



# PNEUMONIA AND HEART DISEASE DETECTION USING MACHINE LEARNING

Prathamesh Jadhav, Trupti Sonawane, Aakanksha Jadhav, Atharv Phadale, Neeta Chavan  
VESIT, Mumbai University.  
Mumbai, India

**Abstract**– The proposed project health care kit involves a heart disease prediction system and Pneumonia detection. Pneumonia is an infection that inflames the air sacs in one of the lungs or both the lungs. The aim is to build an application for pre-detection of pneumonia by using machine learning algorithms to analyze important parameters that lead to the disease which are delivered with consultation of doctors so we are automatically identifying whether possibility of a patient has pneumonia. Various machine learning algorithms are first compared to find the suitable algorithm giving maximum accuracy. If the person shows the possibility to have the disease then, in the second stage where X-rays are required, person to person vision may not be perfect. So, with the help of ensemble learning and CNN classification model we classify the pneumonia infected lungs.

Heart disease, often known as cardiovascular disease, is the leading cause of death globally over the past few decades. It may include a variety of disorders that have severe impact on the heart conditions. Heart disease prediction system seeks to use data mining methods on medical data sets to help in heart disease prediction.

**Keywords**-. pneumonia, heart disease, machine learning, early detection, x-ray, android application.

## I. INTRODUCTION

Pneumonia is responsible for the majority of deaths across the globe, particularly among children under the age of 5 and individuals aged 75 and above. As per the estimations of the Government of India, pneumonia leads to 140,000 deaths of children under the age of five each year in India.. Pneumonia is a respiratory illness that can vary in severity, ranging from mild to extremely severe and life-threatening. The most vulnerable populations are infants, young children, the elderly, and individuals with compromised immune systems or preexisting health conditions. The risk of pneumonia is particularly high for people living in developing countries where energy poverty is prevalent and the majority of people rely on harmful forms of energy. According to the World Health Organization (WHO), more than 4 million premature deaths occur each year due to pneumonia and other diseases caused by household air

pollution. Over 150 million people get infected with pneumonia on an annual basis, especially children under 5 years old. The lack of medical resources and personnel in these regions can further worsen the problem.

In the case of heart disease prediction, cardiovascular diseases can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment and guidance the symptoms of heart disease can be reduced, and the functioning of the heart can be improved. According to WHO An estimated 17.9 million people died from CVDs in 2019, there are no symptoms of the underlying disease of the blood vessels. A heart attack or stroke may be one of the first signs of many underlying diseases. Furthermore, the individual may encounter challenges in breathing or have a sense of breathlessness, feel nauseous or throw up, experience dizziness or faintness, break out in a cold sweat, and their skin tone may appear paler than usual.

The primary objective is to accurately predict the presence of heart disease using a minimal number of tests. Attributes may be considered from the primary basis for tests. A strong dataset is required to get accurate results and to present a durable system. The Mobile application that will automate the task of identifying whether the patient has a possibility to have pneumonia or heart disease. The pre-detection of pneumonia using to see if the chest X-rays are required and then identify pneumonia using the x-rays. The system has the ability to identify and extract concealed insights related to illnesses from a past dataset on heart conditions. The objective of the heart disease prediction system is to utilize data mining methods on medical datasets to aid in foreseeing the occurrence of heart diseases.

## II. LITERATURE SURVEY

Many researchers have introduced methods for heart prediction and chest x-ray classification. Some of these methods use handcrafted feature extraction techniques along with a machine learning algorithm as a classification technique, whereas others use deep learning techniques for feature extraction and classification [1].

By using the logistic regression approach on this dataset, the authors of [2] were able to attain a prediction accuracy of 77%. Authors in [3] focused on the construction of an artificial intelligence-based heart disease detection system



using machine learning algorithms, they discussed the three primary stages of application development: gathering databases, applying logistic regression, and assessing the properties of the dataset. A random forest classifier method is created to more accurately detect cardiac problems. For this application, which is important because it outperforms training data by about 83%, data analysis is required.

In [4] the authors examined a dataset obtained from Kaggle that included characteristics associated with heart disease, like age, gender, blood pressure, cholesterol, and so forth. Additionally, they looked into how accurate other machine learning methods like Support Vector Machines, K-Nearest Neighbor, and Decision Trees were. When used with a huge dataset, the aforementioned algorithms' performance and accuracy are not as good, hence in this case, they sought to increase prediction accuracy utilizing Tensor Flow, Artificial Neural Networks, and Keras.

