PNEUMONIA DETECTION

A PROJECT REPORT

SUBMITTED BY

SREEKUTTY SHAJI

to

The APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree

of

Master of Computer Applications

Department of Computer Applications

St. Joseph's College of Engineering

And Technology Choondacherry, Palai

MAY 2020

DEPARTMENT OF COMPUTER APPLICATIONS

St. Joseph's College of Engineering And Technology, Palai

CERTIFICATE

This is to certify that the report entitled **"PNEUMONIA DETECTION"** submitted by **Sreekutty Shaji** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Master of Computer Application in (MCA) is a bonafide record of the project work carried out by him/her under my/our guidance and supervision.. This report in any form has not been submitted to any other University or Institute for any purpose.

Internal Supervisor(s)

Mr. Jose George

PG Coordinator

Mr. Alex Jose

External Supervisor(s)

Head of the Department

T.D. Jainendrakumar

DECLARATION

I undersigned hereby declare that the project report "**PNEUMONIA DETECTION**" , submitted for partial fulfillment of the requirements for the award of degree of Master of Computer Application of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of **Mr. Jose George**. This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Place: Palai **Signature**

Date **Sreekutty Shaji**

ACKNOWLEDGEMENT

Submitting our project in the divine feet of Almighty God we would like to take this opportunity to express our profound gratitude to all the people who have inspired and motivated us to make this project a success. Words are boundless to express our sincere thanks to our most respected Principal **Dr. J. David.**

We are deeply indebted to **Dr.T.D Jainendrakumar** (Head of the Department Computer Application, St.Joseph's college of engineering and technology, Palai) and **Mr.Jose George** (Internal Guide) for their proper guidance, encouragement and timely suggestions throughout our project work.

Last but not the least we extend our sincere thanks to our dear parents and to our friends who gave us encouragement and self confidence whenever we were fed with heavy schedule.

Sincerely
Sreekutty Shaji

ABSTRACT

Pneumonia accounts for a significant proportion of patient morbidity and mortality.Pneumonia is an infection that influences the air sacs in one or both lungs.Early diagnosis and treatment of pneumonia is critical to preventing complica-tions including death.Chest X-rays are the most common imaging examination tool used in practice,critical for screening,diagnosis and management of a variety of diseases including pneumonia.

However,two thirds of the global population lacks access to radiology diag-nostics,according to an estimate by World Health Organization.Detecting pneumonia in chest radiography can be difficult for radiologists.The appearance of pneumonia in X-ray images is often vague,can overlap with other diagnosis and can mimic many other benign abnormalities.These discrepancies cause considerable variability among radiologists in the diagnosis of pneumonia.

In this project,builds a method that can automatically detect pneumonia from Chest X-ray.For that,here used advanced Machine Learning techniques.By using CNN(Convolutional Neural Network)the features in the image can be automatically extracted.Basically in this project the user can input a chest X-ray and the output is that the input have pneumonia or not.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Need for the Project

During the past decade,technological innovations have progressed such an accelerated pace that they have permeated almost every facet of life..This is especially true in the fields of medicine and health care services.In the current era of modern industrial world,more emphasis is given for the implementation of modern technology and advanced research in every fields. Defined as the use of the principles and techniques of engineering to detect pneumonia from X-ray images.It is to improving accuracy in prediction of disease with less effort.

1.2 Outline of the Project

A general discussion about Existing System,proposed system,algorithm used,system specification and software tools are provided in chapter 2. Module description,working flowchart and UML diagrams are described in chapter 3. Chapter 4 describes the user interface design including screenshots. Chapter 5 presents the usefulness of testing and it includes the functional testing.System implementation which includes the implementation methods and implementation plan are described in chapter6. Conclusion and future scope of the project is summarized in chapter7.

1.3 Motivation

Motivation for this project is through automation any work that will reduce the human work load and also increase the efficiency and accuracy of work. With computer-aided diagnosis,physicians can make chest X-ray diagnosis more quickly and accurately.

