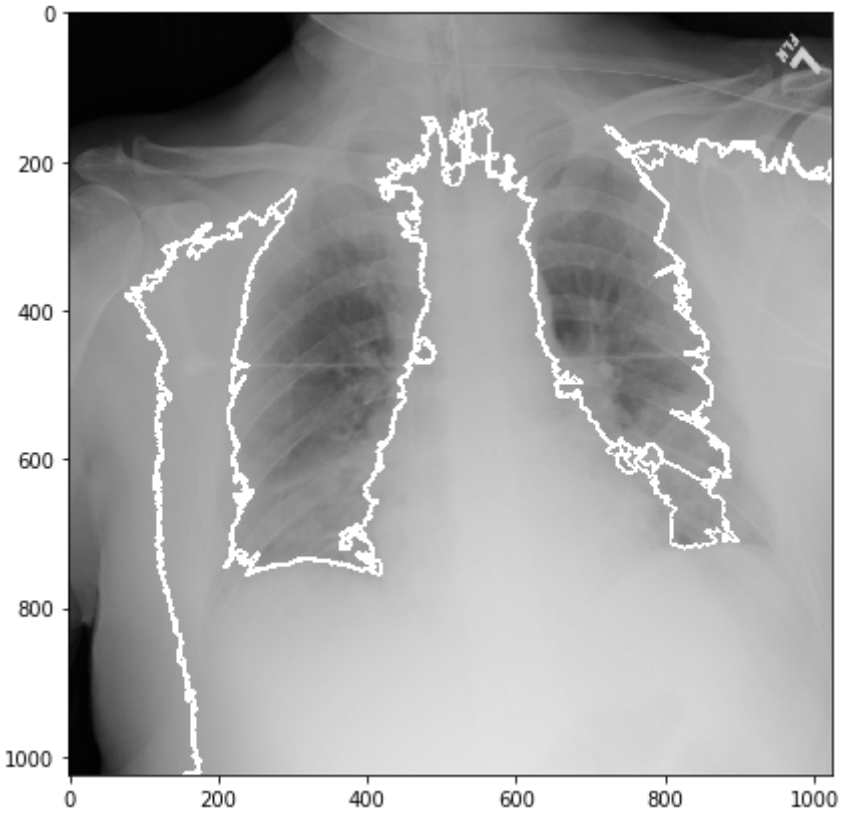
We made use of sobel edge detection and also attempted to use canny edge detection.
Sobel edge detection: It works by calculating the gradient of intensity at each pixel. It finds the direction of the largest increase from light to dark and the rate of change in that direction.

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.



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.



Architecture

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

- Area of opacity
- Perimeter of visible lung regions
- Irregularity index
- Equivalent diameter
- Mean, sd of unenhanced image
- Hu moments

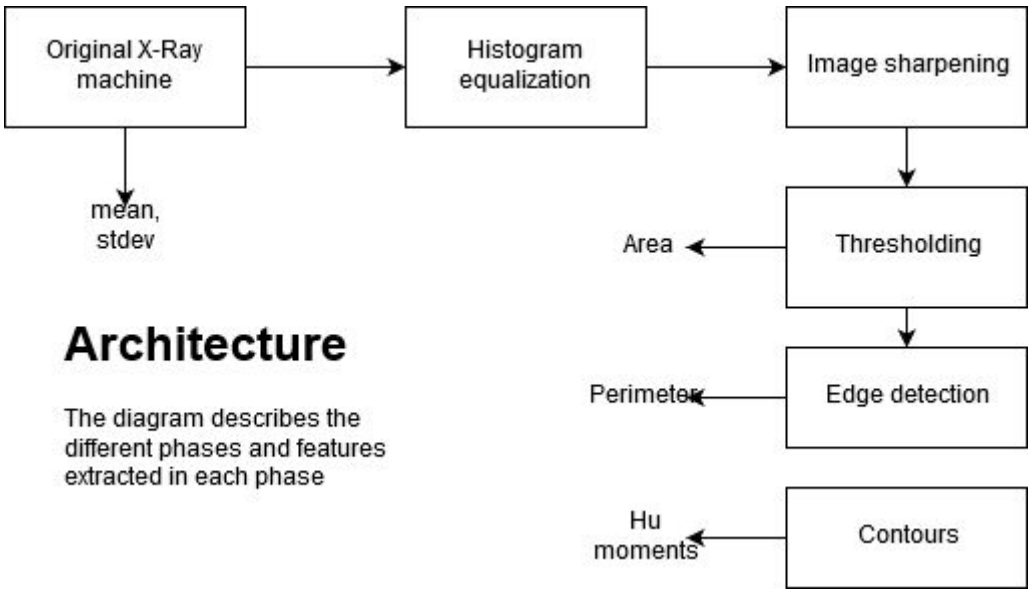


Fig 1. Architecture Diagram

Feature extraction methods

Mean and standard deviation:

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

Area:

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

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.

