A PROJECT REPORT ON

Heart Disease Detection and Patient's Sickness Prediction System

SUBMITTED TO MIT SCHOOL OF ENGINEERING
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE

BACHELOR OF TECHNOLOGY
(Computer Science & Engineering)

BY

Sakshi Singh	MITU19BTCS0271
Tanmay Patil	MITU20BTCSD043
Gurjeet Singh	MITU19BTCS0268
Chaitanya Pandey	MITU19BTCS0197

Under The Guidance of

Prof. Shahin Shoukat Makubhai

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
MIT School of Engineering
MIT Art, Design and Technology University
Rajbaug Campus, Loni-Kalbhor, Pune 412201

MIT SCHOOL OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MIT ART, DESIGN AND TECHNOLOGY UNIVERSITY,
RAJBAUG CAMPUS, LONI-KALBHOR, PUNE 412201

CERTIFICATE

This is to certify that the Project Entitled

Heart Disease Detection and Patient's Sickness Prediction System

Submitted by

Sakshi Singh	MITU19BTCS0271
Tanmay Patil	MITU20BTCSD043
Gurjeet Singh	MITU19BTCS0268
Chaitanya Pandey	MITU19BTCS0197

is a bonafide work carried out by them under the supervision of Prof. Shahin Shoukat Makubhai and it is submitted towards the fulfillment of the requirement of MIT ADT University, Pune for the award of the degree of Bachelor of Technology (Computer Science and Engineering).

Prof. Shahin Shoukat Makubhai Internal Guide Department of CSE	Prof. Dr. Shraddha Phansalkar H.O.D Department of CSE	Dr. Rajneeshkaur Sachdeo Director MIT SoC

DECLARATION

We, the team members

Sakshi Singh	MITU19BTCS0271
Tanmay Patil	MITU20BTCSD043
Gurjeet Singh	MITU19BTCS0268
Chaitanya Pandey	MITU19BTCS0197

Hereby declare that the project work incorporated in the present project entitled **Heart Disease Detection and Patient's Sickness Prediction System** is original work. This work (in part or in full) has not been submitted to any University for the award or a Degree or a Diploma.We have properly acknowledged the material collected from secondary sources wherever required.We solely own the responsibility for the originality of the entire content.

Date: Name & Signature of the Team Members

Sakshi Singh

Tanmay Patil

Gurjeet Singh

Chaitanya Pandey

Prof. Shahin Shoukat Makubhai

Seal/Stamp of the college
Place:
Date:

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MIT SCHOOL OF ENGINEERING
RAJBAUG, LONI KALBHOR
PUNE – 412201

EXAMINER'S APPROVAL CERTIFICATE

1 The project report entitled **Heart Disease Detection and Patient's Sickness Prediction System** submitted by Sakshi Singh (MITU19BTCS0271), Tanmay Patil (MITU20BTCSD043), Gurjeet Singh (MITU19BTCS0268) and Chaitanya Pandey (MITU19BTCS0197) in fulfillment for the award of the degree of Bachelor of Technology (Computer Science & Engineering) during the academic year 2022-23, of MIT-ADT University, MIT School of Engineering, Pune, is hereby approved.

Examiners

Examiner 1 Name and Signature :

Examiner 2 Name and Signature:

Acknowledgments

It gives us great pleasure in presenting the project report on ***'Heart Disease Detection and Patient's Sickness Prediction System'****.*

We would like to take this opportunity to thank my internal guide ***Prof. Shahin Shoukat Makubhai*** *for giving me all the help and guidance I needed. I am really grateful to them for their kind support. Their valuable suggestions were very helpful.*

We are also grateful to ***Prof. Dr. Shraddha Phansalkar****, Head of Computer Science & Engineering indispensable support, suggestions.*

We are also grateful to our technology experts ***Prof. Reena Gunjan*** *and* ***Prof. Suvarna Pawar*** *for their help, support and suggestion.*

Sakshi Singh
Tanmay Patil
Gurjeet Singh
Chaitanya Pandey
(B.Tech. Computer Science & Engineering)

Abstract

This report delves into research conducted on Heart Disease Detection and Patient's Sickness Prediction System, which is a crucial topic given the increasing incidence of heart diseases. Early detection of such diseases is of utmost importance and requires precise and efficient diagnosis. The project focuses on predicting the likelihood of a patient having a heart disease based on their medical history and various medical attributes, using machine learning algorithms such as logistic regression. The proposed model demonstrated strong predictive capabilities and outperformed previous classifiers such as Naive Bayes. The Disease Prediction system, which is based on predictive modeling, utilizes the symptoms provided by the user as input to predict the probability of the disease using the Random Forest Classifier. The project aims to use various supervised machine learning algorithms in disease prediction and heart disease detection through symptoms and medical details as input, demonstrating the potential for these algorithms to aid in the early detection of high-risk diseases.

Contents

List of Figures

CHAPTER 1

INTRODUCTION

1.1 RELEVANCE

- Due to the large volume of data, healthcare professionals face challenges in accurately analyzing symptoms and detecting disease at an early stage. However, supervised machine learning (ML) algorithms have demonstrated great potential over traditional disease diagnosis systems and helping specialists identify high-risk diseases in a timely manner.