1.4 Scope of the Project

Automated detection of diseases from chest X-rays at the level of expert radiologists would not only have tremendous benefit in clinical settings,it would also be invaluable in delivery of health care to populations with inadequate access to diagnostic imaging specialists.In this project,hope to train a model using the dataset to help physicians in making diagnoses of pneumonia in chest X-rays.

Chapter 2

REQUIREMENT ANALYSIS

2.1 System Study

2.1.1 Existing System

Pneumonia is the common type of infection found in the world.More than 1 million adults are hospitalized with pneumonia and around 50,000 die from the disease every year in the US alone(CDC,2017).The infection spreads in the lungs area of a human body.The chest X-ray is performed to diagnose the infection.Physicians use this X-ray image to diagnose or monitor treatment for condition of pneumonia.

2.1.1.1 Drawbacks

- Radiologists have to spend time diagnosing these chest X-ray images to find any potential lung diseases.
- Detecting pneumonia in chest radiography can be difficult for radiologists.
- The appearance of pneumonia in X-ray images is often vague,can overlap with other diagnoses,and can mimic many other benign abnormalities.
- There is a shortage of experts who can interpret X-rays,even when imaging equipment is available,leading to increased mortality from treatable diseases.

2.1.2 Proposed System

With computer-aided diagnosis,physicians can make chest X-ray diagnoses more quickly and accurately.Developing automated system could make huge impact to the patients,who don’t have access to expert radiologists. Pneumonia Detection Using Deep Learning help to detect the pneumonia automatically from chest X-rays.Here, in this proposed system user can input a chest X-ray and the output is that the input have that particular disease or not.Increasing the number of images and number of layers of CNN can significantly increase the accuracy of the model.

2.1.2.1 CNN

In deep learning,a convolutional neural network(CNN,or ConvNet) is a class of deep neural networks,most commonly applied to analyzing visual imagery.It is a feed forward neural network.They are heavily influenced by how we - humans see the surrounding world.CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing.They also known as shift invariant or space invariant artificial neural networks(SIANN),based on their shared-weights architecture and translation invariance characteristics.

2.1.2.2 Components of CNN

CNN has two components :The Hidden layers/Feature extraction part and The Classification part.

1.The Hidden layers/Feature extraction part

In this part, the network will perform a series of convolutions and pooling operations during which the features are detected.

2.The Classification part

Here,the fully connected layers will serve as a classifier on top of these extracted 4 features.They will assign a probability for the object on the image being what the algorithm predicts it is.

2.1.2.3 Layers in CNN

It has basically 4 layers.

- Convolution Layer

  It is the central aspect of processing images in the Convolutional Neural Network.It has a number of filters that perform convolution operation.One image became a stack of filtered images.Ultimate purpose of this layer is to receive a feature map.

- Activation layer

  Once the feature maps are extracted,the next step is to move then to ReLu(Rectified Linear Unit) layer.We will use ReLu activation function that returns ’0’ for every negative value in the input image while it returns the same value for every positive value.

- Pooling Layer

  Pooling layer uses different filters to identify different parts of image like edges,corners etc.It shrinks the image stack:

  ➔ Pick a window size

  ➔ Pick a stride

  ➔ Walk your windows across your filtered images

  ➔ From each window take maximum value

  ➔ Known as max pooling

  Flattening is the process of converting all the resultant 2D arrays from pooled feature map into single long continuous linear vector.

- Fully Connected Layer

  Flattened matrix from the pooling layer is fed as input to the fully connected layer to classify image.The output is gonna come out from this layer.

