A PROJECT REPORT ON

HEART DISEASE PREDICTION SYSTEM USING MACHINE LEARNING

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

BACHELOR OF ENGINEERING
IN
COMPUTER ENGINEERING

SUBMITTED BY

Khan Moin Hayat	Exam No: B190844232
Shaikh Ayaan Shakil	Exam No: B190844237
Kanade Samruddhi Santosh	Exam No: B190844230

Under The Guidance of
Prof. K. D. Dere

DEPARTMENT OF COMPUTER ENGINEERING

JCEI'S JAIHIND COLLEGE OF ENGINEERING
KURAN , PINCODE 410511
SAVITRIBAI PHULE PUNE UNIVERSITY

2023-2024

CERTIFICATE

This is to certify that the project phase-2 report entitled

Heart Disease Prediction System Using Machine Learning

Khan Moin Hayat	Exam No: B190844232
Shaikh Ayaan Shakil	Exam No: B190844237
Kanade Samruddhi Santosh	Exam No: B190844230

is a bonafide student of this institute and the work has been carried out by him/her under the supervision of **Prof. K. D. Dere** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of Bachelor of Engineering (Computer Engineering).

Prof. K. D. Dere	**Dr. A. A. Khatri**
Guide Name	H.O.D
Dept. of Computer Engg	Computer Engg. Department

Dr. S .D. Gunjal	**Dr. D. J. Garkal**
Project Co-ordinator	Principal
Director of J.C.E.I's	Jaihind College of Engineering

Abstract

Hand and face gestures will be recognized using computer vision techniques such as Convolutional Neural Networks (CNNs) for image analysis. This will enable the system to accurately interpret user gestures and map them to specific in-game actions.Hand gesture recognition is part of Human- Computer Interaction as a place for the user interface to be presented.Various machine learning algorithms, including but not limited to, deep learning models (CNNs and Recurrent Neural Networks), decision trees, and ensemble methods, will be explored to classify and interpret the extracted features, translating them into meaningful game commands. Low latency processing will be a priority to ensure that the system responds quickly to user gestures, maintaining a seamless gaming experience. In summary, this research project aims to develop an innovative gesture-based game control model that leverages hand and face gestures, integrating computer vision and machine learning techniques to provide an immersive and responsive gaming experience. This model holds the potential to reshape the future of gaming by offering a more natural and intuitive means of interaction, enhancing user engagement and enjoyment in the gaming world. objects.

Keywords:-Gaming , Gesture-Based Gaming, Hand Gestures, Face Gestures , Gesture Recognition, Computer Vision, Machine Learning

Acknowledgments

With immense pleasure, we present the Project Phase-2 report as part of the curriculum of the B.E Computer Engineering. We wish to thank and express deep sense of gratitude to our guide **Prof. K. D. Dere** for his guidance, inspiration in project work.

We also wish to thank our Project Coordinator **Dr. S. D. Gunjal** for encouragement and he patiently discussed ideas with us and gave indispensable suggestions.

We are grateful to **Dr. A.A. Khatri** , Head of Department of Computer Engineering for always being ready to help with the most diverse problems that we have encountered along the way.

We are also grateful to Principal of Jaihind College of Engineering **Dr. D. J .Garkal** for always being ready for guidance and support.

Khan Moin Hayat	Exam No: B190844232
Shaikh Ayaan Shakil	Exam No: B190844237
Kanade Samruddhi Santosh	Exam No: B190844230

List of Figures

INDEX

Chapter 1

Synopsis

1.1 PROJECT TITLE

HEART DISEASE PREDICTION SYSTEM CONTROL USING MACHINE LEARNING

1.2 PROJECT OPTION

Internal project

1.3 INTERNAL GUIDE

Prof. K. D. Dere

1.4 ABSTRACT

Heart disease prediction is very essential in today's environment, various researches has already done to predict heart disease from large dataset. IoT environment basically generate data from different sensors and predict the disease possibility accordingly. Various synthetic data sets content different body parameters which are extracted by specific sensor values, the major role played by machine learning algorithm. In this research we propose heart disease prediction with the combination of IoT and machine learning approach, the IoT environment has established to extract the data from real-time Body Sensor Network (BSN) with intermediate sensing Sys-

tem and store data in the cloud server adequately. Such audit data has considered synthetic information which is basically used to predict heart disease possibility. In this research, we illustrate various machine learning algorithms as well as some deep learning algorithms to achieve drastic supervision for disease prediction. The experimental analysis shows the effectiveness of proposed deep learning classification algorithms over the classical machine learning algorithms

1.5 TECHNICAL KEYWORDS(AS PER ACM KEYWORDS)

1. Euclidean distance
2. Nearest neighbors
3. Sigmoid function
4. Decision trees
5. Ensemble learning
6. Feature importance

1.6 PROBLEM STATEMENT

The main challenge in heart disease is detection. There are tools that can predict heartdisease, but they are not expensive or effective in calculating heart disease risk. Early detection of heart disease can reduce overall mortality and morbidity. However, in all cases, it is not possible to monitor patients accurately every day, and it is not possible to consult patients 24 hours a day, because it is more convenient, time-consuming and requires experience. Since there is a huge amount of data in the world today, we can use various machine learning algorithms to analyze the data for hidden patterns. Hidden patterns can be used to diagnose health in medical data.