Irregularity index:

Also called compactness is defined as below:

$(4 * \pi * area)/perimeter^{2}$

Equivalent diameter:

$\sqrt{4 * area/\pi}$

Hu moments:

$$I_1 = \eta_{20} + \eta_{02}$$

$$I_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$I_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$I_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$I_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$I_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$I_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2].$$

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

Analysis and results

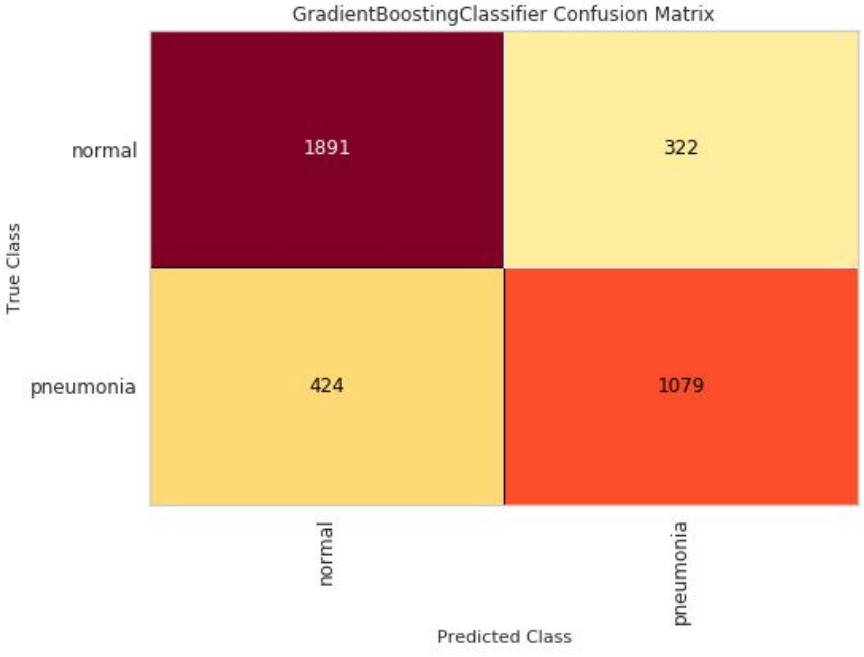
Training Testing ratio: 8:2

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC Score
Logistic regression model	Train: 0.66 Test: 0.66	Train: 0.65 Test: 0.66	Train: 0.34 Test: 0.33	Train: 0.44 Test: 0.44	Train: 0.6 Test: 0.6
Random forest	Train: 0.82 Test: 0.78	Train: 0.81 Test: 0.76	Train: 0.71 Test: 0.67	Train: 0.76 Test: 0.71	Train: 0.8 Test: 0.76
Gradient boosting classifier	Train: 0.87 Test: **0.79**	Train: 0.87 Test: **0.77**	Train: 0.81 Test: **0.71**	Train: 0.84 Test: **0.74**	Train: 0.86 Test: **0.78**
Support Vector Machines	Train: 1.0 Test: 0.59	Train: 1.0 Test: 0.0	Train: 1.0 Test: 0.0	Train: 1.0 Test: 0.0	Train: 1.0 Test: 0.5
KNN	Train: 0.72 Test: 0.7	Train: 0.71 Test: 0.68	Train: 0.55 Test: 0.51	Train: 0.62 Test: 0.58	Train: 0.7 Test: 0.67