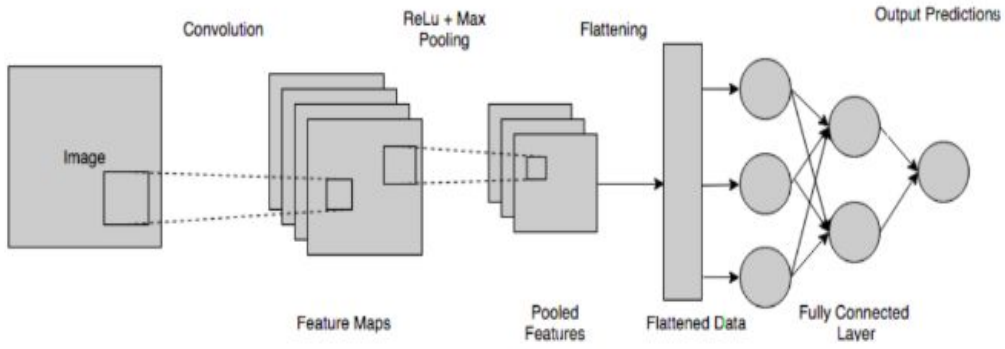


Fig. 2.1. Convolutional Neural Network

2.2 System Specification

The system specification is the work product produced by the system and requirements engineer.It describes the functions and performance of a computer based system and constraints that will govern its development.In requirements specifications, a detailed and precise description of a system is set out to act as a basis for a contract and software developer,so the user requirements are properly organize and documented in the software requirement specification.

2.2.1 Specification for Development

2.2.1.1 Hardware Specification

- Processor:Intel Core i5
- RAM:64GB
- Hard Disk:1TB or higher
- Display:14.1 Colour Monitor(LCD,CRT or LED)
- Peripherals:Normal QWERTY with 104 keys,Mouse

2.2.1.2 Software Specification

- Operating System :windows 10
- Front-End Scripting language:Python
- IDE: Pycharm
- Web Server:Apache Tomcat Server
- Web Browser:Google Chrome or other web browsers
- Framework:Flask
- Packages :Tensorflow,Keras,pillow,numpy,matplotlib

2.2.2 Specification for Implementation

2.2.2.1 Hardware Specification

- Processor:Intel Pentium Iv and above
- RAM:2GB
- Hard Disk:40GB or higher
- Display:14.1 Colour Monitor(LCD,CRT or LED)
- Peripherals:Normal QWERTY with 104 keys,Mouse

2.2.2.2 Software Specification

- Operating System :Windows 10
- Front-End Scripting language:Python
- IDE: Pycharm
- Web Server:Apache Tomcat Server
- Web Browser:Google Chrome or other web browsers
- Framework:Flask

2.3 Software Tools

2.3.1 Python

Python is an interpreted,high-level general purpose programming language.Created by Guido van Rossum and first released in 1991.Python has a design philosophy that emphasizes code readability,notably using significant white space.

Python 3.6.0 is the newest major release of the python language and it contains many new features and optimization.It adds a couple of new features and improvements that will affect the day today work of python coders.

it's new features include:

- Improved numeric literals
- String interpolation
- Type annotations for variables

2.3.2 Pycharm

It is an integrated development environment(IDE) used in computer programming , specifically for the python language. Pycharm is cross-platform,with windows,macos and Linux versions.It is used for development in python and frameworks like DJango.You can customize it with themes and plugins.It lets you to enhance productivity while coding b providing some features like suggestions,Local VCS etc.

2.3.3 TensorFlow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks.It is a symbolic math library,and is also used for machine learning applications such as neural networks.It is used for both research and production at Google.Tensorflow was developed by the Google Brain team for internal Google use.It was released under the Apache 2.0 open-source license on November9,2015.

2.3.4 Keras

Keras is an open-source neural-network library written in python.It is capable of running on top of Tensorflow,Microsoft Cognitive Toolkit,Theano,or PlaidML.Designed to enable fast experimentation with deep neural networks,it focuses on being user-friendly,modular and extensible.It was developed as part of the research effort of project ONEIROS,and its primary author and maintainer is Francis Chollet,a Google engineer.In 2017,Google's Tensorflow team decided to support Keras in TensorFlow's core library.Chollet explained that Keras was conceived to be an interface rather than a standalone machine-learning framework.It offers a higher-level,more intuitive set of abstractions that make it easy to develop deep learning models regardless of the computational backend used.

2.3.5 Flask

Flask is a micro web framework written in python.It is classified as a microframework because it does not require particular tools or libraries.It has no database abstraction layer,form validation,or any other components where pre-existing third-party libraries provide common functions.However,Flask supports extensions that can add application features as if they were implemented in Flask itself.Extensions exist for object-relational mappers,form validation,upload handling,various open authentication technologies and several common framework related tools.Extensions are updated far more regularly than the core Flask program.