1.7 GOALS AND OBJECTIVES

- The purpose of this project is to study the patient's gender, age, chest pain, fasting glucose level, etc. is for screening for cardiovascular disease based on

medical characteristics such as the goal of our heart disease prediction project is to determine whether patients should be diagnosed with heart disease, which is a biased result. Which classification model is the most accurate and to identify correlations in our data.

- To determine which features are most effective in diagnosing heart disease. The main goal of this project is to develop a heart disease prediction system. This system can discover and extract knowledge related to latent diseases from cardiac history databases

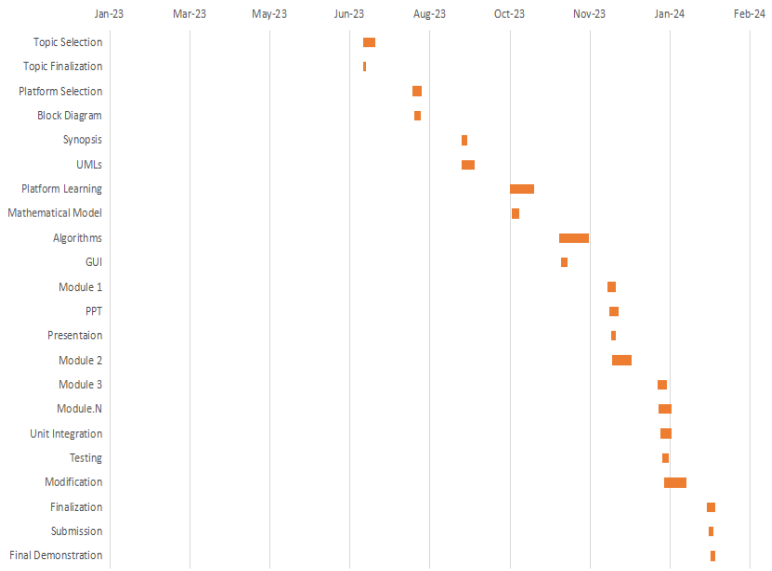
1.8 RELEVANT MATHEMATICS ASSOCIATED WITH THE PROJECT

System Description:

- Input=product,gesture
- Output=detected gesture
- Let S be the | System as the final set
- S = {U,Fr,G,A,F}
- Let U be the set of Users where,
  - S = {U}
  - U ={U1, U2, U3, U4.....| U }
  - where, U1=user1.
  - U2=user2.
- Let Fr be the set of Frame where,
  - S = {Fr}
  - Fr = {Fr1, Fr2, Fr3.....| Fr }
  - where, Fr1=Frame1
  - Fr2=Frame2

- Let G be the set of Gesture where,
  - S = {G}
- G = {G1, G2, G3.....| G}
- Let A be the set of Algorithm where,
  - S = {A}
  - A = {A1, A2, A3.....| A}
- Identify the functions as ?F
- S = {F}
- F = {F1 (), F2(), F3(), F4(), F5(), F6(), F7(),F8(),F9()}
  - F1 (S) = Grab Image
  - F2 (S) = define gesture
  - F3 (S) = Grey Scale
  - F4 (S) = Threshold
  - F5 (S) = Bluring
  - F6 (S) =Image Substraction
  - F7 (S) =Gesture Detection
  - F8 (S) =Add Product
  - F9 (S) =Manage Product

1.9 PLAN OF PROJECT EXECUTION



1.10 REVIEW OF CONFERENCE/JOURNAL PAPERS SUPPORTING PROJECT IDEA

1. Sunil S. Khatal, Dr. Yogesh Kumar Sharma “Analyzing the role of Heart DiseasePrediction System using IoT and Machine Learning” 2020
2. Chipara, Octav, Chenyang Lu, Thomas C. Bailey, and Gruia-Catalin Roman, “Reliable clinical monitoring using wireless sensor networks: experiences in a step-down hospital unit,” In Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, ACM, pp. 155–168, 2010.
3. Khambete, N. D, and A. Murray, “National efforts to improve healthcare technology managementand medical device safety in India,” Appropriate Healthcare Technologies for Developing Countries, 7th International Conference on, IET, pp. 1–5, 2012.
4. Priyan Malarvizhi Kumar and Usha Devi Gandhi,” A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases”, J. Computers and Electrical Engineering, pp. 222-235, 2018.