As for Pneumonia, it is a life-threatening infectious disease affecting one or both lungs in humans commonly caused by bacteria called *Streptococcus pneumoniae*, a study of pre-trained CNN models is proposed by [5] to find out that CNN models employed along with supervised classifier algorithms can be very beneficial in analyzing chest X-ray images, specifically to detect Pneumonia.

Authors of [6] created a computer-aided diagnosis method that uses chest X-ray pictures to automatically detect pneumonia. To address the lack of available data, we used deep transfer learning and created an ensemble of three convolutional neural network models: GoogLeNet, ResNet-18, and DenseNet-121.

A weighted average ensemble strategy was used, and a unique method was used to decide the weights given to the base learners. The proposed method achieved accuracy rates of 98.81% and 86.85% and sensitivity rates of 98.80% and 87.02% on the Kermany and RSNA datasets. Unfortunately, they did not find a decision tree suitable for the data.

To solve the data scarcity problem in biomedical image classification tasks, transfer learning, wherein knowledge gained from a large dataset is used to fine-tune the model on a current small dataset, is currently a frequently used approach [7]. [8] applied purely transfer learning approaches in which different CNN models pre-trained on ImageNet [9] data are used for pneumonia classification to tabulate the development of the state of the art for the pneumonia detection problem.

Modern deep learning techniques for pneumonia detection typically concentrate on using a single CNN model. By combining the conclusions made by various CNN models, ensemble learning [10] [11] efficiently incorporates the key characteristics of all of its base models, gathers supplementary data from various classifiers, and makes decisions more reliable.

### III. PROBLEM STATEMENT

As per the estimates of the Government of India, Pneumonia causes around 140,000 deaths among children under the age of five in India every year. This can be because of the late detection and the wrong diagnosis or neglect of early symptoms. Early Prediction in this case is a huge requirement, The Main Aim is to focus on pre-detection of pneumonia using various features which are delivered with consultation of doctors so that possibility of having pneumonia is pre detected and appropriate actions can be taken

.Building an algorithm to automatically identify whether there is a possibility that the patient has pneumonia or not by analyzing some important symptoms that lead to the disease. Even in the second stage where X-rays are required, person to person vision may not be perfect. The algorithm has to be extremely accurate because lives might be depending upon it.

### IV. PROPOSED SYSTEM

The planned method is intended to make the early detection possible for both pneumonia and heart disease. An android application integrated with machine learning models are included in this system. Comparison of various models to choose the model with most accuracy is done while developing this system. In Stage one for Pneumonia Detection, assessment based on the symptoms like chest pain types, breathlessness, cough, etc. and if there's a possibility of having pneumonia then it progresses to the stage two where final detection from chest x-ray images is done.

For heart disease prediction assessment is done for some symptoms and doctors and tests are assigned if possibility is shown. After detection, a facility to contact a doctor is provided in the application.