Chapter 3

SYSTEM MONITORING

System modelling is the interdisciplinary study of the model to conceptualize and construct systems in business and IT development. A common type of systems modeling is function modeling,with specific techniques such as the UML diagrams.These models can be extended using functional decomposition,and can be linked to requirements models for further systems partition.

3.1 Introduction

3.2 Module Description

The main part of this project can be divided into 3 modules.

- Feature Extraction
- Training and Testing
- Prediction

3.2.1 Feature Extraction

CNN is used to extract feature of the data.It includes different layers that filters and extract each features.Fully Connected layers will serve as a classifier on top of these extracted features.The number of layers in the CNN make a big role in feature extraction.

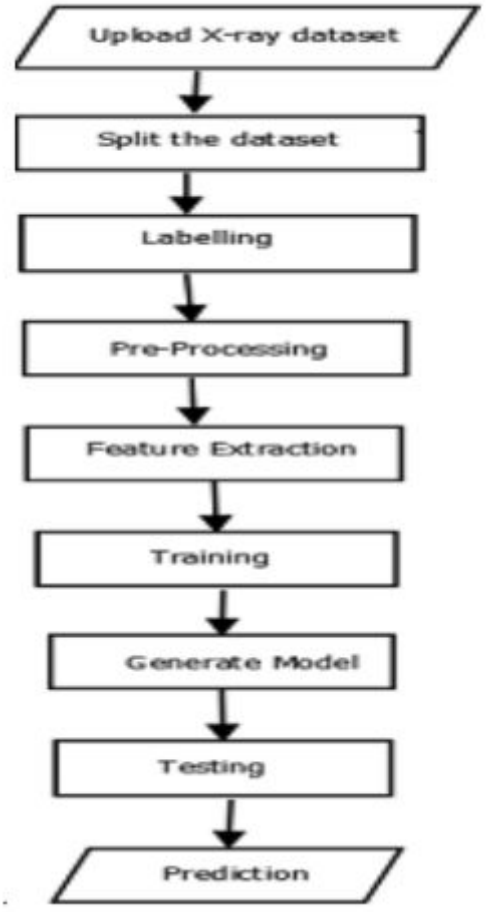
3.2.2 Training and Testing

From the dataset 80% of data is used for training purpose.And the remaining 20% for testing.After training,a model will be created.Once a model is created then we can start testing.