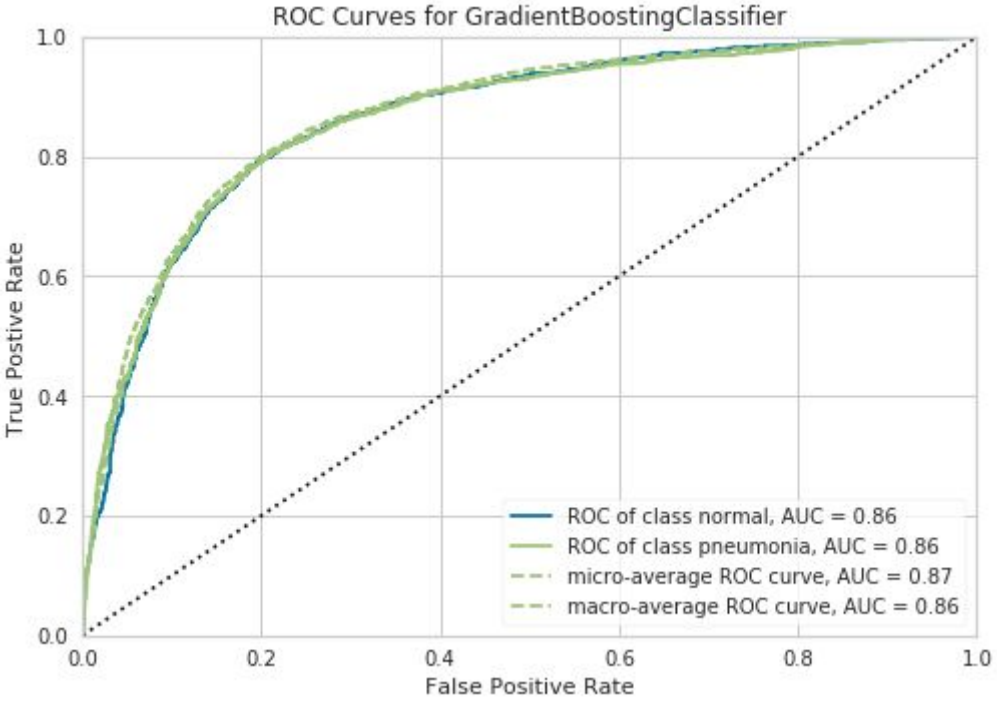
We tried changing the ratios to 7:3 as well as 6:4. There wasn't any significant difference.

Performance of the best model:

