SMART ASSISTANT TO EASE THE PROCESS OF COVID-19 AND PNEUMONIA DISEASE DETECTION

Project ID: 2021-180

Group Project Proposal Report

B.Sc. (Hons) Degree in Information Technology

(Specialization in Information Technology)

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

February 2021

SMART ASSISTANT TO EASE THE PROCESS OF COVID-19 AND PNEUMONIA DISEASE DETECTION

Project ID: 2021-180

Group Project Proposal Report

Supervisor: Ms. Bhagyanie Chathurika

B.Sc. (Hons) Degree in Information Technology

(Specialization in Information Technology)

Department of Information Technology

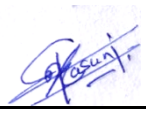
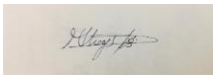
Sri Lanka Institute of Information Technology

Sri Lanka

February 2021

DECLARATION

I declare that this is my own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
Senevirathne K.D. A	IT18115130	
Akalanka B. A	IT18114836	
Nimantha W.A.R	IT18125962	
Dias M.H.V	IT18121698	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor

……

Mrs. Bhagyanie Chathurika

Date

……

ABSTRACT

Covid-19 is a spreading viral disease that infects humans rapidly. Human beings' daily lives, their health, their economy, and the country's economy are affected due to this deadly virus. In the last year, society has observed how the COVID-19 pandemic is frequently affecting many people very fast worldwide. It's a common spreading virus until now. Researchers found Covid-19 infected patients have shown that they are mostly affected by lung diseases. We realize that patients have to face different types of difficulties among all they can't get their report immediately and have to wait more than 12 hours to get the result. To solve this problem, we proposed a computerized exposure of pneumonia and Covid-19 from using a chest X-ray images. Anyone can capture or upload a chest x-ray image from his/her mobile and get the result easily. Diagnosis of Covid-19 and pneumonia using chest X-rays is the less time-consuming and cost-effective method. Compared to the other testing methods Chest X-ray is the best method anyone can easily access. Even though researchers tried out several machine learning methods to build, the Convolutional Neural Network had the best accuracy It is a deep neural network that needs a lot of data. We have intended a combined method using Deep Convolutional Neural Network and image processing to detect Covid-19 and Pneumonia using CXR images in this work.

Keywords: Pneumonia, Convolutional Neural Network, Chest X-ray (CXR)

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

At present, it has become challenging for people to survive this deadly virus. Compared to the developed countries, they have enough resources and facilities. But in developing countries, public awareness of the importance of health is minimal. In this global pandemic, with the growing numbers of patient's doctors cannot handle this easily. But there should be an emerging need to detect Covid-19 patients rapidly. Mostly pneumonia is detect using Chest x-rays. Way of diagnosing pneumonia using chest X-ray images spends time with a specialist.

We used some non-pharmacological strategies like quarantine, social distancing, and personal protective equipment to prevent this deadly coronavirus. But if we have to face to PCR test, we have to wait more than 12 hours to get the report. When it comes to a chest diagnosis like pneumonia, it is a disease that needs to be identified in an early stage. Therefore, it is essential to a make timely diagnosis of chest diseases and take immediate steps to prevent the disease.

With the growing numbers of patients doctors should more work. Because of that there should be a computer system support image classification is needed. Our proposed system will focus on implementing an application to detect pneumonia and covid-19 patients very quickly and accurately. At the first level, we are supposed to detect pneumonia, covid-19 clouds in chest X-rays (CXR) by applying Image processing techniques. Then the patient will be asked about the symptoms that he/she has experienced using a question answering system. After that, the patient will be asked and get the real-time health condition, health history and bad habits question answering system using leveraging Natural Language Processing. We suppose to maximize the accuracy level with the function to go for the CT scan screening procedure.

1.1 Background

When considering the previous work done on this project, several researchers based on covid-19 and pneumonia identification using machine learning. Many of them have used different types of machine learning algorithms and give good results for which algorithms are used.

When looking at the previous work, they tried several methods to identify pneumonia using chest x-ray images. Some use handcrafted characteristic extraction methods and machine learning algorithms to classify chest X-ray, while others use deep learning methods for feature extraction and classification.

When considering previous work CNN was the most popular method for image classification for this domain. The researchers apply mostly handcrafted methods and deep learning methods for medical Imaging to analyze and clarify various diseases like breast cancer [4], tuberculosis, skin cancer etc. They suggest a deep convolution neural network (CNN) that extracts features from chest X-ray image.

Literature survey on methods of image preprocessing

Looking back to the previous work they mainly extract the chest X-ray image first and preprocess the image to decrease the additional area [5]. They used histogram equalization and thresholding methods to detect Pneumonia cloud in Chest X-ray images. In their work, they cut the exact image by a certain percentage.

In one previous work, they coupled various methods to reduce inappropriate data of Chest X-ray images. They applied a novel preprocessing method named Segmentation-based cropping. The process of turning raw data into high-quality data includes all preprocessing techniques [6].

one paper has improved natural algorithms for cropping chest X-ray images to extract the belly and lung boundary classification. In their work, the original data set were 2048 × 2048 pixel in size. They redesigned it to several values and analyzed the computational passing through the bottom line. They flowed vertically up to 2 pixels to the next pixel, and they continued the horizontal scanning until they found groups whose intensity range varied by 100.[3]

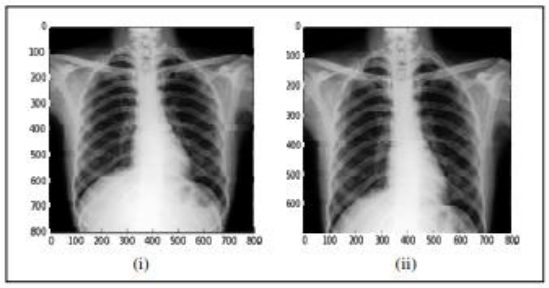


Figure 1.1- image before and after cropping lower abdomen region.

Few examples of pneumonia infected and normal person CXR images from their dataset.

Figure 1.2- Normal CXR.

Figure 1.2- Pneumonia infected CXR.

In this research Agnieszka Mikołajczyk [8] have used different kinds of augmentation methods to overcome the shortage of adequate training data for learning algorithms.

[9] They have shown that neural networks require a considerable amount of data to get a better result. They have used classical image transformation like rotating, cropping, zooming, and other style transfer methods. But they have realized that simple classical operations for image augmentation are not enough. Therefore, they have tried different kinds of new techniques to change the visual features of the images. In this method, their used images were 2048×2048 pixels size. Their example of results is shown below in Table 1.

Table 1.1 Comparison of data using pixel sizes.

Size	Computational time(s)	Result observed
2048x2048	71	Good
1024x1024	9	Good
800x800	4.6	Good
500x500	3.2	Not satisfactory

In their work, they randomly chosen the five Chest X-ray images from their dataset to retrieve the table above, switching their image size and performing their procedure on 5 chest x-ray images of the exact chest. Later they have recorded the threshold models and ratio achieved in every case. The results are acceptable if standard CXR images and pneumonia affected CXR images are adequately modified. They used a picture size of 800 × 800 for their operations and they resized all the C X-ray images into 800 × 800. Their work they applied Histogram Equalization to fix the image sharpness so as improve contrast. This method they have shown that Histogram Equalization enhance the contrast of the image.