5. Mingyu Park and Younghoon Song,JAewonLee,JeongyeupPaek," Design and implementation of a smart chair system for IOT", IEEE,2016.
6. P.M. Kumar, S. Lokesh, R. Varatharajan, C. Gokulnath, P. Parthasarathy," Cloud and IoT baseddisease prediction and diagnosis system for healthcare using Fuzzy neural classifier", Future Generation Computer Systems, 2018.
7. Trevor. (2018). Enterprise Personal Analytics: A Research Agenda. 10.13025/S88H04.
8. Minerva, R., Biru, A., Rotondi, D. (2015),"IEEE- Towards a Definition of the Internet of Things(IoT)".
9. Bilal Afzal, Muhammad Umair, Ghalib Asadullah Shah, Ejaz Ahmed, "Enabling IoT platforms for social IoT applications: Vision, feature mapping, and challenges", Future Generation Computer Systems, 2017.
10. Luminoso, L. (2017). Creative engineering. Design Engineering (Canada), 63(1), 30-31. [11]Holler, J., T siats is, V., Mulligan, C., Avesand, S., Karnouskos, S., Boyle, D. (2014). From
11. Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence".
12. Koppar, Anant R, and Venu gopalachar Sridhar, "A workflow solution for electronic health records to improve healthcare delivery efficiency in rural India," In eHealth, Telemedicine, and Social Medicine, 2009. eTELEMED'09. International Conference on, pp. 227–232, IEEE, 2009.
13. Sharma YK, Khatal Sunil S. Health Care Patient Monitoring using IoT and Machine Learning,in IOSR Journal of Engineering (IOSR JEN) National Conference on "Recent Innovations in Engineering and Technology" MOMENTUM-19
14. Jha, R. K., Henge, S. K., Sharma, A. (2020). Optimal machine learning classifiers for predictionof heart disease. International Journal of Control and Automation, 13(1 Special Issue), 31-37.
15. Patil PR, Sharma Y, Kshirsagar M. U2R Attack Detection Using Machine Learning. Science[ETEBMS-2016]. 2016 Mar;5:6.

16. Sharma YK, Khan V. A Research on Automatic Handwritten Devnagari Text Generation inDifferent Styles Using Recurrent Neural Network (Deep Learning) Especially for Marathi Script.

17. Monika D.Rokade ,Dr.Yogesh Kumar Sharma,"Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic." IOSR Journal of Engineering(IOSR JEN), ISSN (e): 2250-3021, ISSN (p): 2278-8719

18. Monika D.Rokade, Dr. Yogesh Kumar Sharma. (2020). Identification of Malicious Activity for Network Packet using Deep Learning. International Journal of Advanced Science and Technology, 29(9s), 2324 - 2331.

19. Monika D.Rokade ,Dr.Yogesh Kumar Sharma"MLIDS: A Machine Learning Approach forIntrusion Detection for Real Time Network Dataset", 2021 International Conference on Emerging Smart Computing and Informatics (ESCI)

Chapter 2

Technical Keywords

2.1 AREA OF PROJECT

Machine learning.

2.2 TECHNICAL KEYWORDS

1. Euclidean distance
2. Nearest neighbors
3. Sigmoid function
4. Decision trees
5. Ensemble learning
6. Feature importance

- Gesture-based gaming: This refers to a system where a player must ”draw” a specific pattern on the screen in order for a game to initiate.Interfaces that allow the human body to interact with digital resources without the need for standard input devices like a keyboard, mouse, or game controller are known as gesture-based computing..

- Hand Gestures :A natural, innovative, and contemporary method of nonverbal communication is offered by a hand gesture recognition system. Sign language and human-computer interaction are two broad areas of hand gesture applications.

- Face gestures :The capacity to identify and decipher bodily gestures in order to communicate with and operate a computer system without making direct physical contact is known as face gesture control.

- Gesture Recognition:The capacity of a computer or other device to recognise and comprehend human gestures as input is known as gesture recognition. These motions include hand gestures and even symbols made with the fingers.

- Machine Learning :There are several uses for machine learning in the gaming industry.CNN uses the gesture’s fault spots to determine the number of fingers present in the gesture in order to recognise it.

Chapter 3

INTRODUCTION

3.1 PROJECT IDEA

According to the World Health Organization, 12 million people die from heart disease in the world every year. Heart disease is one of the leading causes of morbidity and mortality in the world population. Cardiovascular disease prediction is considered one of the most important topics in data analysis. The burden of heart disease has increased rapidly worldwide in recent years. Many studies have been done to identify the most influential factors of heart disease and accurately predict the overall risk.

Heart disease is also described as a silent killer that kills without obvious symptoms.Early diagnosis of heart disease plays an important role in lifestyle modification decisions in high-risk patients and in turn reduces complications.

Behavioral risk factors for heart disease and stroke include unhealthy diet, physical inactivity, smoking, and alcohol abuse. The effects of behavioral risk factors caninclude high blood pressure, high blood glucose, high blood lipids, obesity, and obesity. These ”intermediate risk factors” can be measured in primary care settings and increase the risk of heartattack, stroke, heart failure and other complications. Machine Learning is used in many areas around the world.

Machine learning can play an important role in predicting the presence/absence of locomotor diseases, heart diseases and more. Such information can provide important insights to doctors who, if predicted, can adjust diagnosis and treatment for their patients.Machine learning helps in making decisions and predictions from the large amount of data generated by the healthcare industry. The project aims to predict future heart attacks by analyzing patient data to determine whether they have heart disease using machine learning algorithms.

Machine learning techniques can have a huge impact on this. Although heart disease can come in many forms, there is a general set of key risk factors that influence whether someone will develop heart disease. In recent years, there has been rapid growth in healthcare services to provide wireless access means of communication between doctors and patients through wearable technology Telemedicine. Nowadays, diabetes is the leading cause of death for all people. 171 million in 2000 people predict it can increase to 642 million in the world by 2040. This increases you should pay attention to this disease. Many health care facilities around the world spend billions of dollars for diabetes health.