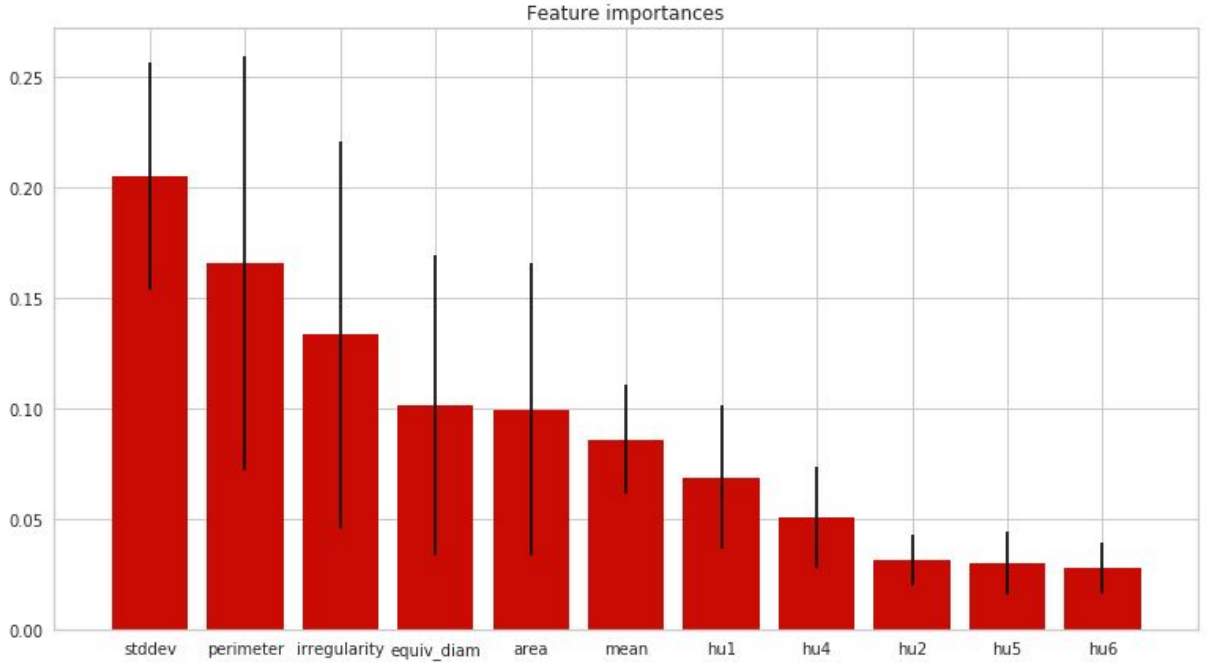
Confusion matrix



Receiver Operator Characteristic

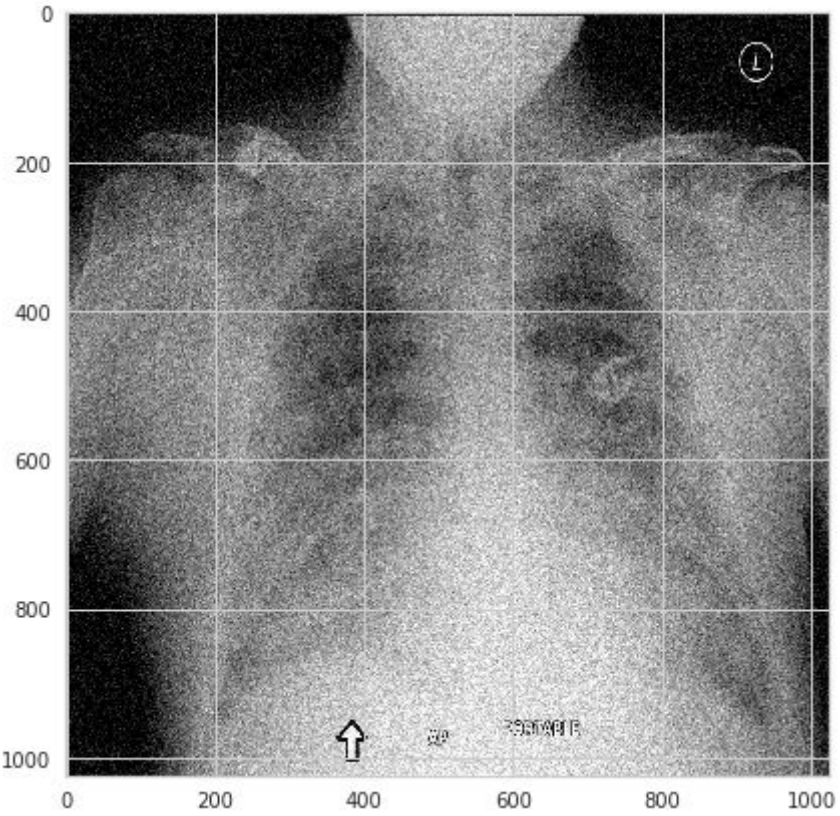


Feature importance



Sensitivity to noise

While most X-ray machines produce denoised image, we tested a few images with gaussian noise and observed the results. We observed that the model was extremely sensitive to noise and produced random results.



Conclusion

We were able to successfully detect pneumonia with impressive results. Our model was able to achieve an roc_auc of 0.78. 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.

Bibliography

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Lungren, M. P. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*.

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.

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.

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

Ayan, E., & Ünver, H. M. (2019, April). Diagnosis of Pneumonia from Chest X-Ray Images Using Deep Learning. In *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)* (pp. 1-5). IEEE.

Barrientos, R., Roman-Gonzalez, A., Barrientos, F., Solis, L., Correa, M., Pajuelo, M., … Zimic, M. (2016). Automatic detection of pneumonia analyzing ultrasound digital images. *2016 IEEE 36th Central American and Panama Convention (CONCAPAN XXXVI)*. doi: 10.1109/concapan.2016.7942375

Vijendran, S., & Dubey, R. (2019). Deep Online Sequential Extreme Learning Machines and its Application in Pneumonia Detection. *2019 8th International Conference on Industrial Technology and Management (ICITM)*. doi: 10.1109/icitm.2019.8710700

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

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. doi: 10.3390/app8101715

Cicero, M., Bilbily, A., Colak, E., Dowdell, T., Gray, B., Perampaladas, K., & Barfett, J. (2017). Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs. *Investigative Radiology*, *52*(5), 281–287. doi: 10.1097/rli.0000000000000341

Khobragade, S., Tiwari, A., Patil, C., & Narke, V. (2016). Automatic detection of major lung diseases using Chest Radiographs and classification by feed-forward artificial neural network. *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*. doi: 10.1109/icpeices.2016.7853683

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*(S2). doi: 10.1186/s12864-018-4467-6