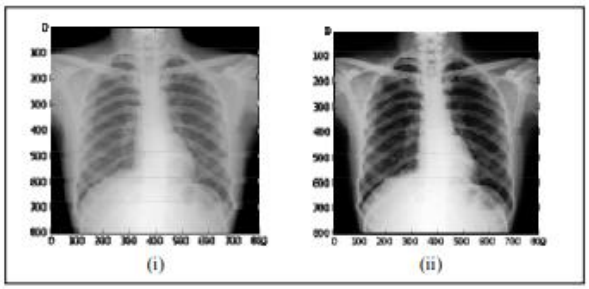


Figure 1.4 – Image before and after histogram equalization

One paper has first shown contrast limited adaptive histogram equalization on CXR image from the Mendeley dataset they have cut 6% vertically on both sides of the picture. After that, they got the two most massive shapes and their leftmost and rightmost points. They have cropped those points to overcome the unnecessary data. Their technique is Histogram equalization used to adjust image intensity to enhance the contrast of the image [7].

Figure 1.5 – The two Largest contours cut along the vertical axis.

In their work, they have filled the preprocessed image into VGG-16 and VGG-19 architectures. They have used five accuracy metrics based to evaluate their result. The five-accuracy matrix are precision, specificity, sensitivity, F1-Score, and accuracy.

Table 1.2- Accuracy levels after using VGG-16 and VGG-19

Accuracy Metric	VGG-16	VGG-19
Precision	0.977	0.971
Sensitivity	0.970	0.974
Specificity	0.939	0.919
F1-score	0.973	0.972
Accuracy	0.962	0.959

Finally, they have found that a deeper compromise network performs better when diagnosing pneumonia, and at the beginning, CXR images should be preprocessed.

Figure 1.6 shows how the chest area is enhanced from the chest CT image using the Otsu's thresholding method.

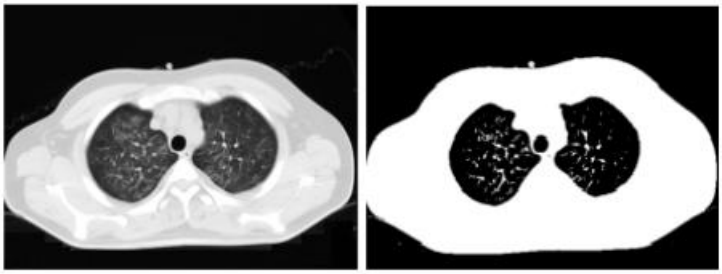


Figure 1.6 Enhancement of lung region by Otsu's thresholding method

In this research [8], they have used the templates offered in cellular neural network (CNN) for early detection of pneumonia symptoms. In addition to developing the algorithm, they found information about the process of understanding the symptoms of pneumonia, characteristics of CT image, CNN and its notations, and other processes such as computer simulation application. The starting position, boundary position, pixel value, field of impact, feedback symmetry, input symmetry, and threshold value are considerations when designing a CNN template.

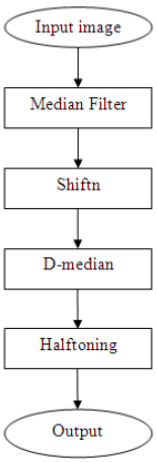


Figure 1.7- Flowchart of the analogic CNN algorithm

1. Input Image – This is the original CT image with pneumonia symptoms. In here it is difficult to identify the area with pneumonia symptoms, and it is difficult to distinguish the affected pneumonia area from the general area.

Figure 1.8- Original CT image with pneumonia symptoms.

2. Median Filter – In this template, impulse noise will be removed from a grayscale image. The pixels are replaced with the median of all the neighboring pixels by the filter.

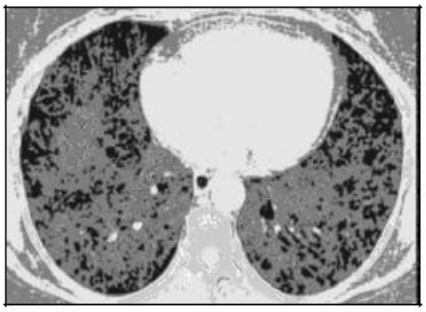


Figure 1.9- Median filter CT image

3. Shiftn Filter - Pixels of the image are shift toward the north direction in this templates. This is where the write contribution to every line is made.

Figure 1.10 - Shiftn filter CT image

4. D-median Filter - In this template, the noise will removed without affecting the sharp edges.

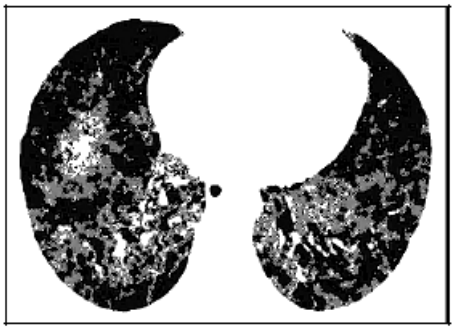


Figure 1.11 - D-median filter CT image

5. Halftoning Filter - The image halftoning for the grayscale image is the function of this halftoning template.

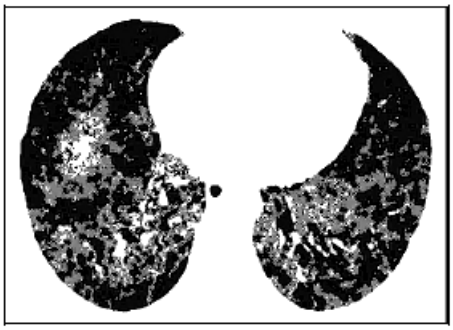


Figure1.12 – Halftoning filter CT image

6. Output Image - After a complete simulation of the designed analogy CNN algorithm successfully obtains the output image of pneumonia symptoms.

Figure1.13 - Output CT image

Here a cellular neural network (CNN) is used to develop an automated method of identifying pneumonia symptoms. What can be understood as the output result is the image segmentation based on the black, grey and white colour. Black indicates the normal area, while gray indicates the pneumonia symptom area and white for pneumonia's critical areas. Each template used has its own function, which are to remove the unwanted elements and filter only the areas with pneumonia from the output image.

With the rapid advancement of computer technology, it has become easier to identify diseases among the common people. Artificial intelligence(AI) has played a major role in the field of computers today. A model can be trained through an algorithm in machine learning, a sub part of artificial intelligence, to obtain accurate results. That model further enhances its accuracy through experience. Therefore, author use machine learning to predict and analysis risk level of covid-19 and pneumonia using symptoms.

In [9] authors done his research using unstructured data. Authors used two algorithms to trained the models for the research. One is KNN(K-Nearest Neighbor) and Other one is CNN (Convolutional neural network). They told CNN model is faster than KNN with 94.8% accuracy.

This [10] authors doing using three algorithms Gaussian Naïve Bayes, Decision Tree and random forest to tuberculosis and pneumonia prediction. Of these three algorithm, the random forest has 97.64% accuracy. It is a highest one of these algorithms.

In [11] the authors used KNN, SVM (Support Vector machine) and Naive Bayes for the research. The model which is trained using SVM algorithm shows 98.7% accuracy in predicting tuberculosis than other algorithms.

In [12] author done a research using SVM algorithm. Authors' trained model shows 96.68 of accuracy. Trained model diagnosis whether a patient has Tuberculosis or not. The dataset in this paper includes 38 properties present in patient discharge reports.

In [13], the authors conducted a study and developed an Individual-Level Fatality Prediction of COVID-19 Patients Using AI Methods. Of the thirty-seven deaths in the test database thirty-six were accurately predicted to bring an Auto-encoder model that would reach 97% accuracy. The authors have told to do Fatality prediction of Covid-19 using chronic disease in the case of health habit or psychological factors, and symptoms as future works.