V. METHODOLOGY

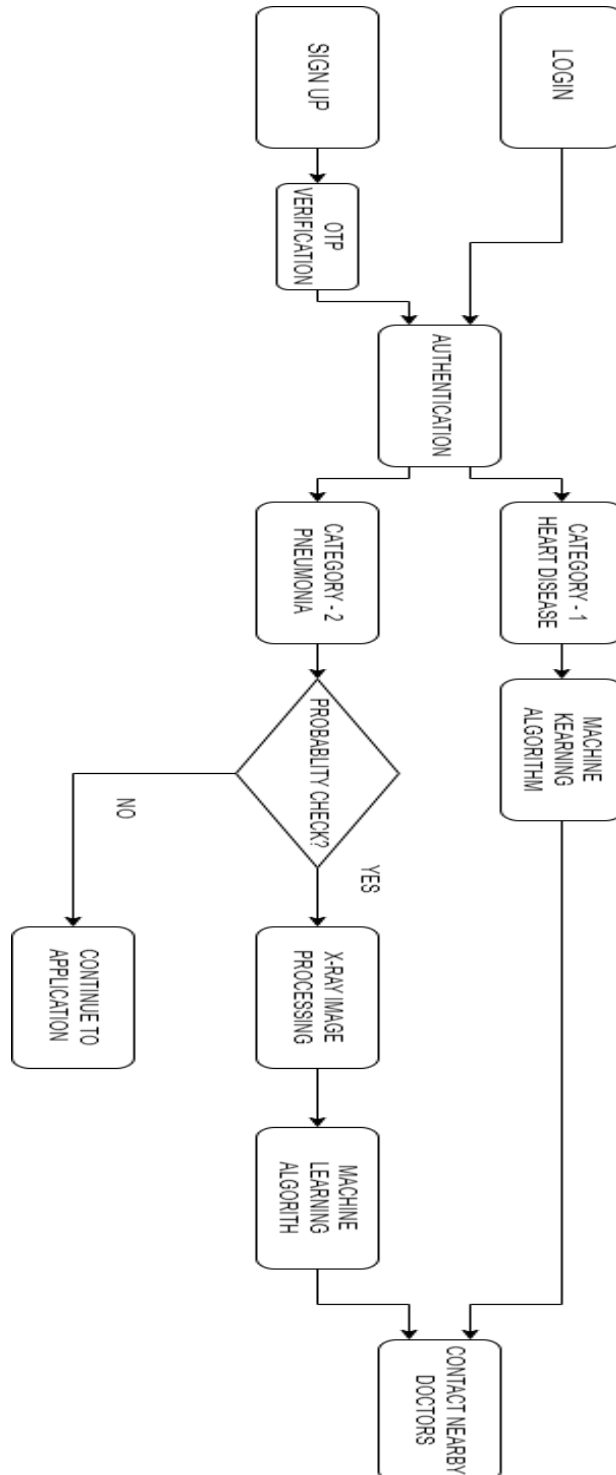


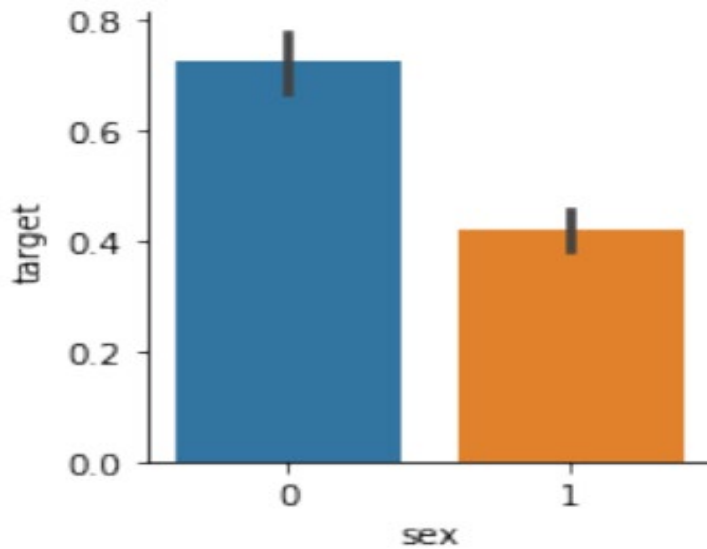
Figure 1. Flowchart of the system



**1. HEART DISEASE ASSESSMENT MODEL**

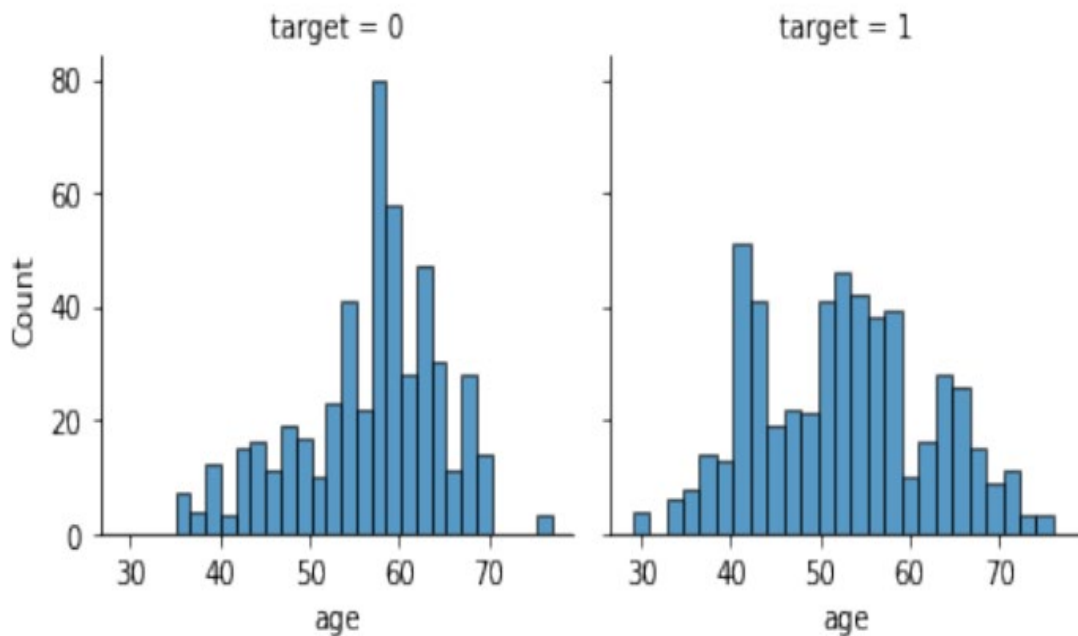
The user or patient will be asked a set of questions with the help of researched parameters and the machine learning algorithm will work on that to predict if the patient has heart disease or not. The diagnosis of heart disease is usually based on signs, symptoms and physical examination of the patient. There are several factors that increase the risk of heart disease, such as smoking habit, body cholesterol level,