1.2 MOTIVATION OF THE PROJECT

- Identifying and predicting diseases is critical to preventing their severity and fatal consequences. In India, the majority of deaths are due to heart attacks, especially among elderly people affected by cardiovascular diseases. To solve this problem, our prediction system uses various machine learning algorithms to predict the risk level of these diseases. The advent of artificial intelligence (AI) has enabled computer systems to act intelligently, like humans, by perceiving, thinking, and making decisions. AI encompasses multidisciplinary areas such as machine learning, computer vision, deep learning, and natural language processing. By applying optimization, statistical, and probabilistic techniques to past data, ML algorithms can learn and assist healthcare professionals in decision making, contributing to treat the patient appropriately.

1.3 PROBLEM STATEMENT

- Developing a system to detect heart disease and predict disease for patients with help with medical insights and initial symptoms using machine learning algorithms with highest possible accuracy.

1.4 OBJECTIVES

goal is to use supervised ML algorithms to improve healthcare through accuracy and early detection of diseases that become harmful at a later stage. ML models will be used to predict diseases ranging from common to severe, just

to name a few that are located in the heart, kidney, breast and brain. For disease prediction, we used Random Forest Classification and logistic regression to detect heart disease.

1.5 SCOPE

- The incidence of heart disease is steadily increasing around the world, including in our country. To address this, we used logistic regression to detect and predict heart disease, achieving 81 accuracy. Although this accuracy is satisfactory, it can still be improved by exploring alternative methods for attribute selection and increasing the size of the dataset to avoid overfitting and improve performance. Alternatively, we may consider leveraging the power of deep learning algorithms to achieve even more accurate results in the future.

1.6 ORGANIZATION OF THE REPORT

- This report covers the methodology used to develop the prediction system, as well as the technologies and tools used to create it. In addition, it includes relevant research papers that were reviewed during our investigation to explore the different systems that exist. The future prospects of the project are also discussed in this report. Finally, the conclusion section provides a summary of the entire report.

CHAPTER 2

LITERATURE SURVEY

2.1 RELATED WORK

Much research and development efforts have been made to develop disease detection and prediction models. Commonly used supervised machine learning algorithms include Naive Bayes (NB), Decision Trees (DT), and K-Nearest Neighbors (KNN). According to the study results, Support Vector Machine (SVM) is suitable for detecting kidney disease and Parkinson's disease, while Logistic Regression (LR) is highly effective in predicting heart disease. In addition, Random Forest (RF) and Convolutional Neural Networks (CNN) successfully predicted breast and common diseases, respectively. Despite advances in computing, doctors still need technology for a variety of purposes, such as surgical imaging and X-ray imaging. However, technology has yet to keep up with doctors' levels of knowledge and experience. This is because many different factors, such as medical records, weather conditions, air, blood pressure, and others, must be considered in order to understand the entire process of labor. Medical decision support systems can help address this challenge by helping doctors make the right decisions. We used machine learning to analyze the hospital's comprehensive data, allowing us to build models that can quickly analyze data and deliver results faster. By leveraging machine learning, physicians can make critical decisions regarding their patient's diagnosis and treatment options, thereby improving patient care services. . The healthcare industry is a great example of how machine learning is revolutionizing the medical field. [7]

2.2 COMPARISON OF EXISTING WORK

To begin with, the ML algorithms used by most systems tend to have lower accuracy when compared to the ones we have used. Decision Tree Classifier and KNN are commonly used in existing systems, but they do not offer the same level of accuracy as our algorithms. Another reason for this difference in accuracy is the use of complex deep learning algorithms, which require a large amount of data to predict a disease, such as medical images taken from different angles, which in turn require medical expertise and other medical diagnosis details.[5]

2.3 GAP IDENTIFICATION

We compared all the models for classification and prediction and the one with the highest accuracy is chosen for the further development of the project. We are using more than one machine learning algorithms to achieve the highest accuracy

CHAPTER 3

SOFTWARE REQUIREMENT SPECIFICATION

3.1 INTRODUCTION

Due to the substantial amounts of data, medical doctors are facing challenges to analyze symptoms accurately and identify diseases at an early stage. However, Supervised ML algorithms have showcased significant potential in surpassing standard systems for disease diagnosis and aiding medical experts in the early detection of high-risk diseases.

3.2 PURPOSE AND SCOPE OF DOCUMENT

The goal of the project is to use medical data to help doctors and patients identify diseases a person may have based on symptoms and medical information provided.

3.3 GENERAL DESCRIPTION

The system's disease prediction module is designed to predict disease based on symptoms and is intended for end users. Machine learning is leveraged by the system to make predictions, with a Random Forest classification algorithm used to predict the disease. The heart disease detection module, on the other hand, looks at medical attributes such as gender, age, chest pain, and fasting blood sugar to determine if a patient is susceptible to cardiovascular disease. To do this, a logistic regression algorithm is used.