Type 1 diabetes is divided into four types of diabetes, pre- diabetes, type 2 diabetes, and pregnancy. Type 1 is caused by insulin deficiency young and old. Prediabetes is type 2 and gestational diabetes occurs before the onset of diabetes women during pregnancy. All these patients can be diagnosed at different levels to check your glucose level. A1C means a higher glucose level test used to diagnose Type 1 and pre-diabetes. The fasting glucose test is used to diagnose type 1 diabetes and type 2 diabetes. OTG - Oral glucose test is used to diagnose type 2 diabetes and gestational diabetes. High levels of Glucose can affect a person's health and cause serious conditions such as vision loss and kidney disease Neuropathy, liver, heart problems and foot problems. Too much sugar, diabetes. It is necessary to diagnose retinopathy, which can cause vision loss and night blindness.

Using data mining techniques in the medical field, a large amount of

data is generated every day and we can find hidden patterns that can be used for clinical diagnosis. Therefore, data mining plays an important role in the field of medicine, which can be proven by the work done in the last decade. We will work closely with predicting heart disease and for this we will look at the heart data database, get different insights that help us understand the weight of each feature and how they relate to each other.

The most important clinical parameters indicate an impending heart attack. This was determined using ROK analysis. Blood sugar (BS) - Fasting and blood sugar (BS) - Postprandial positive association with heart disease. However, respiratory rate (RP), heart rate (HR), blood pressure (BP): systolic interval (SR), Blood Pressure (BP): Diastolic Range (DR) and Body Temperature (BT) were found to be negatively associated with heart. Disease The test results show that the respiratory rate (RP) is around 50 and 12 is a very important indicator of heart disease. Heart palpitations are 160 symptoms of heart disease. Likewise, blood pressure (BP): systolic interval and body temperature (BT) of 140 and 37 are considered very important variables to indicate heart problems.

3.2 MOTIVATION

1. The main motivation of this project is to present a heart disease prediction model to predict the occurrence of heart disease. In addition, this project aims to determine the best classification algorithm to determine the probability of heart disease in a patient.
2. Today, most hospitals use hospital information systems to manage healthcare or patients' information. These systems generate large amounts of data, usually in the form of numbers, text, charts, and images. Unfortunately, these data are rarely used to make clinical decisions.
3. This raises an important question: "How can we turn data into useful information that allows healthcare professionals to make intelligent clinical decisions?" This is the main motivation for this project.

3.3 LITERATURE SURVEY

This paper according to Sunil S. Khatal Dr. Yogesh Kumar Sharma in "Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning." Prediction of heart disease is very important in today's environment; various studies have been done to predict heart disease from large database. The IoT environment basically generates data from various sensors and predicts the probability of disease accordingly. A variety of synthetic data, the main role of machine learning algorithms, is a variety of body parameters obtained by certain sensor values. In this study, we propose a combination of IoT and machine learning approaches to predict heart disease, an IoT environment designed to extract data from real-time body sensor networks (BSN) and remote sensing systems and store the data efficiently in a cloud server. Such studies are considered synthetic data used to predict the likelihood of heart disease. In this study, we present a variety of machine learning algorithms as well as some deep learning algorithms to achieve better control over the prediction of various diseases. Experimental analysis shows that deep learning classification algorithms are more effective than classical machine learning algorithms.

This paper According to Chipara et. all in "Reliable clinical monitoring using wireless sensor networks: experiencesin a step-down hospital unit" presents the design, deployment, and empirical investigation of a wireless clinical monitoring system that collects pulse and oxygen saturation readings from patients. The main contribution of this paper is an in- depth clinical trial evaluating the potential of a wireless sensor network for patient monitoring in a public hospital. We present a detailed analysis of system reliability in a seven-month long-term hospital setting involving 41 patients in a step-up cardiology unit. Retrospective data analysis showed that the system provided adequate temporal resolution to ensure transfer to the intensive care unit in three patients who experienced clinically significant events. These results show the potential and promise of using wireless sensor networks for continuous patient monitoring and detection of clinical deterioration in general hospital wards.

This paper describes according to Khambete, N. D A. Murray in "National efforts to improve healthcare technology management and medical device safety in India" In the practice of contemporary modern medicine, effective and safe use of healthcare technology is acknowledged worldwide as essential for any healthcare system.Achieving these goals can be particularly challenging in developing countries such as India, where an estimated of medical technology is imported and studies have shownthat almost of medical equipment is out of service. Furthermore, concerns regarding medical equipment safety have been raised in newspaper reports and also reported by a pilot study. However, recently, substantial efforts are being made to introduce changes in the health care system that will help in improving this situation. Discussions on these issues were initiated at two 'International Clinical Engineering Workshops' (Trivandrum 2009 and Pune 2011) and two 'Regional Clinical Engineering Workshops' (Latur and Mumbai 2011).

This paper according to Priyan Malarvizhi Kumar Usha Devi Gandhi in "A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases" Among the applications developed by the Internet of Things (IoT), continuous health monitoring systems are very important. Sensor devices that can be used in IoT health monitoring systems generate large amounts of data continuously. The data generation rate of IoT sensor devices is very high. Thus, the amount of data obtained from IoT-based health monitoring systems is also very high. To overcome this problem, this paper proposes a scalable three-tier architecture for large-scale wearable sensor data storage and processing. Tier-1 focuses on collecting data from IoT-enabled sensor devices. Tier-2 uses Apache HBase to store large amounts of wearable IoT sensor data in cloud computing.