family history of heart disease, obesity, high blood pressure, and lack of physical exercise. The model used in Heart disease prediction is a machine learning model named Random Forest. with Test accuracy over 96%. Analysis on the dataset obtained from kaggle is simplified so that one can understand the possibilities according with each attribute.



**Figure 2. Comparison of male/female attribute**

- Provided Dataset contains 72% data of females and 42% data of males



**Figure 3. Comparison of age**



- The age group between range(40-60) have greater probability facing heart diseases  
 Likewise all the attributes were tested before training the model.

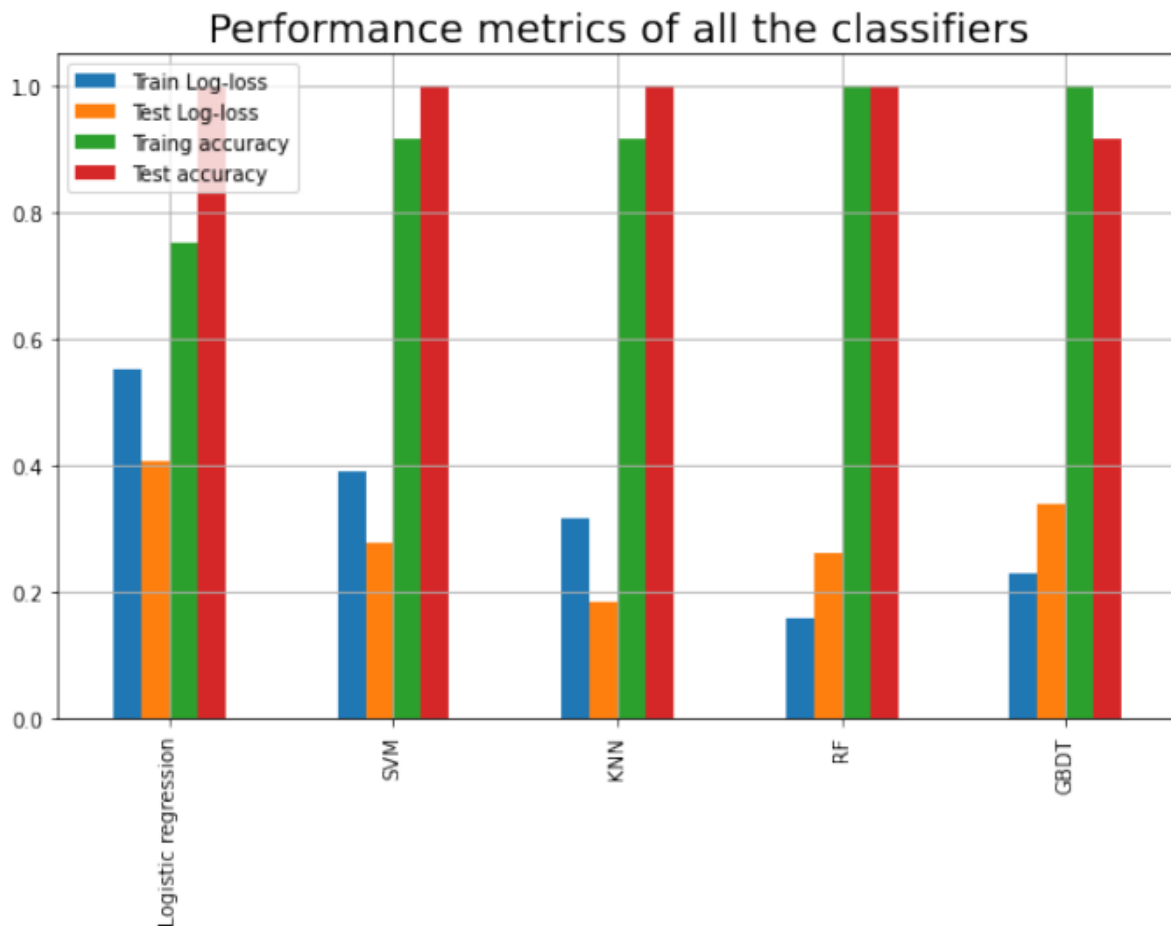
```

age sex cp trestbps chol fbs restecg thalach exang oldpeak
slope \
0 52 1 0 125 212 0 1 168 0 1.0
2
1 53 1 0 140 203 1 0 155 1 3.1
0
2 70 1 0 145 174 0 1 125 1 2.6
0
3 61 1 0 148 203 0 1 161 0 0.0
2
4 62 0 0 138 294 1 1 106 0 1.9
    
```