Chronic diseases like diabetes, cholesterol, heart problems are directly affected in a negative manner when they are suffering from pneumonia and covid-19. If the patient has bad health habits like smoking, large consumption of junk food, surgeries which have been done before also directly affect to the patient's severity level. So, these facts are not concerned in the existing applications. So, it's better to develop an application which considers all of these factors. Then the accuracy level of the applications will be increase. Chronic diseases and breath-holding time are most important factors when identifying COVID-19. Therefore, we suppose to implement a machine learning model to predict the severity level for COVID-19 based on those factors.

1.2 Research Gap

Currently, Covid-19 is one of the most infectious disease compared to the past decade. During this pandemic, there should be an emerging need to identify Covid-19 patients rapidly. Even though the 3rd world country like Sri Lanka has minimum resources to utilize to identify Covid-19 infected patients as soon as possible. On the other side, pneumonia is an infection that can easily affected to Covid-19. Because of this need, we try to propose a system to detect Covid-19 infected patients and Pneumonia infected patients by using a mobile app.

Comparing to the last project, they implement some methods to detect pneumonia using image processing and machine learning techniques. But still there is a vast gap between the existing applications researchers do not implement applications for detect Covid-19 using CXR and CT scan images. When looking at the previous apps, most of them are for chest x-ray interpretation. They do not provide a radiological guide to the Imaging of acute and chronic chest conditions. In terms of accuracy level and speed most existing applications are weak at those levels.

Mostly Coivid-19 is diagnosed by PCR test, but we have to wait at least 12 hours to get the result. Meanwhile, developing countries like Sri Lanka have to wait more than two-three days to get the results because we do not have enough equipments and laboratory facilities. Covid-19 patients should be assessed within a short period of time by several clinical physicians and minimal resources. With the introduction of medical Imaging, we trying to identify patients with COVID-19 and pneumonia with the help of chest X-rays.

Compared to different methods, Chest X-ray (CXR) tools required for this assessment is less annoying and lightweight and can be transported. Chest x-ray test takes approximately 15 seconds per patient, so CXR is a less time-consuming and cost-effective device.

When comparing to the previous work done, existing apps don't have the technology to recognize or to differentiate between these two lung diseases. There exists a vast gap between the existing apps and the proposed system of ours.

	App 1 (Pneumonia [14])	App2(corona check [15])	App 3 (x-ray classifier [16])	Proposed app
Develop a mobile app for upload chest X-ray image and detect Covid-19	×	×	×	ü
Develop a mobile app for upload chest X-ray image and detect pneumonia	×	×	ü	ü
Capture Chest x-ray image by using phone camera	×	×	ü	ü
Image enhancement for speed up the image preprocessing	×	×	×	ü
Doctor recommendation/health precautionary methods	ü	×	×	ü

TABLE 1.3: Comparison with other apps

Although there is plenty of x-ray classifier apps, those apps are not provided with the facility of medical intelligent bot. Most of the applications only have the classifier process and patient will not directed to medical recommendation and so on. Using the proposed app patients can tell their real time health condition, health history, other medications. So, the final output of the application will not only depend on the x-ray, CT scan classification but also depends on the patient's real time health condition and history. According to the literature survey, I could not find a proper app with all these facilities. So, there is a huge gap for an innovation.

For the illustration purpose three existing apps were downloaded and tested to get a clear understanding of the research gap to be filled. Most of the applications concerned

with the classification of other chest diseases compared to detection of pneumonia and covid-19. Those applications are limited to the x-ray classification and not concerned about the patient's real time health condition and health history. Therefor the accuracy level seems very low because they only depend on x-ray classification.

1.2 Research Problem

Image diagnosis is a vital problem in medicine. Now a day's ongoing pandemic of Covid-19 has become the major infectious diseases among all other diseases. As we

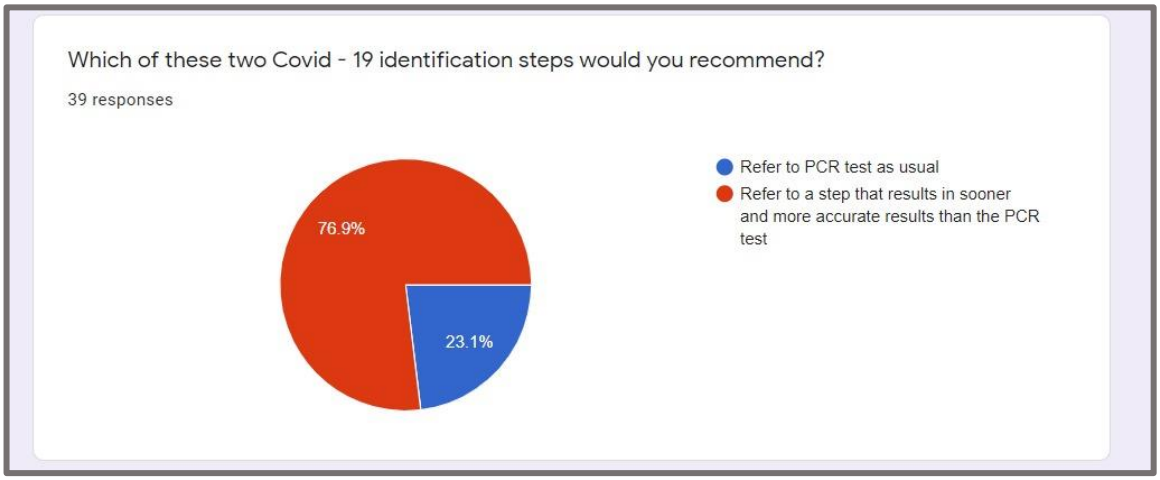


Figure 1.14 Summary of method people like to select when doing test.

all know, Covid-19 is a fast-spreading viral disease. Researchers found that patients with Covid-19 have proved that they are mostly affected by lung diseases. Covid-19 is diagnosed using RT-PCR Testing, CT Scans and Chest X-Ray (CXR) images. But the problem is RT-PCR test takes at least more than 12 hours to get a result. Covid-19 patients should be identified and tracked as soon as possible, and it requires specific material equipment. In developing countries, there is a lack of resources to do these tests. Even though methods that can generate faster results exist, these methods are too expensive for third-world countries like Sri Lanka.

It shows that most people like to get their reports in minimum time. In this kind of a situation patients must wait for a long time, and it is stressful. If there is a proper and speed method that gives the result in very accurate way that will be helpful.

Below image illustrates the average time which takes to a PCR test we gathered from our survey. It appears that most PCR test results are obtained after 3 days.

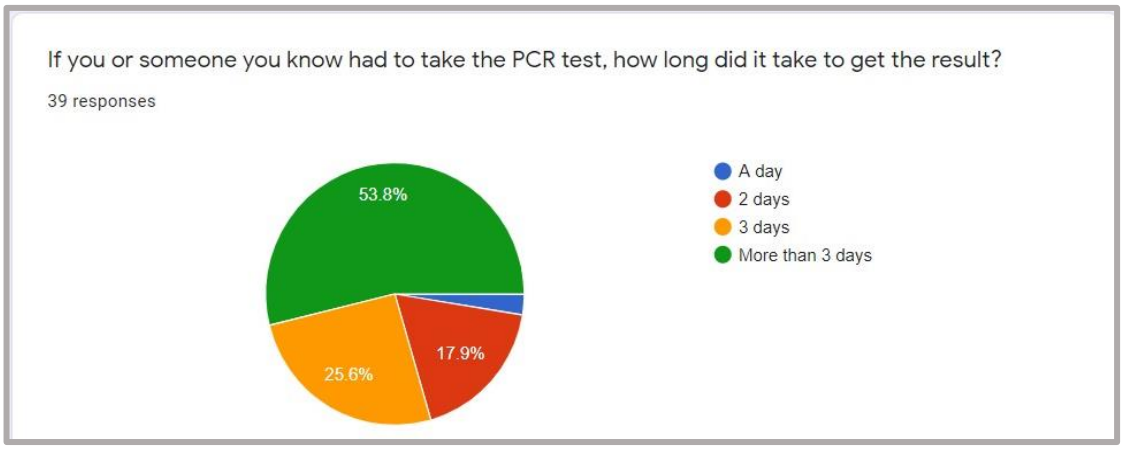


Figure 1.15- Average time takes to do a PCR test.