3.2.3 Prediction

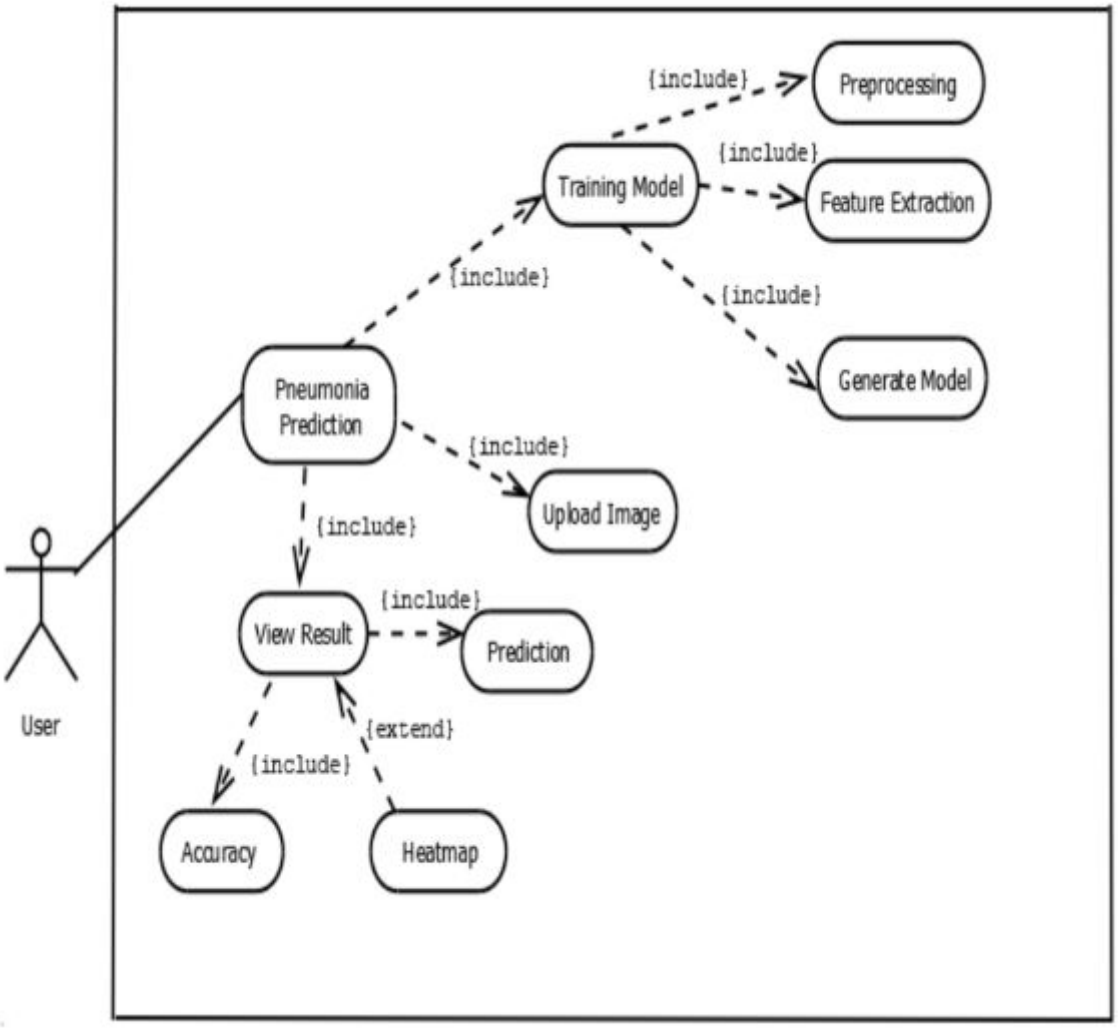
Pneumonia detection task is a binary classification problem.Where the input is a frontal view chest X-ray image x and output is a binary label y.Value of y be either 0 or 1,indicating the absence or presence of pneumonia respectively.