**Figure 4. Python output of All Attributes**

**2. PNEUMONIA ASSESSMENT MODEL**

The data is collected from patients using common attributes, So that efficient prediction of the model can be achieved. There are total 13 attributes which includes: Name, Age, Breathlessness, Dehydration, Type of cough (dry, wet), Heart Rate, Chest Pain, Color of eyes(Brown, Normal), Smoking. We used different classifiers to see which one gives the best accuracy. From Logistic regression, SVM, KNN, Random Forest, GBDT it is observed that Random forest gave the most accuracy.



**Figure 5. Comparison of models**

**3. DETECTION USING X-RAY MODEL**

This is stage two of the pneumonia detection where if the user already has a chest x-ray image or they have gotten the x-ray done because of having been detected there is a possibility to have pneumonia. We have used Transfer

learning technique for training pneumonia detection model, The model used is Mobilenet V2. MobileNet V2 model was developed at Google, pre-trained on the ImageNet dataset with 1.4M images and 1000 classes of web images. Which gives test accuracy of 85%.

We used feature extraction for extracting the meaning features of the infiltrates from the x-ray image.

Both the shape descriptors and texture descriptors were obtained by analyzing the external boundary of the x-ray image.

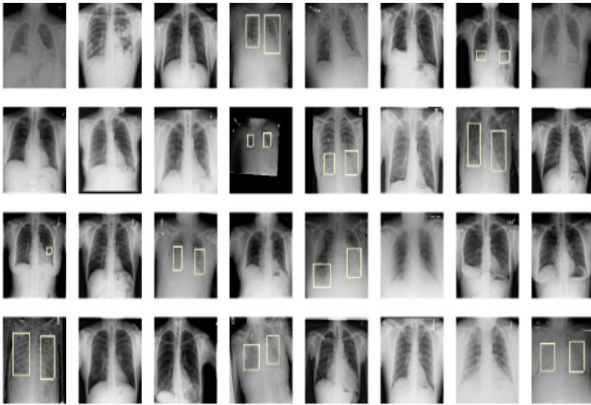


Figure 6.feature Extraction from X-ray images

#### 4. IMPLEMENTATION

We built an application which is integrated with a machine learning model to detect pneumonia and heart disease. To use this system the user needs to first install the Application which is called “Remote care” on their handset. They need to sign up, that is they have to fill the required data and create an account. After signing up the user just needs to login with their credentials to enter the application. Once the user is in the application depending upon the symptoms, they are feeling they can choose between two given options which are Pneumonia assessment and Heart Disease Assessment.

##### Pneumonia Detection :

- Users will be asked if they have any recent chest x-ray or not.
- If pressed yes, they will be moved to the x-ray image classification page.
- If selected no, they will be moved to the assessment page.
- After taking responses, the results will be provided in percentage form.
- If it's showing a higher possibility of a patient having pneumonia then they will be suggested for a chest x-ray.
- Next step will be chest x-ray classification to find out strong results of the diagnosis.
- If pneumonia is detected then only the user will be provided with a list of nearby doctors and hospitals.

##### Heart Disease Detection:

- Users will be asked a set of questions which they will have to answer according to their physics and surroundings.
- They will get the result of the possibility of having a heart disease.
- The next step is to get a consultation of nearby doctors for further tests.

#### 5. COMPONENTS

##### a) Google Colab

We used Google Colab which is a cloud-based platform that allows users to write, run, and share Python code using Google's infrastructure. Google Colab provides free access to GPUs and TPUs, which can significantly speed up training and inference for deep learning models. It allows users to easily share their notebooks with others by generating a shareable link or exporting the notebook as a Jupyter notebook or PDF. It is easy to share collaboratories with an active community allowing developers to work together efficiently.