Because of that, chest X-ray images become the less cost and time-effective tool for take the decisions. Compared to other Chest X-ray methods, it is a lower-cost process, and anyone can easily access this method. In this global pandemic, frequent patients will need to be assessed in short periods by a few clinicians and very few resources.

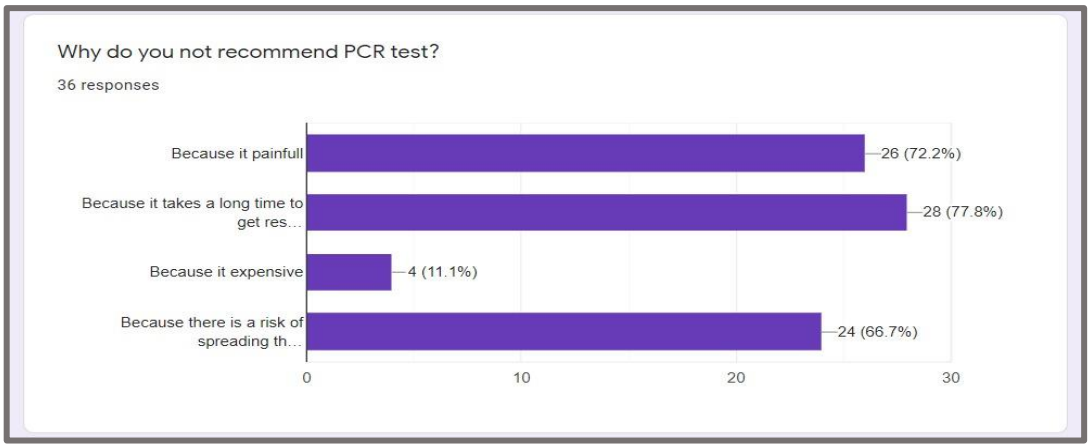


Figure 1.16 – why do not people choose PCR tests.

Not only Covid-19 but also Pneumonia is an infectious disease that needs to be diagnosed at an early stage. With the introduction of imaging, the diagnosis of diseases using X-ray images of the chest has been accelerated. But it also requires a specialist and a qualified radiologist to read the chest X-ray image correctly. To this purpose, we are suggesting a joined method using image processing and Convolution Neural Network for automated detection of covid-19 and pneumonia by using chest X-ray images.

2 OBJECTIVES

2.1 Main Objective

Our main objective is to implement an app to identify pneumonia/covid-19 in a minimum time and give patients health advice and medical recommendation. This app will reduce the cost which spends to do PCR tests and patients will get a high accurate result through the app. Patients will be asked about their symptoms, health history and advice according to their severity level.

One of the main objectives of implementing this app is identify positive or negative for pneumonia/covid-19 in a minimum time with high accuracy level. We suppose to screen x-rays and identify whether the patient suffering from pneumonia/covid-19. Main objective of screening x-ray images as it is a cheap method compared to PCR tests. People must spend lots of money when doing a PCR test. So, one of over main objectives is reduce the cost for medical tests.

Doing the test and give the result in minimum time is another major objective of this research. In developing countries people must wait so long to have the results as there's lack of facilities. But in this research, we target to give the results of the test in within a minimum time. Then patients do not stress and wait for the result.

If a patient considers not only the money, they spend but also the high accuracy level, they can go for CT scan screening procedure. With that function we suppose to maximize the accuracy level. According to the literature survey that we have done, CT scans gives high accurate results when compared to PCR tests. So, giving a high accurate result will be full fill by this option that we have planned to do.

Most important objective is after the patient is identified positive for pneumonia or covid-19 he/she should be guided to the medication or advices is a specific objective of this research. To advice the patient correctly we should first have a real understanding about the patient's current health condition and past health history. Otherwise, we cannot predict any advices.

2.2 Specific Objectives

- Understanding correct objects and automatically process the data of that specific object and ready for the disease identification process.
- Segment the captured chest x-ray image to identify the image.
- Adding color filtering algorithm to the captured image
- Organize processed Data for disease identification algorithms.
- Developing an algorithm to analyze a sample of captured chest x-ray images to predict the disease.
- Developing a method for classifying lung diseases.
- To develop a model for image acquisition and enhancement.

The uploaded image of the chest CT scan will be analyzed and prepared for segmentation. Various filters and tools are used for removing noises, change contrast, brightness and also other properties of the entire image or part of the image

- To develop a model for image segmentation.

The image of the chest CT scan will be analyzed and objects in the image will be separated. The aim of this objective is to separate the important parts from the background of the image.

- To develop a model for recognizing the Covid-19 infected chest CT scan images using Machine Learning techniques.

The AI model is developed with machine learning techniques to analyze and identify hidden patterns and relationships between the uploaded image and the trained data set of the AI model.

- To develop a mobile application.

The aim of this objective is to develop the mobile application. The mobile application is used for taking photos of a chest CT scan and upload it to the cloud-based AI model for analyzing.

Predict the severity level, give doctor validation if necessary is one of the most important section to be implement after the x-ray, CT scan classification. It is dominant to note that even though machine learning model can offer identify the severity level, valuable facts, and symptoms, they aren't qualified and don't have the authority to give an official treatment for diagnosis. Become the first point of contact before any human involvement is the main proposition behind this methodology.

- Implement user friendly reliable communication method to interact with patient and gather information:

  To give an accurate result we must get a clear understanding about the patient's current health condition and past. To do that. I developed a machine learning model to predict the severity level of the patient. This is a chunk of software that administer a survey kind of procedure to extract the information from the patient. chronic diseases, pregnancy, bad health habits, surgeries etc., are the basic information that we planned to ask from the patients.

- To monitoring current breath-holding time:

  To get an idea about what a patient is going through, the app asks about their breathing information. Through the app it counts the breath holding time and take it to the analyzation process.

- Doctor validation.

Analyze the severity level of the patient considering about all the data that gathered via the mobile app. According to the severity level patient will be notified about their current risk level and a doctor validation will be given if the patient make a request via the mobile application.

Training three model and select model that have highest accuracy to predict risk level and classify the covid-19 and covid-pneumonia using patient's symptoms and patient's age and sex within this year.

- To develop model for classify covid-19 and pneumonia

In this train machine learning model to classify covid-19 or pneumonia.

- To improve accuracy of trained algorithm.

Algorithm should be trained and improved accuracy until come to expected accuracy.

- To develop user friendly mobile application

In this objective develop a mobile application to collect symptoms and upload it to the cloud for do the prediction and classification through the Machine learning model.

3 METHODOLOGY

X-ray image Classification

We plan to use a publicly available Covid-19 and Pneumonia Detection dataset of Chest X-rays in Kaggle [17]. Detection of Covid-19 and Pneumonia by chest X-ray images can be divided into two main categories. The first step is image preprocessing, and the other method is the deep learning model for detecting Covid-19 and Pneumonia.