3.4 PRODUCT PERSPECTIVE

This product is for detecting heart disease and predicting any disease in its early stage.

3.5 PRODUCT FEATURES

- High accuracy.
- Integration of two types of disease prediction.
- Easy to use.

- User friendly interface (simple and attractive UI).
- Reduce the cost of medical test.

3.6 ASSUMPTION,DEPENDENCIES CONSTRAINTS

We have selected a dataset from the UCI repository containing patients' medical history and attributes to predict the likelihood of a patient having a heart disease or any other disease. We use various medical attributes of a patient to classify and predict the presence of a heart disease or any other disease. Two algorithms, namely Logistic Regression and Random Forest, are currently being used to train these medical attributes and sickness attributes.

3.7 DOMAIN REQUIREMENTS

To implement the project, it is necessary to have Python or Anaconda, Streamlit Server, Sypder and Jupyter Notebook installed on the computer. Additionally, a graphics card with at least 2GB of memory is required.

3.8 INTERFACE REQUIREMENTS

Currently in our project we have used:

- Streamlit : It is an open- source app framework for machine learning and data science teams. We have used this for the frontend development and it also act as a server.[2]
- Jupyter : Jupyter notebook for training and testing data and for creating models.
- Spyder : Spyder for implementing the logic.

3.9 DESIGN CONSTRAINTS

The quality of the data used for training the machine learning models is crucial for accurate predictions. The data must be reliable, accurate, and representative of the

target population.The machine learning models must be designed to be interpretable, so that clinicians and patients can understand the factors that contribute to the predictions. This is important for building trust in the models and ensuring that they are used appropriately.Bias in the data can lead to biased predictions. It is important to ensure that the training data is diverse and unbiased, so that the model can accurately predict heart disease risk for all individuals.

3.10 OVERVIEW OF RESPONSIBILITIES OF DEVELOPER

The first step involves data collection, followed by the extraction of significant values in the second stage. The third stage is the data preprocessing stage, which involves exploring the data. Data preprocessing includes dealing with missing values, cleaning the data, and normalizing the data based on the algorithms used. Once the data is preprocessed, feature selection is performed. After feature selection, a classifier is used to classify the preprocessed data. In the proposed model, the classifiers used are Logistic Regression and Random Forest Classifier. Finally, the proposed model is evaluated based on its accuracy and performance using various performance metrics.