Chapter 4

SOFTWARE AND HARDWARE REQUIREMENTS

4.1 SOFTWARE REQUIREMENT

1. Operating System: Windows 10
2. IDE: Jupyter Notebook, Spyder, VS code
3. Programming Language: Python

4.2 HARDWARE REQUIREMENTS

1. RAM: 4 GB As we are using Machine Learning Algorithm and Various High Level Libraries Laptop
2. RAM minimum required is 4 GB.
3. Hard disk: 256 GB
4. Processor: Intel i3 Processor.

4.3 USER INTERFACE

Application Based Heart disease detection.

4.4 SOFTWARE REQUIREMENT SPECIFICATION

A software requirements specification document describes the goals, requirements, and nature of the software to be developed. It also includes revenue and software costs. A software requirements specification (SRS) describes the nature of a project, software, or application. Simply put, an SRS is a manual prepared before a project begins. Software documents are generallyprepared for any project, software, or application type.

4.4.1 Purpose and scope of document

1. The central outline of the proposed algorithm is the implementation of Heart Disease Prediction System.
2. System predicts the heart disease successfully using machine learning algorithms.
3. By analyzing patient data, machine learning algorithms can detect early signs of heart disease that human doctors might miss. This early detection can prevent the disease from progressing to a more serious stage.
4. The system analyzes multiple data points to identify early signs of heart disease, such as irregular heartbeat, chest pain, and shortness of breath. Early detection of heart disease allows doctors to take steps to prevent infection and improve patient outcomes.
5. The system uses a variety of machine learning algorithms such as Extra Trees Classifier, KNN Classifier, Logistic Regression to determine heart disease.

4.4.2 Overview of responsibilities of developer

- One of the most important tasks in a machine learning project is preparing the data for analysis. Developer responsible for collecting data, cleaning it, and converting it into a format that machine learning algorithms can use.

- Developer responsible for developing machine learning algorithms for use in predicting heart disease. This may include selecting appropriate models, setting hyper parameters, and optimizing algorithms to improve accuracy.
- After machine learning algorithms are developed, developers must evaluate their performance against a validation database to ensure that they are accurate and reliable. Developers can use various metrics to evaluate models, such as precision, recall, F1- score, classification report.
- After the model is trained and validated, the developer can be responsible for deploying it in the production environment. It can enable the integration of this model with other systems and use by health professionals.
- Ensure that the project is completed within allotted budget and timeliness.

4.5 NON-FUNCTIONAL REQUIREMENTS

4.5.1 Performance requirements

The function and performance of each module must be good. The overall performance of the software will allow the user to work efficiently. Data encryption performance should be fast. Theperformance of providing a virtual environment can quickly become a security requirement. Theprogram is structured in modules that allow you to easily find and fix errors. It makes it easy to install new functionality and update as needed.

4.5.2 Safety requirements

The program is organized in modules that can detect and easily correct errors. This makes it easy to install new functionality and update as needed.

4.5.3 Software quality attributes

Our software has many qualities attribute that are given below : -

- Adaptability: This software is adaptable for all users.
- Availability: This software is freely available to all users. The availability of the software is easyfor everyone.
- Maintainability: After the deployment of the project if any error occurs then it can be easily maintained by the software developer.
- Reliability: The performance of the software is better which will increase the reliability of theSoftware.
- User Friendliness: Since the software is a GUI application, the output generated is much userfriendly in its behavior.
- Integrity: Integrity refers to the extent to which access to software or data by unauthorizedpersons can be controlled.
- Security: Users are authenticated using many security phases so reliable security is provided.
- Testability: The software will be tested considering all the aspect

Chapter 5

Project Plan

5.1 PROJECT ESTIMATES

The first Process Model to be introduced was the Waterfall Model. Another name for it is the linear-sequential life cycle model. It is quite easy to use and comprehend. Every stage in a waterfall model needs to be finished completely before moving on to the next. This kind of paradigm is typically applied to small-scale projects with well-defined needs. Every phase of the project is reviewed to see whether it is headed in the correct direction and to decide whether to move forward with it or scrap it altogether. According to this strategy, testing doesn't begin until after development is finished. Phases in a waterfall model do not cross over.

1. **Planning and requirement analysis**
   This process involves planning the suggested system structure and gathering the necessary requirements.
2. **Designing project architecture**
   The architects are really being designed by the developers at the second stage of the software development life cycle. We create the work flow, and the architecture is created in line with it.
3. **Development and programming**