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine*, *15*(11). doi: 10.1371/journal.pmed.1002683

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. doi: 10.1513/annalsats.201612-947oc

Naam, J., Harlan, J., Nurcahyo, G. W., Arlis, S., Sahari, S., Mardison, M., & Rani, L. N. (2017). Detection of Infiltrate on Infant Chest X-Ray. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *15*(4), 1943. doi: 10.12928/telkomnika.v15i4.3163

Gordienko, Y., Gang, P., Hui, J., Zeng, W., Kochura, Y., Alienin, O., … Stirenko, S. (2018). Deep Learning with Lung Segmentation and Bone Shadow Exclusion Techniques for Chest X-Ray Analysis of Lung Cancer. *Advances in Intelligent Systems and Computing Advances in Computer Science for Engineering and Education*, 638–647. doi: 10.1007/978-3-319-91008-6_63

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

Guerra, M., Crichiutti, G., Pecile, P., Romanello, C., Busolini, E., Valent, F., & Rosolen, A. (2015). Ultrasound detection of pneumonia in febrile children with respiratory distress: a prospective study. *European Journal of Pediatrics*, *175*(2), 163–170. doi: 10.1007/s00431-015-2611-8

Ker, J., Wang, L., Rao, J., & Lim, T. (2017). Deep learning applications in medical image analysis. *Ieee Access*, *6*, 9375-9389.

Sha, Y., & Wang, M. D. (2017). Interpretable Predictions of Clinical Outcomes with An Attention-based Recurrent Neural Network. *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics - ACM-BCB 17*. doi: 10.1145/3107411.3107445

Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S. (2016). Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, *35*(5), 1207–1216. doi: 10.1109/tmi.2016.2535865

Rosenberg, L., Lungren, M., Halabi, S., Willcox, G., Baltaxe, D., & Lyons, M. (2018). Artificial Swarm Intelligence employed to Amplify Diagnostic Accuracy in Radiology. *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. doi: 10.1109/iemcon.2018.8614883

Nichols, J. A., Chan, H. W. H., & Baker, M. A. (2019). Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophysical reviews*, *11*(1), 111-118.

Cisneros-Velarde, P., Correa, M., Mayta, H., Anticona, C., Pajuelo, M., Oberhelman, R., ... & Lavarello, R. (2016, August). Automatic pneumonia detection based on ultrasound video analysis. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4117-4120). IEEE.

Veeramani, S. K., & Muthusamy, E. (2016). Detection of abnormalities in ultrasound lung image using multi-level RVM classification. *The Journal of Maternal-Fetal & Neonatal Medicine*, *29*(11), 1844-1852.

Barrientos, F., Roman-Gonzalez, A., Barrientos, R., Solis, L., Alva, A., Correa, M., ... & Oberhelman, R. (2016, November). Filtering of the skin portion on lung ultrasound digital images to facilitate automatic diagnostics of pneumonia. In *2016 IEEE 36th Central American and Panama Convention (CONCAPAN XXXVI)* (pp. 1-4). IEEE.

Kumar, T. G., Asha, V., Manish, T. I., & Muthulakshmi, G. (2018). Empirical system of image enhancement for digital microscopic pneumonia bacteria images. *Bratislavske lekarske listy*, *119*(8), 522-529.

Dietert, K., Nouailles, G., Gutbier, B., Reppe, K., Berger, S., Jiang, X., ... & Witzenrath, M. (2018). Digital Image Analyses on Whole-Lung Slides in Mouse Models of Acute Pneumonia. *American journal of respiratory cell and molecular biology*, *58*(4), 440-448.

Soldati, G., & Demi, M. (2017). The use of lung ultrasound images for the differential diagnosis of pulmonary and cardiac interstitial pathology. *Journal of ultrasound*, *20*(2), 91-96.

Zenteno, O., Castañeda, B., & Lavarello, R. (2016, August). Spectral-based pneumonia detection tool using ultrasound data from pediatric populations. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4129-4132). IEEE.

https://en.wikipedia.org/wiki/Image_moment

https://scikit-image.org/docs/stable

https://en.wikipedia.org/wiki/Invariant_(mathematics)

https://www.kaggle.com/suryathiru/dip-different-approches

https://www.kaggle.com/c/rsna-pneumonia-detection-challenge

https://www.kaggle.com/kmader/training-u-net-on-tb-images-to-segment-lungs