CHAPTER 4

PROJECT DESIGN AND IMPLEMENTATION

4.1 ARCHITECTURAL DIAGRAM

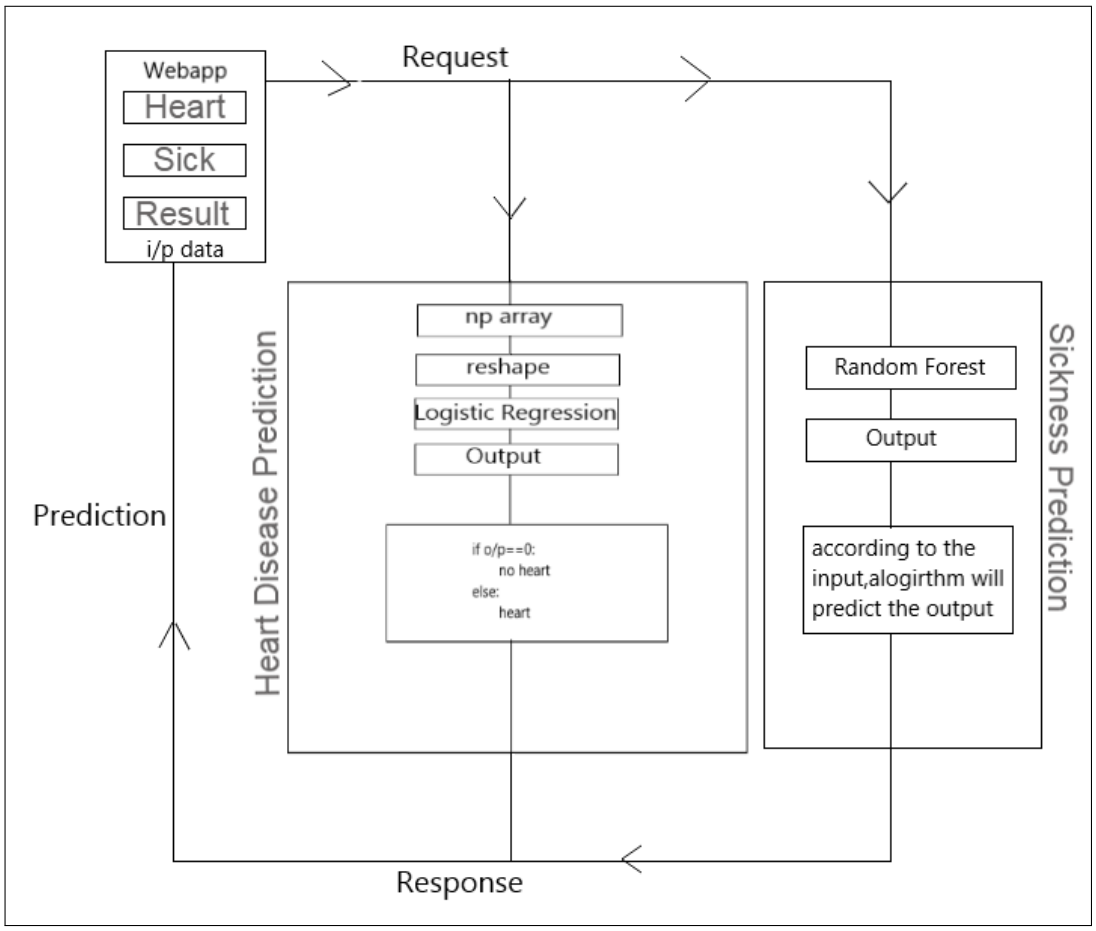


Figure 4.1: Architecture diagram

4.2 METHODOLOGY

Heart Disease Prediction

The methodology for heart disease prediction using machine learning typically involves the following steps:

- Data collection: Collecting the relevant data for heart disease prediction, which could include medical history, patient attributes such as age, gender, etc., and any other relevant information.
- Data preprocessing: Cleaning and processing the collected data, dealing with missing values, and normalizing the data to make it suitable for use with ma-

chine learning algorithms.

- Feature selection: Selecting the most relevant features that are likely to contribute to the accurate prediction of heart disease.
- Model training: Training a machine learning model on the preprocessed and feature-selected data. The model is typically trained on a portion of the data, known as the training set.
- Model evaluation: Evaluating the trained model using various performance metrics such as accuracy, precision, recall, and F1-score, among others. The model is evaluated on another portion of the data, known as the test set.
- Model optimization: Tuning the model's hyperparameters and optimizing the performance of the model by selecting the best combination of parameters and features.
- Deployment: Deploying the trained and optimized model in the clinical setting to assist doctors and medical professionals in the accurate prediction of heart disease in patients.

S.No	Attributes	Value type
1.	age	Numerical
2.	sex	Nominal
3.	cp	Nominal
4.	trestbps	Numerical
5.	cho	Numerical
6.	fbs	Nominal
7.	restecg	Nominal
8.	thalach	Numerical
9.	ca	Nominal
10.	target	Nominal

Figure 4.2: Heart Disease Dataset

Logistic Regression is a Machine Learning Algorithm that predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.Logistic regression involves the following data and operations:[1]

- Gather columns.
- Splitting Data.
- Normalization.
- Fitting into Model.
- Prediction.
- Model Evaluation.

Requirements

The following hardware and software are required for experimental analysis and heart disease prediction results. A system with an i5 quad-core processor, 6 GB of RAM, and software packages such as pandas, python, SciPy, StatsModels, and Matplotlib is required. Test analysis was performed in the Jupyter and Spyder web application environments. The analysis is done at two levels: first, the dataset is cleaned using pandas tool, and second, the cleaned data is passed through classifiers to predict heart disease.

Patient's Sickness Prediction

An Excel sheet was created from an open-source dataset, listing all the symptoms for respective diseases.The dataset contained approximately 230 diseases with over 1000 unique symptoms. The symptom of an individual were used as inputs for various machine learning algorithms.

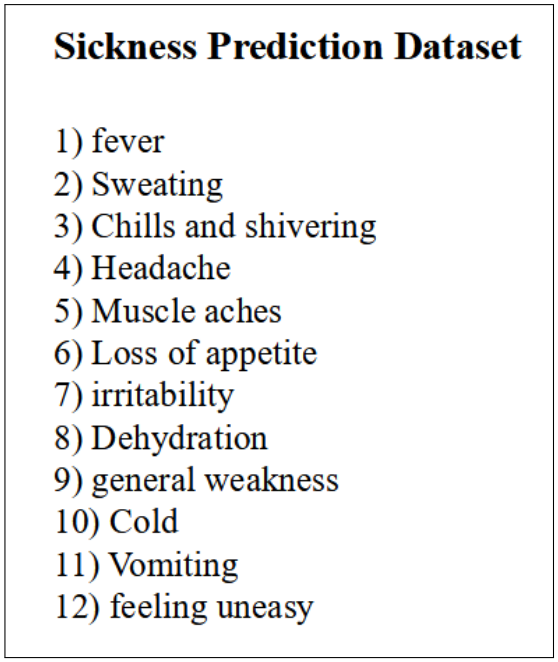


Figure 4.3: Sickness Dataset

Random Forest Classifier

Random forest, as its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model' prediction.[4]

Working

- The website will collect input data from users and utilize a training dataset to determine the outcome.
- When the user clicks the result button, it will trigger a request to the Streamlit server containing their inputs, which the server will subsequently restructure.
- Subsequently, the inputs will be fed into a trained model.
- The model will analyze the provided data and generate a forecasted output.
- The server will transmit the forecasted output to the web application as a response, and the web application will exhibit the projected outcome.

Such is the methodology we shall employ to predict heart disease and patients' illnesses using trained models.