Image pre-processing part has been performed as follows:

- Image acquisition:

  Obtaining the image is the first step of image processing. It is the way of collecting the image from the physical device. It is one of the major roles of this process because everything depends on the input image. There are three methods to take the image in acquisition model and they are using a single sensor, line sensor and array sensor.

- Image enhancement

  Image enhancement is used to process the image according to the specific application. That enhanced image is more appropriate than the original image we obtained. Image enhancement contains two categorical methods, and they are spatial domain method and the Frequency domain method. In enhancement process we must increase the contrast and then we need to block the unrelated parts of the images including objects and the color to identify the relevant objects using threshold operations.

- Image Segmentation

  It is the process of partitioning into meaningful zones relative to the image application. In here image Segmentation here is mainly used to crop the lung nodule area in the CXR image. Using Contrast Limited Adaptive Histogram Equalization chest x-ray images are replaced as cropped image. The segmentation is based on various measurement taken from the image, and it can be grey level, color, texture, or motion.

There are two main techniques used in the image segmentation and they are, Thresholding and the Clustering. Both are based on grey level histograms. All images are available modified to optimal size for computer purposes.

In our work we used histogram equalization to enhance contrast of the Chest X-ray (CXR) image.

1. Histogram equalization
   This algorithm is used to adjust image intensity so we can enhance the contrast of the CXR image.
2. Copping of the chest X-Ray lower region

In the computer vision image acquisition, image enhancement, image segmentation has a vital role. It helps for the identifying objects from the complex backgrounds and any kind of image vision application.

Next step is to train the model with prepared data set. In order to do this CNN is used. After the pre-processing part image is entered into the convolution layer. This system consists of several compromised systems blended with nonlinear and pooling layers.

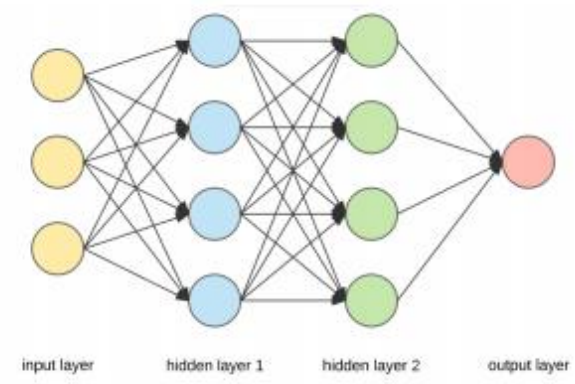


Figure 3.1 CNN Layer architecture

CXR image Recognition using CNN (Convolutional Neural Network)

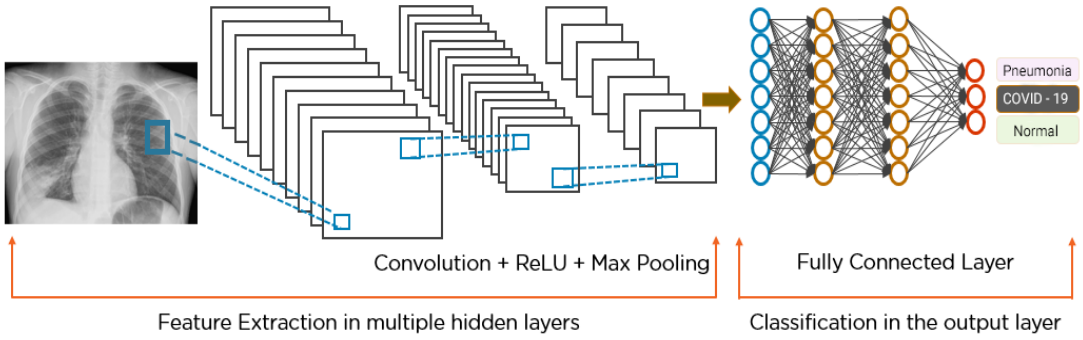


Figure 3.2 Basic CNN Structure in the system

Image Classification task has been performed as follows:

In here CNN layer contains three layers.

1. Convolution Layer
2. Pooling Layer
3. Fully Connected Layer

Convolution layer is used for extracting the features from CXR image. After that pooling layer is used. The purpose of this layer is a smaller image comparison. The Pooling layer calculates the maximum or ordinary value of an appropriate feature over the input chest x-ray image region. Fully connected layer is the final layer of the Convolutional Neural Network (CNN). In this layer, object classification is happening. This layer take input from the previous layer. It may be Conv layer, Relu layer or pooling layer. There is a N dimensional vector as an output in this layer. Where N is the number of classes network has to choose. Fully connected layer uses different approaches to represent the output like softmax approach or multiclass Support Vector Machine (SVM) under linear classification.

Backpropagation algorithm plays a key role in any neural network. Backpropagation is a generally applied algorithm for preparing feed-forward neural networks. We are planning to use Backpropagation algorithm to minimize the errors and update the weights according to the learning rate.

Then we used Activate function. It converts an input signal to an output signal.

Next step is implementing the mobile application for the model to be used. Since most mobile phones do not have the required computing power to work with such a deep learning model so we have to implement a separate web service to run the model.

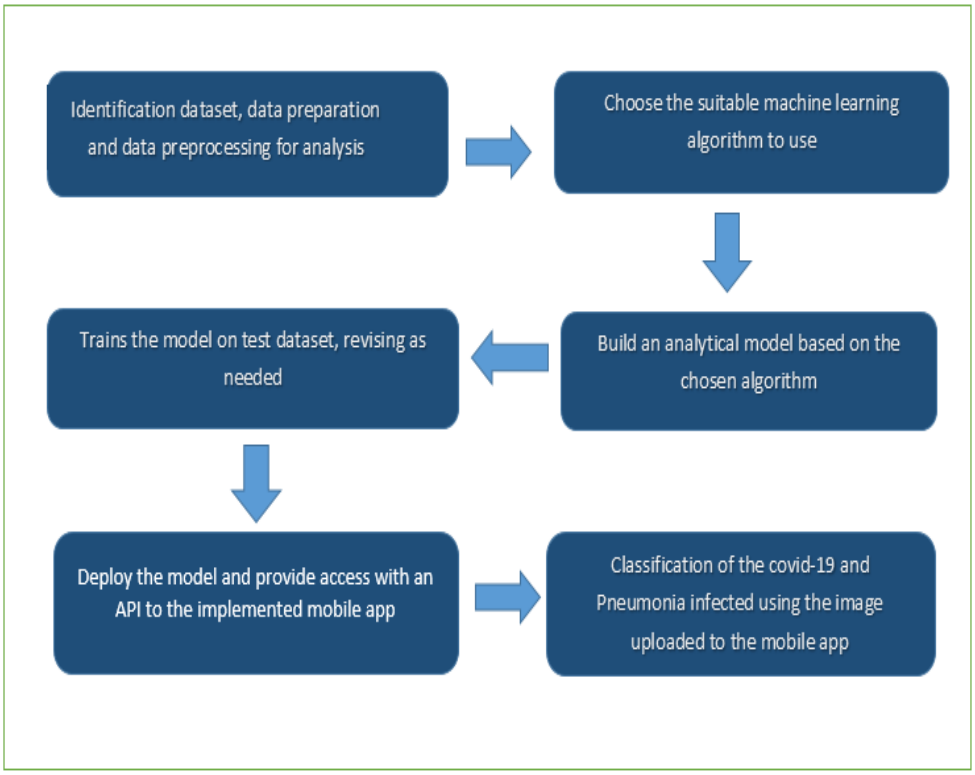


Figure 3.3- implementation procedure

We are planning to use Flask library to implement the API so that mobile application can communicate with model using the exposed API endpoints. In the mobile application we are planning to implement a functionality to upload a chest x-ray image from the gallery or to capture an image from the camera while using mobile app itself.