3.3 Flowchart

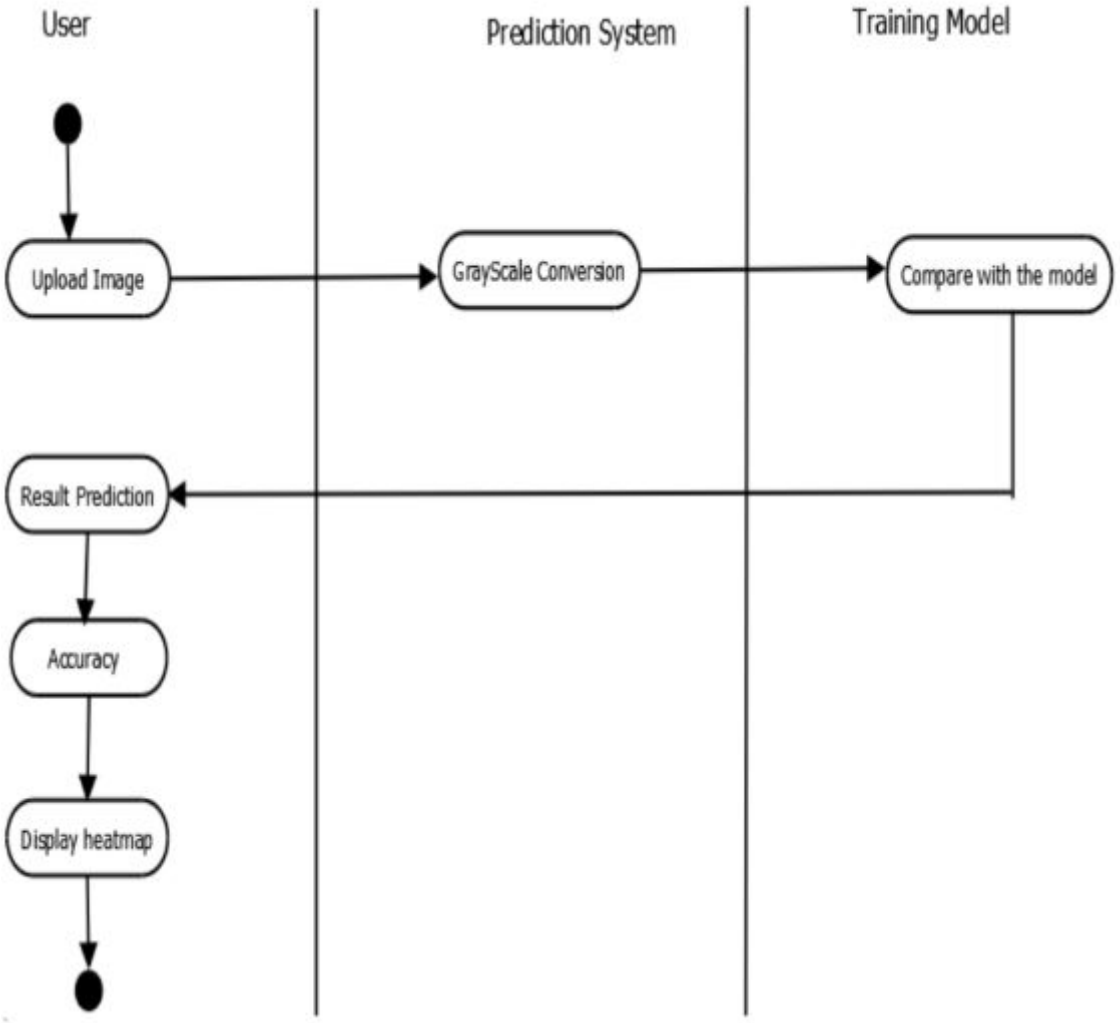


3.4 UML Diagram

3.4.1 Use Case Diagram



3.4.2 Activity Diagram



Chapter 4

System Design

4.1 Introduction

system design is the process of defining the architecture,modules,interfaces and data for a system to satisfy specified requirements.System design could be seen as the application of systems theory to product development.