##### b) Android Studio

Android Studio is a popular Integrated Development Environment (IDE) used by developers to build Android applications. Android Studio can be integrated with other tools like Git, Gradle, and Firebase, allowing developers to streamline their workflow. We developed our application through the various functions provided by android studio. It Also has a large and active community of developers who are willing to share their knowledge and help others with their development tasks.

#### VI. RESULTS AND ANALYSIS

The user will need to install “Remote Care” on their handsets in order to use this detection system. The models used in the proposed system gives high accuracy, sensitivity, and specificity in detecting pneumonia from chest X-ray images and using the symptoms as well as in detecting heart disease.

After getting the results of possibility the patients are provided with contact of doctors along with their clinic locations. Random Forest Provides accuracy of around 95% for symptoms assessment. While for chest X-ray detection overall has 85% accuracy.

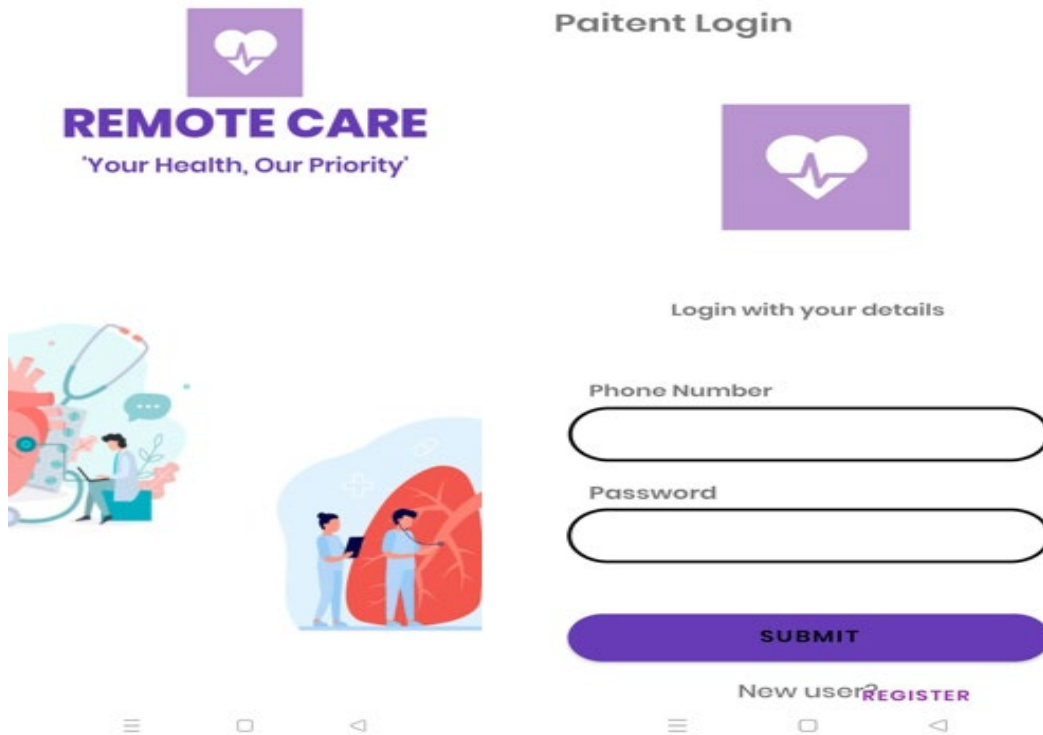


Figure 7 Splash Screen. Patient Login

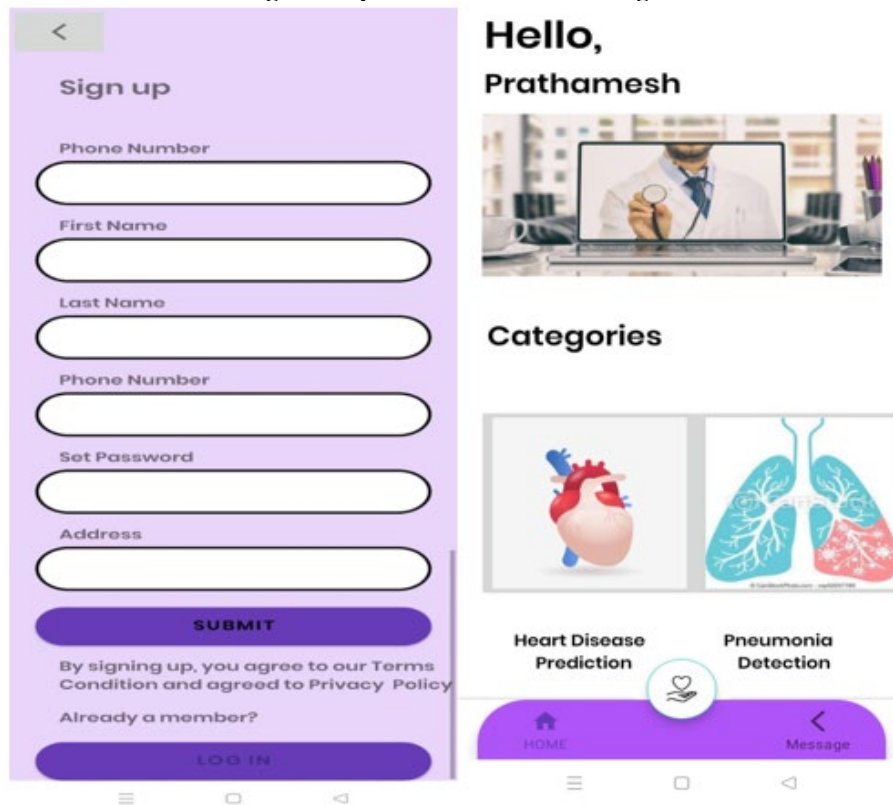


Figure 8 Registration page, Home Screen

### Heart Disease Prediction



Please Enter the following details :

Age:

Sex:

Chest pain type:

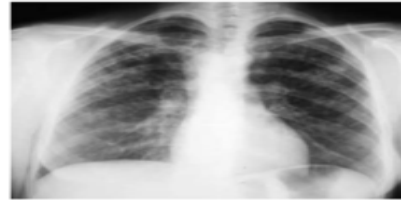
trestbps(mm Hg):

Chol(mg/dl):

fbs:

**SUBMIT**

### Phenomena Detection



Upload your chest X-ray Image

Fill out the following information

Shortness of Breath?  Yes  No

Fever, sweating and shaking chills?  Yes  No

Chest pain when you breathe or cough?  Yes  No

Fatigue or Nausea?  Yes  No

**SUBMIT**



Figure 9. Detection Pages

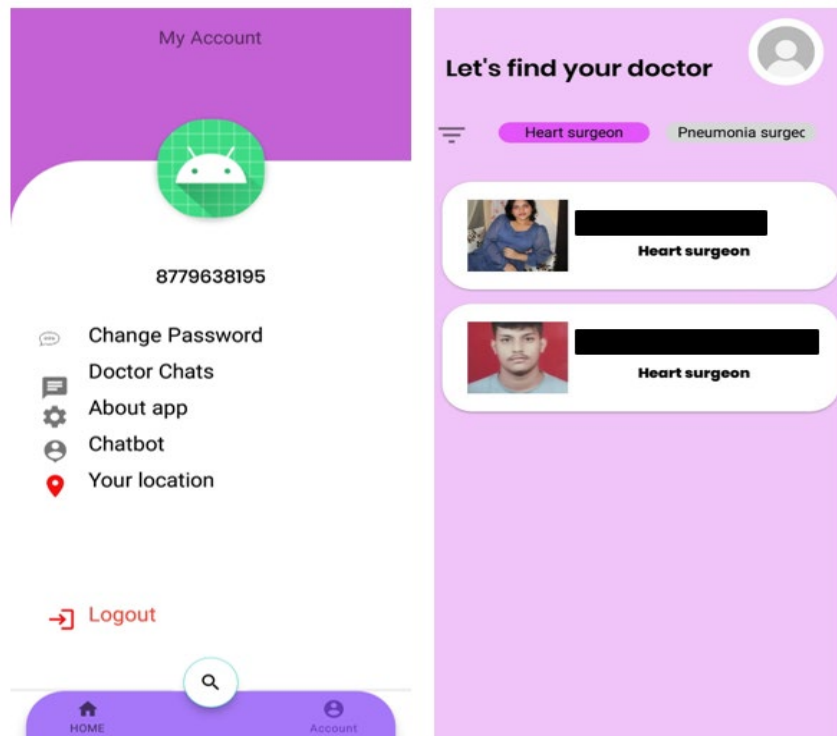


Figure 10. Account Info page, Contact Doctor Page





## VII. CONCLUSION

Heart Disease conditions are increasing rapidly with all the commotions and ignored health factors. Also Pneumonia is a life-threatening condition that affects the lungs. Prediction of any disease helps prevent or cure it faster. In order to provide a solution this system proposes the detection system where the model can correctly identify those with pneumonia and possibility of having a heart disease while also avoiding false positives. Further this system can be improvised to find the different types of heart diseases like coronary heart diseases and Conjunction heart diseases. Potential benefits of this system are significant and could lead to better patient health outcomes and improved healthcare delivery. In the proposed system Random Forest gives 100% accuracy for both pneumonia detection and heart Disease Detection. Mobile net gives 85% accuracy for x-ray detection in pneumonia detection.

## VIII. REFERENCES

- [1]. Elshennawy, Nada M, and Dina M Ibrahim. "Deep-Pneumonia Framework Using Deep Learning Models Based on Chest X-Ray Images." *Diagnostics* (Basel, Switzerland) vol. 10,9 649. 28 Aug. 2020, doi:10.3390/diagnostics10090649
- [2]. Detrano R, Janosi A, Steinbrunn W, Pfisterer M, Schmid JJ, Sandhu S, Guppy KH, Lee S, Froelicher V. International application of a new probability algorithm for the diagnosis of coronary artery disease. *Am J Cardiol.* 1989 Aug 1;64(5):304-10. doi: 10.1016/0002-9149(89)90524-9. PMID: 2756873..
- [3]. Victor Chang, Vallabhanent Rupa Bhavani, Ariel Qianwen Xu, MA Hossain, "An artificial intelligence model for heart disease detection using machine learning algorithms,," *ELSEVIER*, vol. 02, November 2022.
- [4]. Syed Nawaz Pasha, Dadi Ramesh, Sallauddin Mohmmad, A. Harshavardhan and Shabana "Cardiovascular disease prediction using deep learning," *IOP Conf. Series: Materials Science and Engineering*, p. 7, 2020.