Once the web service receives the chest x-ray image and fended into the model result will be returned to mobile application.

CT-scan image classification

Image enhancement is the technique of processing the chest CT scan image according to the specific application. That enhanced image is more suitable than the original

image that we acquired. There are two categorical method in image enhancements, and they are, spatial domain method and the Frequency domain method.

Image Segmentation is the process of partition chest CT scan image into meaningful regions concerning image application. The segmentation is based on different measurement taken from the image, and it may be grey level, color, texture, depth, or motion. There are two main techniques used in the image segmentation and they are, Thresholding and the Clustering. Both are based on grey level histograms.

Object detection in a complex background is an important and a difficult task. To detect these objects, color methods and shape detection techniques are being used with thresholding and Circular Hough Transformation. There are many object detection challenges such as target object may have interfaced with other objects and chest CT scan image can be contained noise and difficult to object detection process and object may be overlapped.

Image Recognition and CNN (Convolutional Neural Network)

There are three layers in the convolutional neural network. They are,

- Convolution Layer
- Pooling Layer
- Fully Connected Layer

Convolution Layer

The main purpose of the Convolution layer is to extract the features. To extract the features, we need to implement a mask or convent. Number of filters will be equal to the number of features.

Pooling Layer

The main function of the pooling layer is to shrink the chest CT scan image to small size. Pooling layers computes maximum or average value of particular feature over the region of input data.

Fully Connected Layer

Fully connected layer is the final layer of a Convolutional Neural Network. In this layer, object classification is happening. This layer take input from the previous layer. It may be Conv layer, ReLU layer or Pooling layer. There is a N dimensional vector as an output in this layer. Where N is the number of classes network has to choose. Fully connected layer uses different approaches to represent the output like SoftMax approach or multiclass Support Vector Machine SVM under linear classification.

Back propagation algorithm

Back propagation algorithm plays a major role in any neural network. It updates the weights using optimizations algorithms like Gradient decent. It minimizes the errors and update weights according to the learning rate with higher learning rate and low learning rate. With low learning rate, neural network gives reliable results, but it will take long time.

Activation Function

Activate function is activating the neurons in the neural network using the threshold value. It converts an input signal to an output signal. This output signal is used as an input signal in next layer. There are many activations functions in the deep learning. When we are implementing a network, we need to decide which activation function is use for neural network.

Dataset

A suitable data set that matches the defined purpose will be chosen from (kaggle.com) It consist of 2500 chest CT-Scan images, which has 1260 of COVID-19 Pneumonia images and 1240 of non-COVID19 Pneumonia images.

Symptom classification

In this part we predict covid-19 disease. And we predict risk level of user have corona-pneumonia level. Risk stage is predicted using user age gender and symptoms. Willing to use three algorithms at the first stage. Then select one algorithms among them that have highest algorithm. After that increase accuracy of trained model.

- Get data

The first step in this project is find suitable data [patient details (age, sex) with symptoms] for train our models. Find data set using www.kaggle.com and datasetsearch.research.google.com .

- Clean, prepare and manipulate Data

  Raw data has unorganized noise or missing rows and columns. Before train model the data set all missing and noise data must clean.

- Trains model

  Logistic regression, SVM, random forest, SVM single class model, isolate forest and outlier factor are most commonly using for trained models. SVM, Random forest and Auto-Encoder are most used classification and prediction algorithms.

  - Auto-encoder

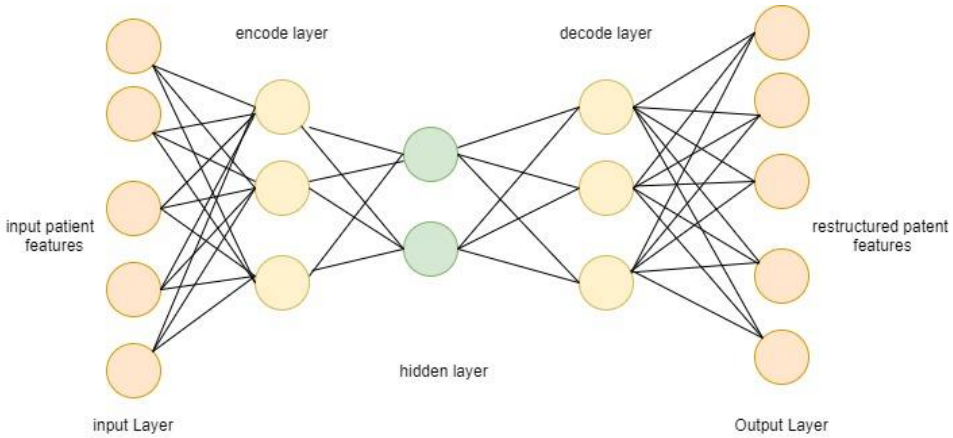


Figure3.4: Auto-encoder neural network

Auto-encoder is an artificial neural network that learns the representative codes from the inputs. Then map these codes to the inputs. This format is usually used to encode inputs and output a compressed version of that inputs.

  - SVM

    SVM is supervised ML (machine learning) classification algorithm. SVM attempts to find the hyperplane with the maximum distance between two classes.

  - Random forest classifier

Random Forest is a tree-based learning algorithm that uses decision trees that emerge from a randomly selected training subset to solve a taxonomy problem. The new class of SVM algorithms typically use only "normal" data to determine whether a new-data record is similar to the training kit.

- Validate model
    - Train, test & validation split

30% of whole dataset use to test the trained model. Another 70% of data use to train model and validate the trained model.

Health History and breath-holding time classification

- Data gathering:

  A suitable data set that matches the defined purpose will be chosen from https://www.kaggle.com etc. Chronic diseases, pregnancy, smoking, obesity, surgeries are some of the special information consider when selecting the data set.

- Clean and prepare for machine learning process:

  In machine learning projects, data cleaning is the first essential step. If there are erroneous records it should be recognize first. Then it should be correct (or removing) or erroneous records from a record set, table, or database and refers to point out partial, incorrect, imprecise, or inapplicable parts of the data and then replacing, modifying, or deleting the dirty or unrefined data will be appraise in this level.

- Trains the model:

  Inquiring about the patient's disease is done by prognosis of work carried out on data of various domains and sizes. Required time to train and test the accuracy may also differ according to the different algorithms [12]. We get the class-wise circulation of the record by training the dataset of non-identical count while SVM holds proper command over scattered categorization tasks [12]. K-nearest neighbor algorithm (KNN) and naïve holds only quick and

simple classification. CNN, RNN and HAN in deep learning algorithms are deployed for training and testing to pinpoint the emotion from textual conversation had with the patient.

- Classifiers:

  To emotion classification from textual outputs, three classifiers of deep neural networks viz, CNN, RNN, and HAN are used in this level.

  CNN: It will be used to feed-forward artificial neural networks (NL) where relation between nodes do not form a cycle and use a distinction of multilayer perceptron's designed to aspect require preprocessing [12].

  RNN: Behavior for a time sequence [13] will be manifest in this level.

  HNN: Structure of hierarchical document and different instructive words in the textual conversation are the two primary aspects captured in this level. HAN has gained high accuracy and acts better than CNN and RNN if we have massive dataset.