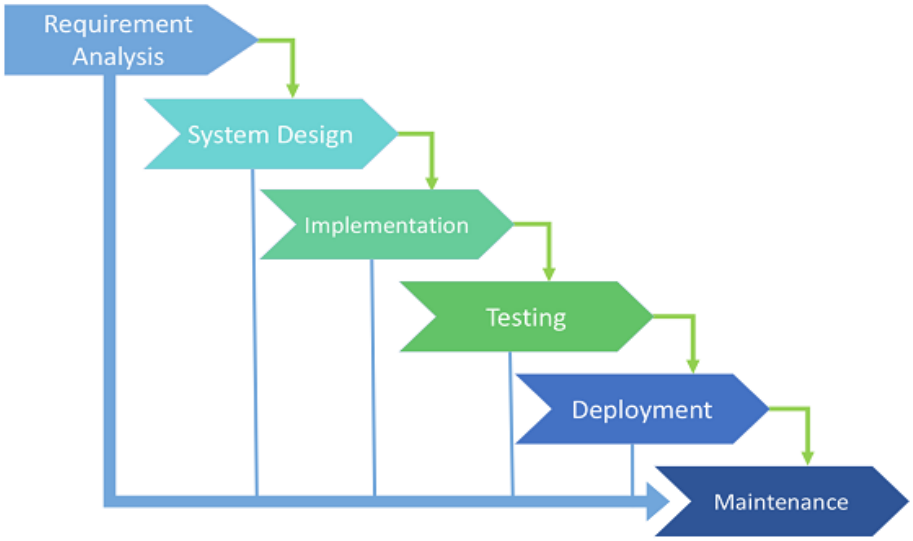


Figure 5.1: Waterfall Model

Following approval of the requirements, the procedure moves on to the real development phase. Python is being used to code the back-end algorithms for predictions.

- Source code writing
- Testing and debugging

4. **Testing**

   The debugging procedure is a part of the testing step. After evaluating each of the different inputs, the entire system is tested.

5. **Deployment**

   In this phase the overall proposed system is passed for deployment which means the system is ready for user usage.

5.2 RECONCILED ESTIMATES

5.2.1 Effort Estimate Table:

Task	Effort weeks	Deliverables	Milestones
Analysis of existing systems & compare with proposed one	4 weeks		
Literature survey	1 weeks		
Designing & planning	2 weeks		
System flow	1 weeks		
Designing modules & itś deliverables	2 week	Modules: design document	
Implementation	7 weeks	Primary system	
Testing	4 weeks	Test Reports	Formal
Documentation	2 weeks	Complete project report	Formal

Table 5.1: Effort Estimate Table

5.2.2 Project Description:

Phase	Task	Description
Phase 1	Analysis	Analyse the information given in the IEEE paper.
Phase 2	Literature survey	Collect raw data and elaborate on literature surveys.
Phase 3	Design	Assign the module and design the process flow control.
Phase 4	Implementation	Implement the code for all the modules and integrate all the modules.
Phase 5	Testing	Test the code and overall process weather the process works properly.
Phase 6	Documentation	Prepare the document for this project with conclusion and future enhancement.

Table 5.2: Project Scheduling

5.2.3 Reconciled Estimates

Cost Estimates

Cost of Project

C=N*Cp

C=3*4000

C=12000

The Cost of the project is approximately up to 12000

Time Estimates

Estimating LOC for this project is difficult at estimation stages this project is of innovative type project. Average estimation of this project is 10000 to 15000 line of code. LOC based Estimation: Efforts in Person in months

E = 3.2 (KLOC)

1.05 E = 3.2 9.0

1.05 to 11.0 4.2

1.05

Function	**Estimated KLOC**
GUI Design	1.1-1.3
Logical Code	1.5-2.0
Location Based Code	1.1-1.3
Directory Matching Code	1.0-1.3
Business Logic	2.2-2.5
Testing	1.1-1.2
Re-correct Code	1.0-1.2
Total	9.0-10.11

Table 5.3: Man Month Utilization

Project Resources

Project resources includes People,Software like Python IDE

5.3 RISK MANAGEMENT

5.3.1 Risk Identification

Business Impact Risk

- The quantity and caliber of documentation that has to be created and sent to the client; the client will receive a comprehensive online help file and Game Forge user manual. As a side note, since the client will be grading the project, they will also have access to all Game Forge development docs.
- There are no known government restrictions on the product's development.
- Costs associated with a delayed delivery A late delivery will prevent any employee from issuing a letter of approval for the product, which will result in an incomplete grade for the course.
- The expenses linked to a faulty product Not yet known.

Customer Related Risks

- Expenses related to aHave you previously collaborated with the client? Yes, each member of the team has finished at least one project for the client, although none of them has been as large as the one they are working on now.efficient product Not yet known.
- It is accurate to say that the customer may view both the system requirements and the software requirements specifications. The item Will the client consent to investing time in formal requirements collecting sessions in order to determine the scope of the project? Not sure. The consumer hasn't made the query yet, but they probably will if asked.

Process Risks

- Does senior administration uphold a composed strategy proclamation that underlines the significance of a standard cycle for programming improvement? N/A. Dad Programming doesn't have a senior administration. It

ought to be noticed that the organized strategy has been embraced. Toward the finish of the task, it still up in the air in the event that the product technique is OK as a standard cycle, or on the other hand on the off chance that changes should be carried out.

- Has your association fostered a composed portrayal of the product cycle to be utilized on this project? Yes.
- Are staff individuals ready to utilize the product interaction? Yes. The product interaction was settled upon before improvement work started.
- Is the product interaction utilized for different items? N/A. Dad Programming has no different tasks at present.

Technical Issues

- Are worked with application detail methods used to support correspondence between the client and the engineer? The advancement group will hold successive gatherings straightforwardly with the client. No proper gatherings are held (all casual). During these gatherings the product is examined and notes are taken for future audit.
- Are explicit techniques utilized for programming investigation? Extraordinary techniques will be utilized to examine the product progress and quality. These are a progression of tests and surveys to guarantee the product really depends on speed. For additional data, see the Product Quality Affirmation and Programming Design The board reports.
- Do you involve a particular technique for information and structural plan? Information and structural plan will be generally object situated. This takes into consideration a more serious level information exemplification and measured quality of code.