4.2 User Interface

4.2.1 Screenshots

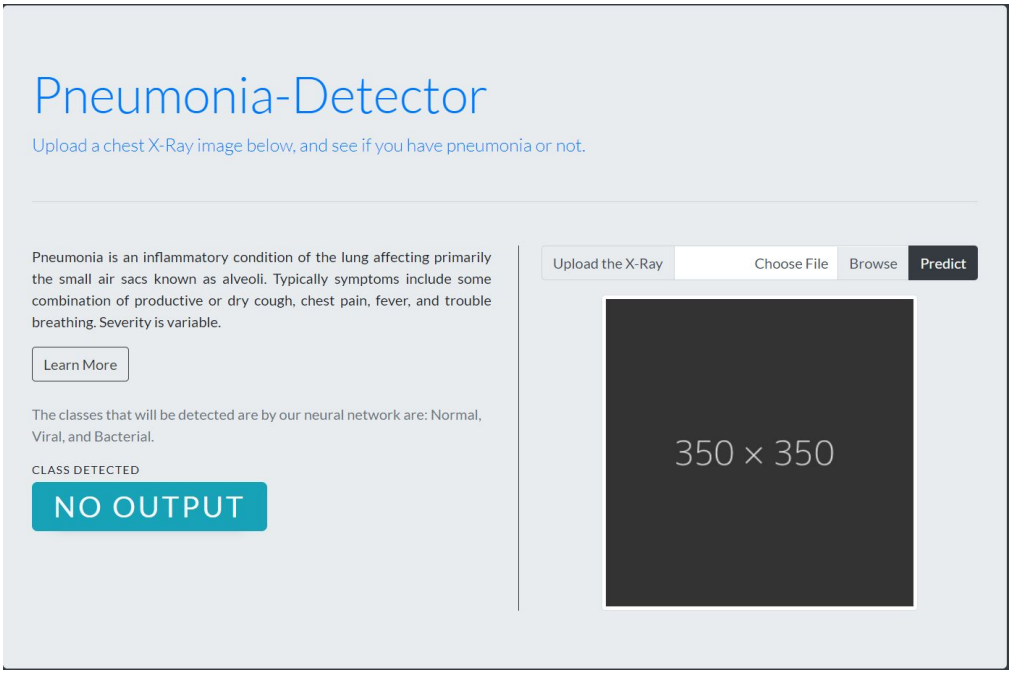


Fig 4.1 HomePage

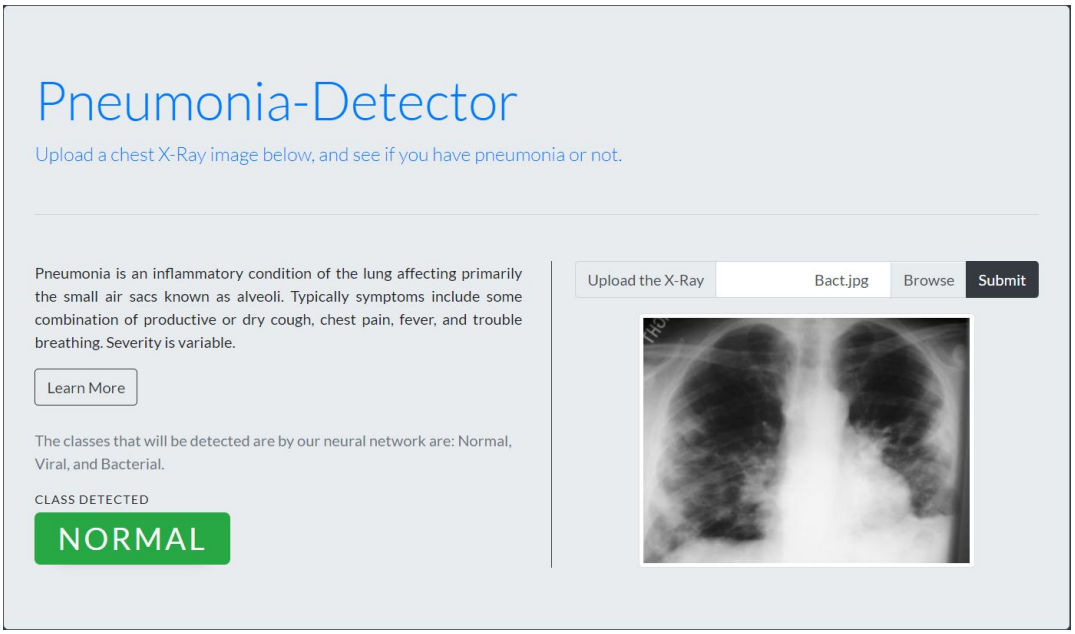


Fig 4.2 Normal

Chapter 5

SYSTEM TESTING

5.1 Introduction

System testing is the stage of implementation,which is aimed at ensuring that the system works accurately and efficiently before live operation commences.Testing is the process of executing the program with the intend of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied.The ultimate aim is quality assurance.Tests are carried out and the results are compared with the expected document.In the case of erroneous results, debugging is done.

5.2 Functional Testing

Functional Testing is a type of software testing whereby the system is tested against the functional requirements/specifications.

Functions(or features) are tested by feeding them input and examining the output.Functional testing ensures that the requirements are properly satisfied by the application.

Typically, functional testing involves the following steps:

- Identify functions that the software is expected to perform.
- Create input data based on the function’s specifications.
- Determine the output based on the function’s specifications.
- Execute the test case.
- Compare the actual and expected outputs.