- Severity level identification and doctor recommendation:

  After the process of gathering the data of each textual conversation with the patient, which is the procedure of allocating one of the severity levels to the textual conversation according to its content. Then the advices, precautionary methods, doctor recommendations will be given to the patient based on the severity level.

3.1 System Overview Diagram

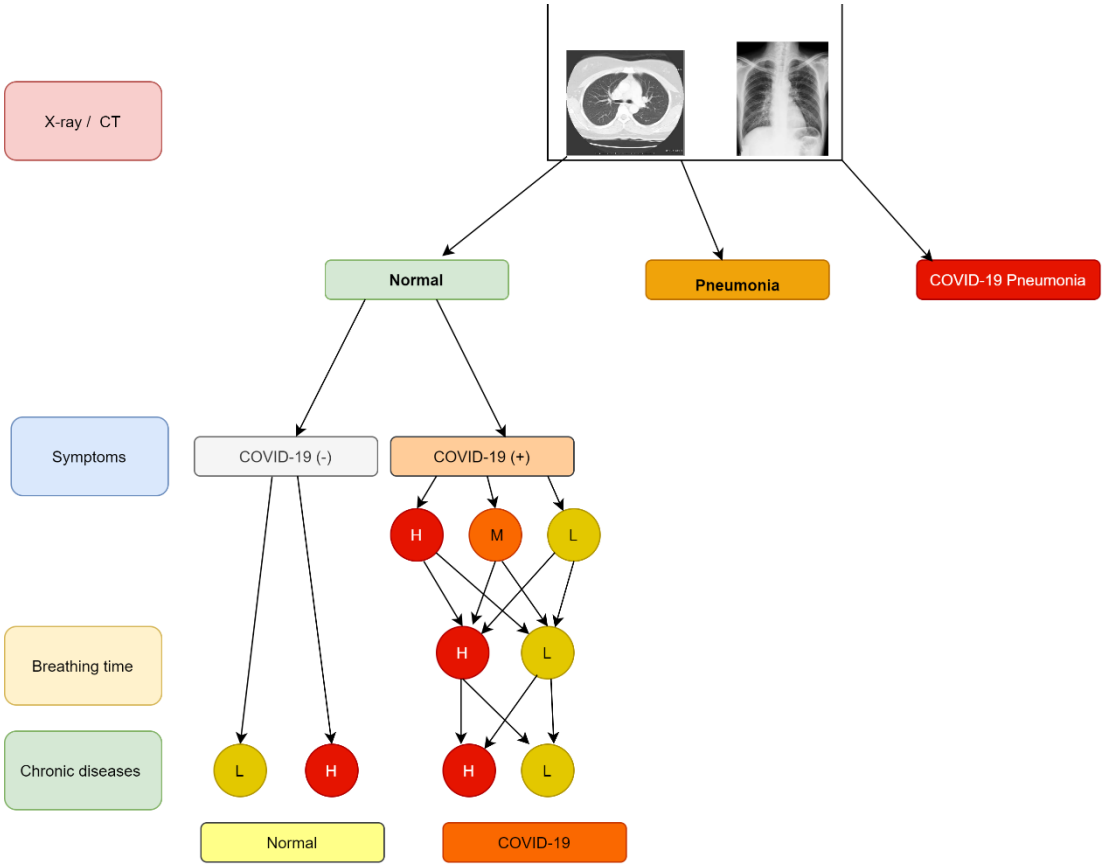


Figure 3.5 – System Overview Diagram

3.2 Work Breakdown Architecture

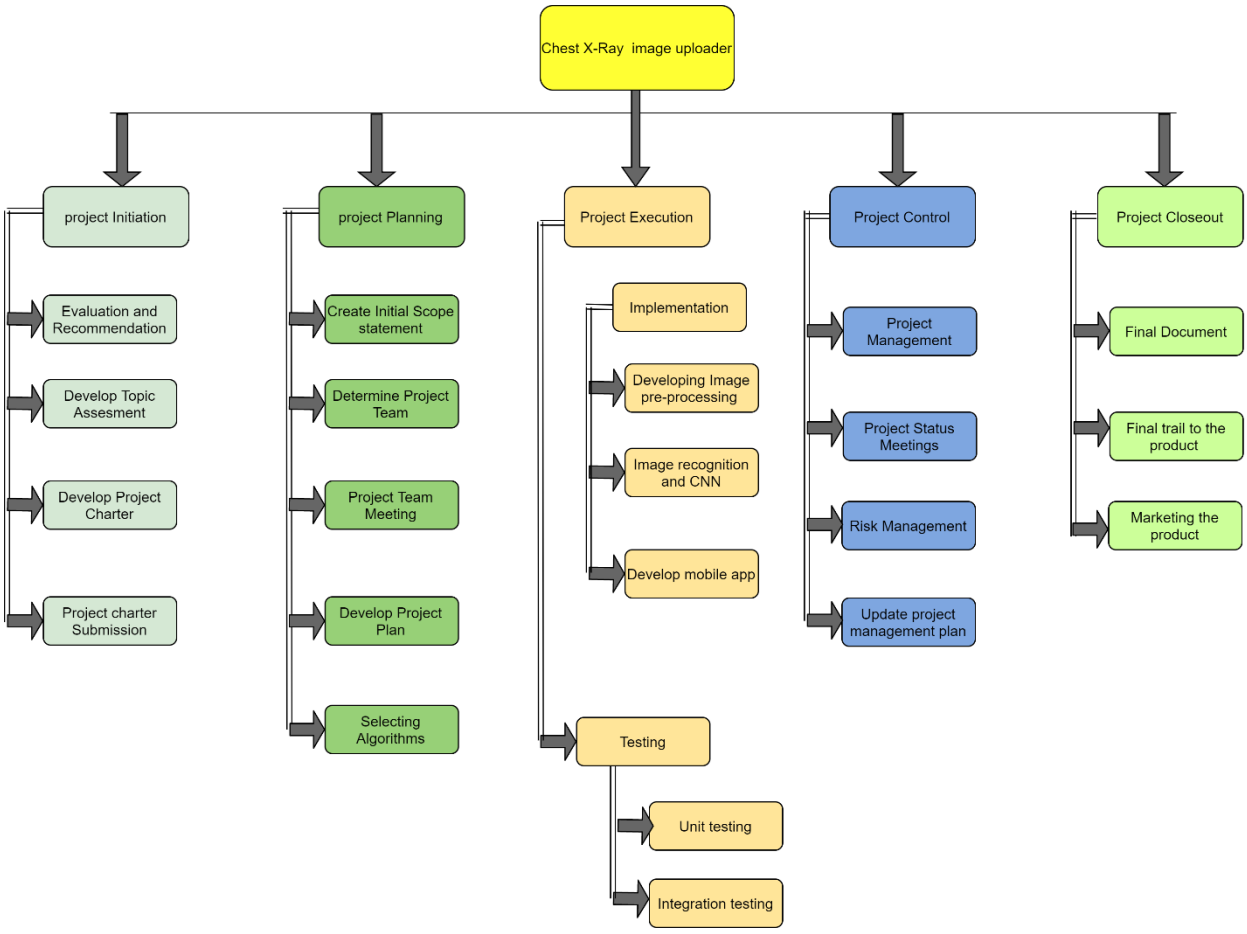


Figure 3.6-Work breakdown Architecture

3.3 Gantt Chart

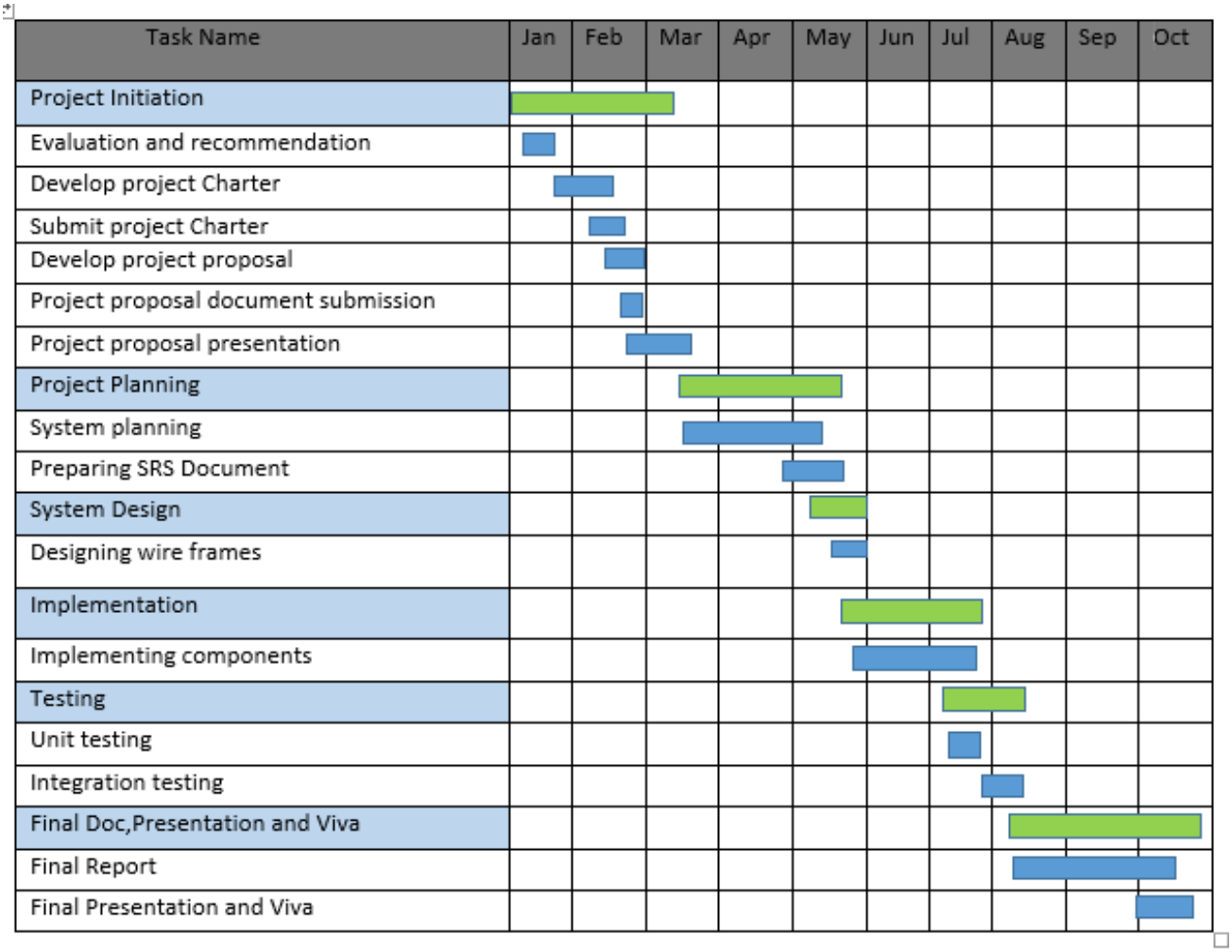


Figure 3.7- Gantt Chart

4. PROJECT REQUIREMENTS

Hardware requirements

Windows 10

64-bit (*64) processor

2.70-GHz or faster processor

8GB RAM

Software requirements

OpenCVSharp

OpenCV Library

Android Studio

TensorFlow

Proto.io

Firebase

Functional requirements

System should use for teenagers and as well as adults.

Display the user's awareness level.

It should be easily understandable to everyone.

There should be a proper way to upload or capture the chest X-ray image.

Non-Functional requirements

Availability – user can access this anywhere at any time.

Reliability – always show the accuracy result to the user.

Usability – anyone can easily access the app.

Performance – fast uploading when it is opening.

5 BUDGET AND BUDGET JUSTIFICATION

TABLE 5.1 Budget Calculation

Item	Price (Rs.)
Sample X-ray	2500
Documentation	3000
Printing photocopies	2000
Total	7500

REFERENCE

[1] Tabik, S.;A. Gómez-Ríos, J. L. Martín-Rodríguez, I. Sevillano-García, M. Rey-Area, D. Charte,, "COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-Ray Images," IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS,, 2020.

[2] H. Sharma, J. Sethia Jain, P. Bansa and S. Gupta, ""Feature Extraction and Classification of Chest X-Ray Image using CNN to detect Pneumonia"," 2020.

[3] Md. Mehed Hasan, Mir Md. Jahangir Kabir, Md. Rakibul Haque, Mohiuddin Ahmed, "A Combined Approach Using Image Processing and Deep learning to detect Pneumonia from chest X-rays," *IEEE,* 2019.