Technology Risk

- Is the innovation to be fabricated new to your association? Yes
- Does the product communicate with new or dubious equipment? No a specific UI requested by the item necessities? Yes.

Development Environment Risks

Is a specific UI requested by Is a product project the board instrument accessible? No. No product devices are to be utilized. Because of the current cutoff time, the advancement group felt it would be more useful to start carrying out the undertaking than attempting to learn new programming apparatuses. After the fruition of the undertaking programming devices might be executed for future projects.the item necessities? Yes.

5.3.2 Risk analysis

It is difficult to Examine risk. There will never be sufficient data you can accumulate. Obviously, a ton of that information is intricate, however most ventures have best practices, which can assist you with your gamble investigation. You may be shocked to find that your organization as of now has a structure for this interaction. At the point when you evaluate project risk you can at last and proactively address many effects, for example, keeping away from likely prosecution, resolving administrative issues, consenting to new regulation, decreasing your openness and limiting effect.

5.3.3 Overview of Hazard Alleviation, Observing, Management

Risk management organizational role

Every individual from the association will embrace risk the board. The advancement group will reliably be checking their advancement and undertaking status as to distinguish present and future dangers as fast and precisely as could really be expected. With this said, the individuals who are not straightforwardly engaged with the execution of the item will likewise have to keep their eyes open for any potential dangers that the advancement group didn't detect. The obligation of hazard the executives falls on every individual from the association, while William Ruler keeps up with this record.

5.4 PROJECT SCHEDULE

5.4.1 Project Task Set

- Getting Basic Knowledge of Python
- Going through the previous existing system
- Environment setup
- Load all required libraries
- Building GUI for better outlook
- Dividing the assign task among group members.

5.4.2 Time Line Chart

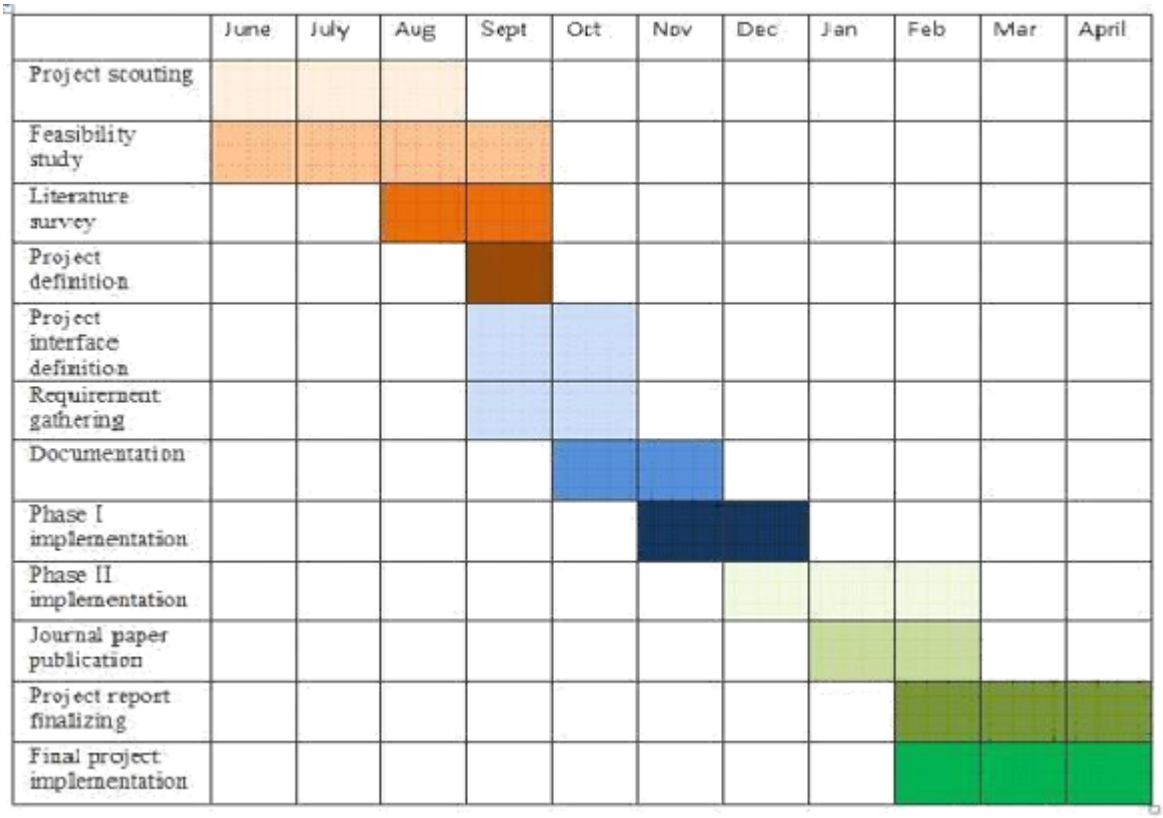


Figure 5.2: Time Line Chart

5.5 TEAM ORGANIZATION

5.5.1 Team Structure

Name	**Role**
Ayaan Shaikh	Coding
Khan Moin	Documentation
Samruddhi Kanade	Designing