4.3 TOOLS AND TECHNOLOGIES USED

- Python: It is utilized as the back-end programming language to implement the prediction models.
- Streamlit: It is an open-source Python library used for building interactive web applications for data science and machine learning projects. It simplifies the process of creating a web application by allowing developers to create a user interface with simple Python scripts. We have used this for the frontend development
- Jupyter Notebook: It is often used in data science and machine learning projects for data exploration, analysis, visualization and for training and testing data and for creating models.[10]
- Spyder: It is an open-source integrated development environment (IDE) for scientific computing in Python. It provides a powerful editor for writing code, a console for executing code, and tools for debugging and profiling code.
- Libraries : NumPy to deal with huge data-set in numerical form, Pandas to analyze the data, Pickle for creating model.sav files, Stream-lit and streamlitoptionmenu for creating website, Sklearn for traintestsplit , logistics regression for accuracy and Sklearn for random forest classifier

4.4 USAGE SCENARIO

Heart disease Prediction

- Early Detection: It can predict the likelihood of heart disease in a patient before the disease becomes severe. This can help in early detection and timely intervention to prevent or manage the disease.

- Risk Assessment: Healthcare providers can use model to assess the risk of developing heart disease in patients based on their medical history, lifestyle factors, and other risk factors. This can help in developing personalized treatment plans and preventive strategies.

Patient's Sickness Prediction

Suppose a provider wants to predict the likelihood of a patient being diagnosed with a certain disease based on their current symptoms. The provider collects patient's current symptoms.[3]

It can be used to predict the likelihood of a patient being diagnosed with a certain disease based on their current symptoms. This information can be used by healthcare professionals to identify high-risk patients and provide early interventions or treatment plans. Additionally, the model can be continuously updated with new data to improve its accuracy over time.

4.4.1 Use-cases

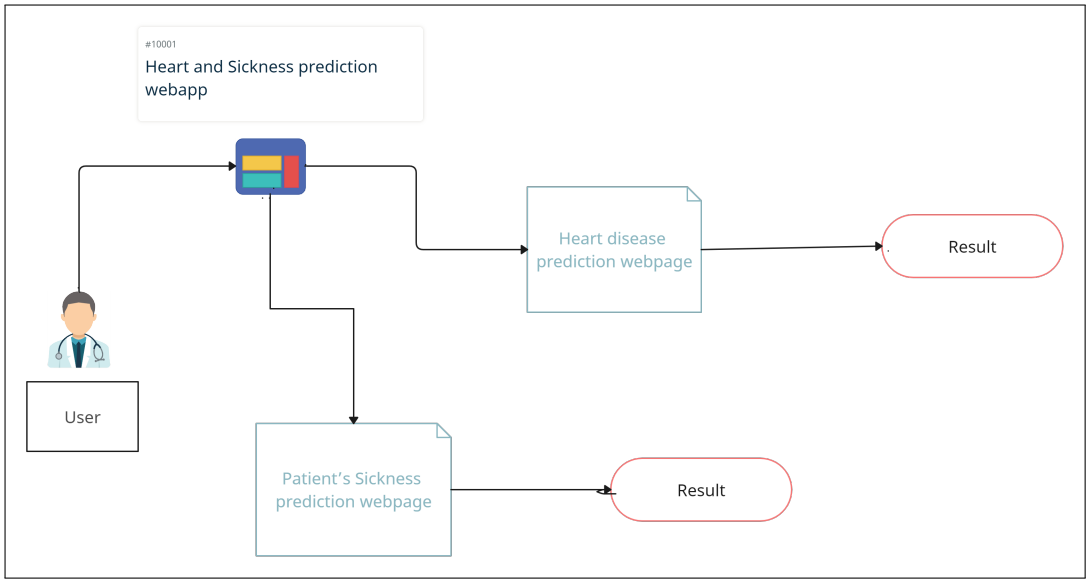


Figure 4.4: Webapp Use Case Diagram

CHAPTER 5

RESULTS

Our web-app is designed to predict both the sickness and heart disease of patients. The system is built using machine learning algorithms such as Random Forest Classifier and Logistic Regression.

Random Forest Classifier is used for predicting various types of diseases by inputting the symptoms of the patients. The system has an accuracy of 95 percent which makes it highly reliable.

For predicting heart disease, we have used Logistic Regression which is known for its effectiveness in training models with large datasets and getting a discrete value as an output. In comparison to other algorithms, the accuracy of our heart disease prediction model is better, especially when patients provide some essential medical details such as blood pressure, cholesterol, diabetes, LDL, etc.[6]

Our web-app is user-friendly and easy to use, doctors can input the medical details and symptoms of patients and the web-app will provide the predictions accordingly. Additionally, the it provides results in a timely manner which makes it highly useful for medical professionals and patients alike.

CHAPTER 6

SYSTEM TESTING

We have used ITERATIVE MODEL. In this type of model possible key steps in the process start with a simple implementation of a subset of the system requirements and iteratively enhance the evolving versions until the full system is implemented.

6.1 WHITE BOX TESTING

This type of testing ensures that:[9]

- All independent paths have been exercised at least once.
- All logical decisions have been exercised on their true and false sides.
- All loops are executed at their boundaries and within their operational bounds.
- All data structures have been exercised to assure their validity.