- [5]. D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan and A. Mittal, "Pneumonia Detection Using CNN based Feature Extraction," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-7, doi: 10.1109/ICECCT.2019.8869364.
- [6]. Kundu R, Das R, Geem ZW, Han GT, Sarkar R. Pneumonia detection in chest X-ray images using an ensemble of deep learning models. *PLoS One.* 2021 Sep 7;16(9):e0256630. doi: 10.1371/journal.pone.0256630. PMID: 34492046; PMCID: PMC8423280.
- [7]. Liang G, Zheng L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Comput Methods Programs Biomed.* 2020 Apr;187:104964. doi: 10.1016/j.cmpb.2019.06.023. Epub 2019 Jun 26. PMID: 31262537.
- [8]. Ibrahim AU, Ozsoz M, Serte S, Al-Turjman F, Yakoi PS. Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19. *Cognit Comput.* 2021 Jan 4:1-13. doi: 10.1007/s12559-020-09787-5. Epub ahead of print. PMID: 33425044; PMCID: PMC7781428.
- [9]. J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.
- [10]. Kundu R, Basak H, Singh PK, Ahmadian A, Ferrara M, Sarkar R. Fuzzy rank-based fusion of CNN models using Gompertz function for screening COVID-19 CT-scans. *Sci Rep.* 2021 Jul 8;11(1):14133. doi: 10.1038/s41598-021-93658-y. PMID: 34238992; PMCID: PMC8266871.
- [11]. Manna, Ankur & Kundu, Rohit & Kaplun, Dmitry & Sinitca, Aleksandr & Sarkar, Ram. (2021). A fuzzy rank-based ensemble of CNN models for classification of cervical cytology. *Scientific Reports.* 11. 14538. doi: 10.1038/s41598-021-93783-8.
- [12]. D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan and A. Mittal, "Pneumonia Detection Using CNN based Feature Extraction," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-7, doi: 10.1109/ICECCT.2019.8869364.
- [13]. Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam, Mohammad Hasan Imam, "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System," *World Journal of Engineering and Technology.*, pp. 854-873, 2018.
- [14]. Yar Muhammad, Muhammad Tahir, Maqsood Hayat & Kil To Chong, "Early and accurate detection and diagnosis of heart disease using intelligent computational models," *Sci Rep* , p. 10, 2020.
- [15]. M Sheshikala, D Kothandaraman , R V. P. and G Roopa,2019. Natural language processing and machine learning classifier used for detecting the author of the sentence *International Journal of Recent Technology and Engineering* 8(3),936-939.