Test case ID	Test Objective	Precondition	Steps:	Test data	Expected result	Post-condition	Actual Result	Status(pass/Fail)
TC_01	Training	1. Upload a set of X-ray images 2. Define the numer of epochs	1. Upload large dataset of X-ray images 2.Run the Train.py file	Set of X-ray images	Training completed and shows an accuracy and loss graph	Generate a model	Error "Resource Exhausted Error:OOM when allocatingtensor with shape[401408,512]& type float 0ttask:0device:CPU:0by allocator CPU	Failed
							Shows the accuracy and loss graph	Success
TC_02	GrayScale Conversion	1. Images should be of black and white 2. Figures should be in a particular size	1. Upload the image	Nolabel.png	Prediction	Feature Extraction	Error "Value Error:Error When checking input:	Failed
							Predicted	Success
TC_03	Plotting the bouding_box values on the images	1. Create .csv file for uploaded images	1.Upload the images 2.Run the code linesplitnew.py	Set of X-ray images	Show the bounding box by a rectangle	Generating .h5 model after completing the training session	Error "Value Error:invalid literal for int() with base10	Failed
							Plotted the particular area	Success
TC_04	Plotting the bouding_box values on the images	1. Create .csv file for uploaded images	1.Upload the images 2.Run the code linesplitnew.py	Set of X-ray images	Show the bounding box by a rectangle	Complete the training and generate the model	Error "CV2.error:Opencv350	Failed
							Plotted boundingbox on image	Success

Fig 5.1 Functional Testing

Chapter 6

SYSTEM IMPLEMENTATION

Implementation is the stage in the project where the theoretical design into a working system and is giving the confidence on the new users that it will work effectively and efficiently.

6.1 Implementation Method

For the implementation of the project collected 42GB of NIH X-ray dataset from Kaggle.com.For the purpose of training and testing split the dataset.Hence for the actual dataset it took nearly 12GB RAM,and two days of training.After the training a model will be created with a good accuracy,as well as 2 graphs that shows the accuracy and loss and also predict whether affected by pneumonia or not.Inorder to demonstrate a model,we collected a sample of images and train it,then test it.As the number of epochs increases accuracy will also increase.We obtain the 2 graphs along while training.Then plot the affected area by bounding box.csv which also used in drawing heatmap.Heatmap is a coloured visual representation of data to visualize the areas of the image most indicative of the disease using Class Activation Mappings.

6.2 Implementation Plan

- Initially used simple neural networks to demonstrate and study the basic working of CNN.For that used dog vs cat image classification with deep learning.
- Collect small dataset which contain x-Ray images.With that train and obtain a model,which predicts successfully and demonstrates the working easily along with a graph of loss and accuracy.
- Use the original dataset of x-Ray images.Train it in a supercomputer and obtain a model using the model it can predict more accurately and using the heatmap can visualize the affected area successfully.

Chapter 7

CONCLUSION AND FUTURE SCOPE

Pneumonia is lung inflammation caused by infection with virus,bacteria,fungi or other pathogens.According to National Institute of Health(NIH),chest x-ray is the best test for pneumonia diagnosis.However,reading X-ray images can be tricky and requires domain expertise and experience.Pneumonia detection using deep learning will help to detect the pneumonia without an expert radiologists with an accuracy of 65 %.

Automated detection of diseases from chest X-rays at the level of expert radiologists would not only have tremendous benefit in clinical settings,it would also be in valuable in delivery of health care to populations with inadequate access to diagnostic imaging specialists.And my project can be easily extends to find 14 thoracic diseases in the dataset.

Chapter 8

REFERENCES

1) Mohd Nizam Saad, Noraidah Sahari Ashaari, Hamzaini Abdul Hamid, Image Segmentation for Lung Region in Chest X-ray Images using Edge Detection and Morphology, 2014 IEEE International Conference on Control System Computing and Engineering, 28 to 30 November 2014.
2) Detecting Thoracic Diseases from Chest X-ray images by Binit Topiwala,Mariam Alawadi,Hari Prasad Stanford University
3) Er O., Yumusak N., Temurtas F. Chest diseases diagnosis using artificial neural networks. Expert Systems with Applications. 2010
4) Pranav Rajpurkar, Jeremy Irvin, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X- rays with Deep Learning, https://arxiv.org/pdf/1711.05225.pdf 2017.
5) Gao Huang, Zhuang Liu, Laurens van der Maaten. Densely Connected Convolutional Neural Networks https://arxiv.org/abs/1608.06993