6.2 CONDITIONAL TESTING

In this part of the testing each of the conditions were tested to both true and false aspects. And all the resulting paths were tested. So that each path that may be generate on particular condition is traced to uncover any possible errors.

6.3 DATA FLOW TESTING

This type of testing selects the path of the program according to the location of definition and use of variables. This kind of testing was used only when some local variable was declared. The definition use chain method was used in this type of testing.

6.4 TEST RESULTS

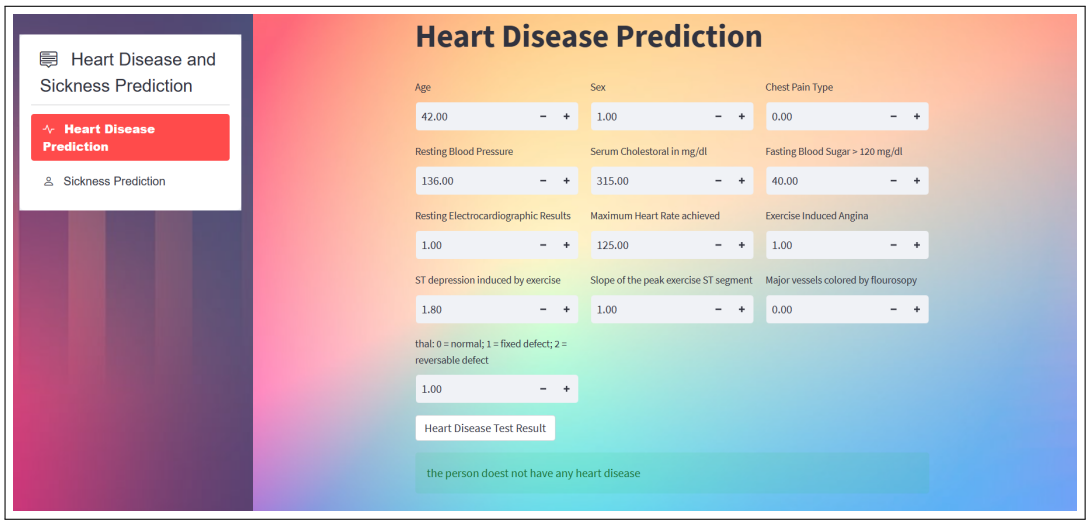


Figure 6.1: Patient With Healthy Heart

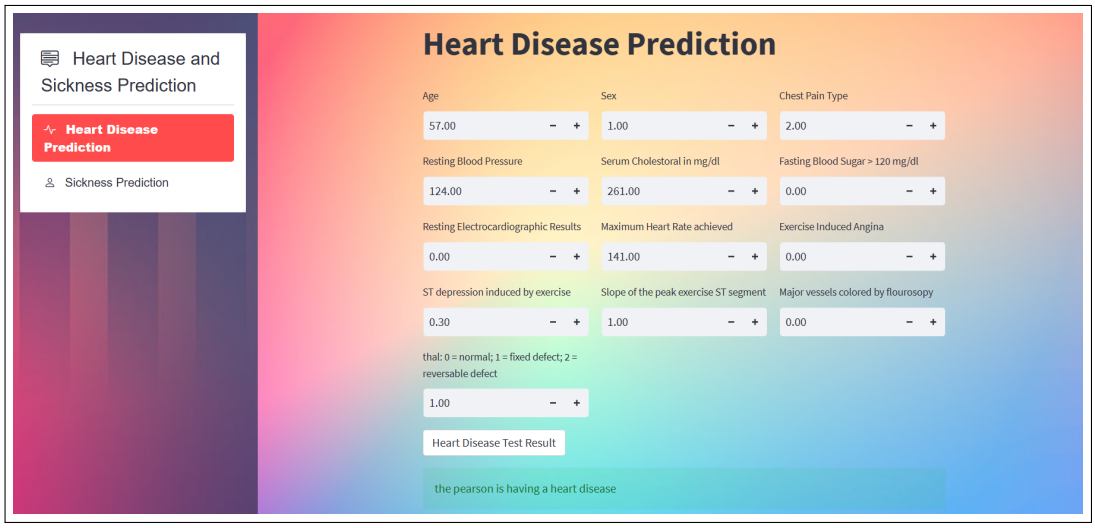


Figure 6.2: Patient With Unhealthy Heart

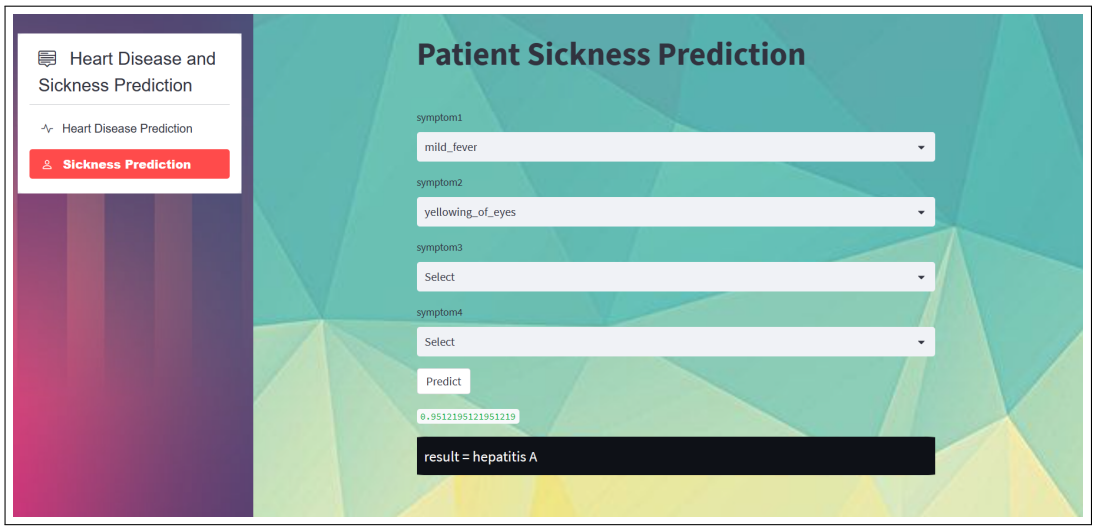


Figure 6.3: Patient's Sickness Prediction

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

The project aids in disease prediction at both personal and public levels. It provides information about the potential risk factors, thus reducing the need for unnecessary tests and associated costs. Moreover, the system can extract and analyze data from patients to reveal hidden patterns, which can be useful for future research and medical advancements.

During the training and testing phase of our project, we evaluated multiple machine learning models for predicting common diseases. The Random Forest Classifier model emerged as the most effective and accurate model with a precision of 95 percent. For the prediction and detection of heart disease, we utilized Logistic Regression which yielded an accuracy of 81 percent. There are various possibilities to further improve the accuracy of this model in the future.

The Heart disease detection model using Logistic Regression and the sickness prediction model using Random Forest Classifier showed superior results. These algorithms are not only more accurate but also cost-efficient and faster compared to the algorithms used by previous researchers. The Random Forest Classifier achieved a maximum accuracy of 95 percent, and Logistic Regression accuracy was 81 percent, which is either greater or almost equal to the accuracy's obtained from previous research. We can conclude that our accuracy has improved by using additional medical attributes from the dataset we used and can also improve more in future updates.

7.1 CONCLUSION