[4] D. A. Ragab, et al., "Breast cancer detection using deep convolutional neural networks and support vector machines," PeerJ 7:e6201, 2018.

[5]. Sharma, A., Raju, D., & Ranjan, S., Detection of pneumonia clouds in chest X-ray using image processing approach." 2017 Nirma University International Conference on Engineering (NUiCONE). IEEE, 2017.

[6] J. Luengo et al., Big Data Preprocessing - Enabling Smart Data. Springer, 2020.

[7]. Gonzalez, R. C., "Digital Image Processing/Richard E." Woods. Interscience, NY (2017), ch. 3, pp. 126-127.

[8] Azian Azamimi Abdullah, Norafifah Md Posdzi, Yoshifumi Nishio, “ Premilinary Study of Pneumonia Symptoms Detection Method using Cellular Neural Network,” *IEEE*, 2011.

[9] Dhiraj Dahiwade, 2Prof. Gajanan Patle, 2Prof. Ektaa Meshram (2019). Designing Disease Prediction Model Using Machine Learning Approach

[10]. Aiyesha Sadiya, Anusha V Illur, Aekhata Nanda, Eshwar Rao, Vidyashree K P, Mansoor Ahmed (2019). Differential Diagnosis of Tuberculosis and Pneumonia using Machine Learning

[11]. T, A., Natarajan, S., & Murthy, K. N. B. (2011). A Data Mining Approach to the Diagnosis of Tuberculosis by Cascading Clustering and Classification. Journal of Computing

[12]. Yahiaoui, A., Er, O., &Yumusak, N. (2017). A new method of automatic recognition for tuberculosis disease diagnosis using support vector machines. Biomedical Research, 28(9), 4208– 4212.

[13]. Individual-Level Fatality Prediction of COVID-19 Patients Using AI Methods (2019). https://www.frontiersin.org/articles/10.3389/fpubh.2020.587937/full#B8

[14] Kadira Apps.2019. Chest X-Ray Interpretation (version 1.0). [Chest X-Ray Interpretation]. [Accessed by 11 July 2019]

[15] Aga Khan University.2020. CoronaCheck (version 1.1). [CoronaCheck]. [Accessed by 30 Mar 2020]

[16] AeroTech Studios.2020. Chest X-ray Classifier (version 1.0). [Chest X-ray Classifier]. [Accessed by 30 Sept 2020]

[17] Chest X-ray images Available online https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia

APPENDICES