We worked as a team, and gained some experience on how professional programmers work in the industry. There is always room for improvement, and this application we created can also be improved. This is especially because we had to create it within a limited time due to other projects, quizzes and Exams. The main goal of our system is to help detect any illness a person may be carrying early, allowing doctors to keep the treatment on track. This is especially helpful for people with heart problems because they can determine if they are at risk for heart disease. By using our system, individuals can benefit from early detection and treatment, which ultimately leads to better health outcomes.

7.2 FUTURE SCOPE

There are several ways in which we can further enhance the accuracy of our system. For instance, we can explore the utilization of deep learning algorithms,[8] consider alternative techniques for attribute selection, and potentially expand the size of the dataset in the future to address issues related to overfitting and to improve overall performance.

CHAPTER 8

BIBLIOGRAPHY

[1] Lynne Connelly. Logistic regression. *Medsurg Nursing*, 29(5):353–354, 2020.

[2] Laxmi Deepthi Gopisetti, Srinivas Karthik Lambavai Kummera, Sai Rohan Pattamsetti, Sneha Kuna, Niharika Parsi, and Hari Priya Kodali. Multiple disease prediction system using machine learning and streamlit. In *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 923–931. IEEE, 2023.

[3] P Hamsagayathri and S Vigneshwaran. Symptoms based disease prediction using machine learning techniques. In *2021 Third international conference on intelligent communication technologies and virtual mobile networks (ICICV)*, pages 747–752. IEEE, 2021.

[4] Aakash Parmar, Rakesh Katariya, and Vatsal Patel. A review on random forest: An ensemble classifier. In *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*, pages 758–763. Springer, 2019.

[5] Anjan Nikhil Repaka, Sai Deepak Ravikanti, and Ramya G Franklin. Design and implementing heart disease prediction using naives bayesian. In *2019 3rd International conference on trends in electronics and informatics (ICOEI)*, pages 292–297. IEEE, 2019.

[6] Montu Saw, Tarun Saxena, Sanjana Kaithwas, Rahul Yadav, and Nidhi Lal. Estimation of prediction for getting heart disease using logistic regression model of machine learning. In *2020 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–6. IEEE, 2020.

[7] Devansh Shah, Samir Patel, and Santosh Kumar Bharti. Heart disease prediction using machine learning techniques. *SN Computer Science*, 1:1–6, 2020.

[8] Sumit Sharma and Mahesh Parmar. Heart diseases prediction using deep learning neural network model. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9(3):2244–2248, 2020.

[9] Akanksha Verma, Amita Khatana, and Sarika Chaudhary. A comparative study of black box testing and white box testing. *International Journal of Computer Sciences and Engineering*, 5(12):301–304, 2017.

[10] Jiawei Wang, Li Li, and Andreas Zeller. Better code, better sharing: on the need of analyzing jupyter notebooks. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: New Ideas and Emerging Results*, pages 53–56, 2020.

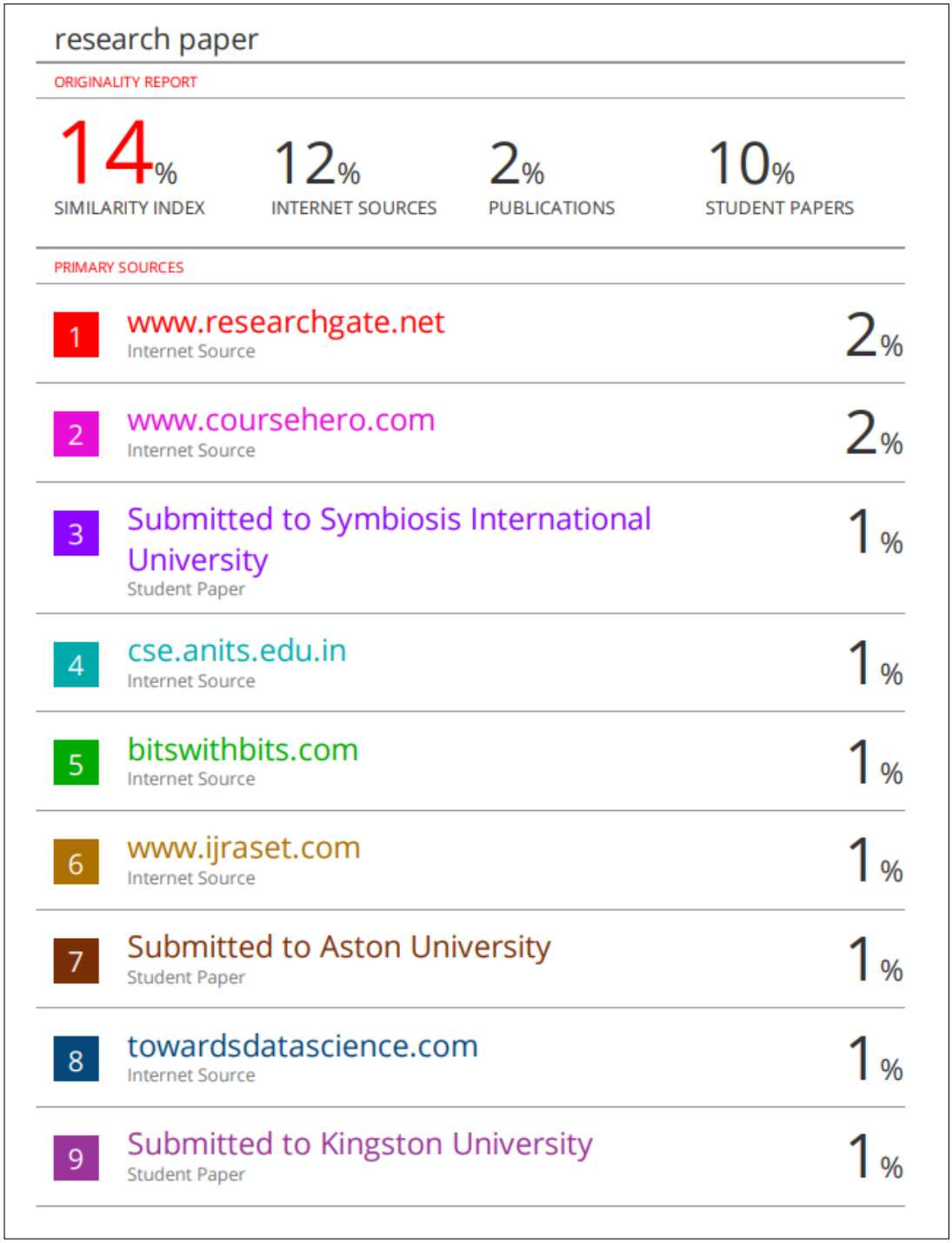
ANNEXURE A

LIST OF PUBLICATIONS AND RESEARCH PAPER

1. Paper Title : Heart Disease Detection and Patient's Sickness Prediction System
2. Name of the Conference/Journal where paper submitted : International Organization of Scientific Research
3. Paper accepted/rejected : Accepted
4. Review comments by reviewer : Quality of Manuscript is good.
5. Corrective actions if any : None

ANNEXURE B

PLAGIARISM REPORT



research paper

ORIGINALITY REPORT

14%	12%	2%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	2%
2	www.coursehero.com Internet Source	2%
3	Submitted to Symbiosis International University Student Paper	1%
4	cse.anits.edu.in Internet Source	1%
5	bitswithbits.com Internet Source	1%
6	www.ijraset.com Internet Source	1%
7	Submitted to Aston University Student Paper	1%
8	towardsdatascience.com Internet Source	1%
9	Submitted to Kingston University Student Paper	1%