Table 5.4: Team Structure

Chapter 6

SYSTEM DESIGN

6.1 INTRODUCTION

In this research to design and develop a system for dynamic and secure heart disease prediction system using machine learning. The first step in designing a system is identifying data sources and processing the data to ensure that it is suitable for analysis. This may involve cleaning the data, converting it into a format suitable for machine learning algorithms, and engineering it to extract relevant features. After the data is processed, an appropriate machine learning algorithm should be selected for the system. This may include choosing algorithms such as decision trees, randomforests, support vector machines, or neural networks, depending on the nature of the data

Selected machine learning algorithms are trained and validated using pre-processed data. This may include using techniques such as k-fold cross-sectional testing to ensure reliable and accurate models. After machine learning models are trained and validated, they must be integrated into systems that can be used in clinical settings. This may include developing a web-based interfaceor integrating the model with an electronic health record system. Systems must be tested and evaluated to ensure that they meet project goals and stakeholder needs. This may include conductinguser testing, evaluating system performance, and identifying areas for improvement.

6.2 SYSTEM ARCHITECTURE

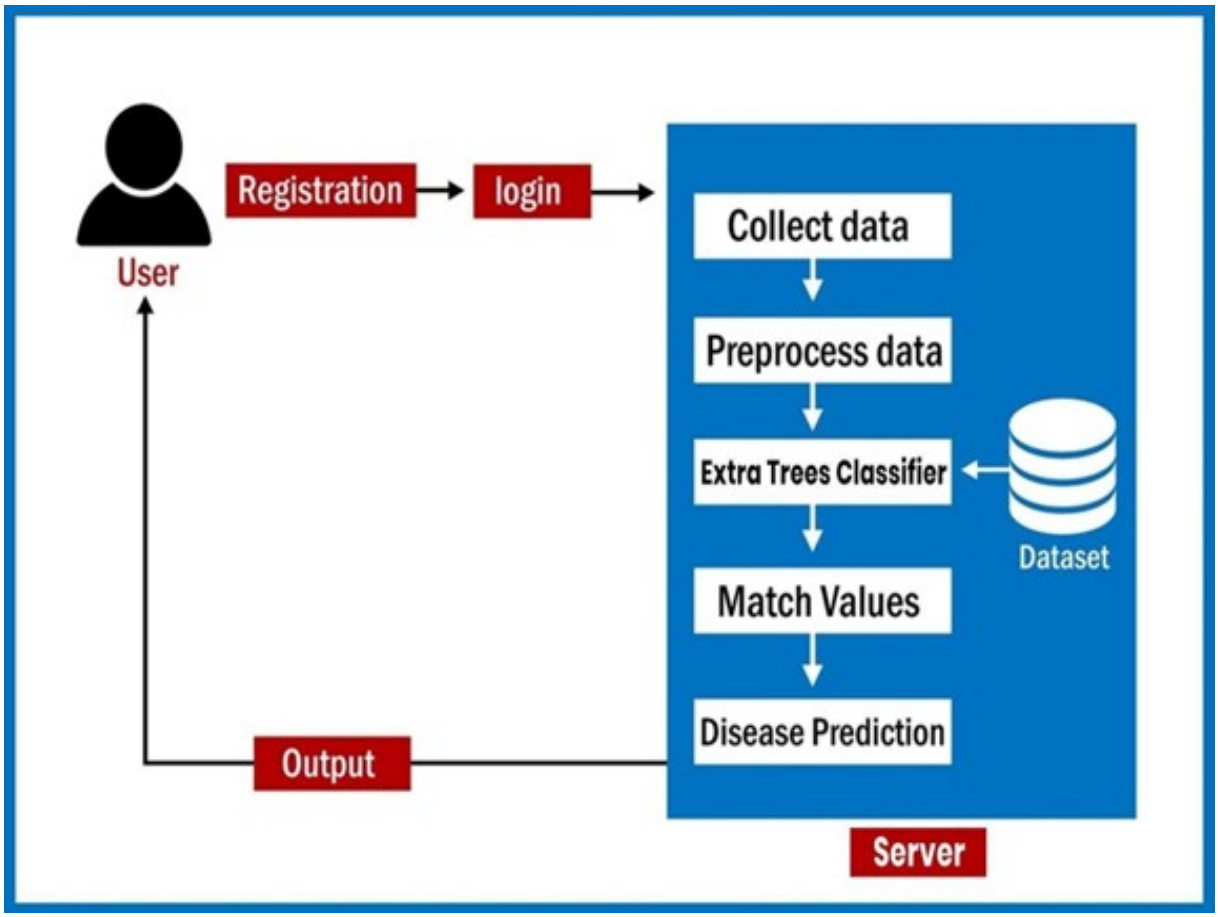


Figure 6.1: Proposed system architecture

6.3 MOTIVATION

The main motivation of this project is to present a heart disease prediction model to predict the occurrence of heart disease. In addition, this project aims to determine the best classification algorithm to determine the probability of heart disease in a patient. Today, most hospitals use hospital information systems to manage healthcare or patients' information. These systems generate large amounts of data, usually in the form of numbers, text, charts, and images. Unfortunately, these data are rarely used to make clinical decisions.

6.4 DATA FLOW DIAGRAM

In the data flow diagram, we show the rectangular current input of the data flow in our DFD0 system, as well as the data flow diagram (DFD) graphical representation of the data flow through the data system. Process aspect. Data is processed by the system in terms of input and output as determined by the system.

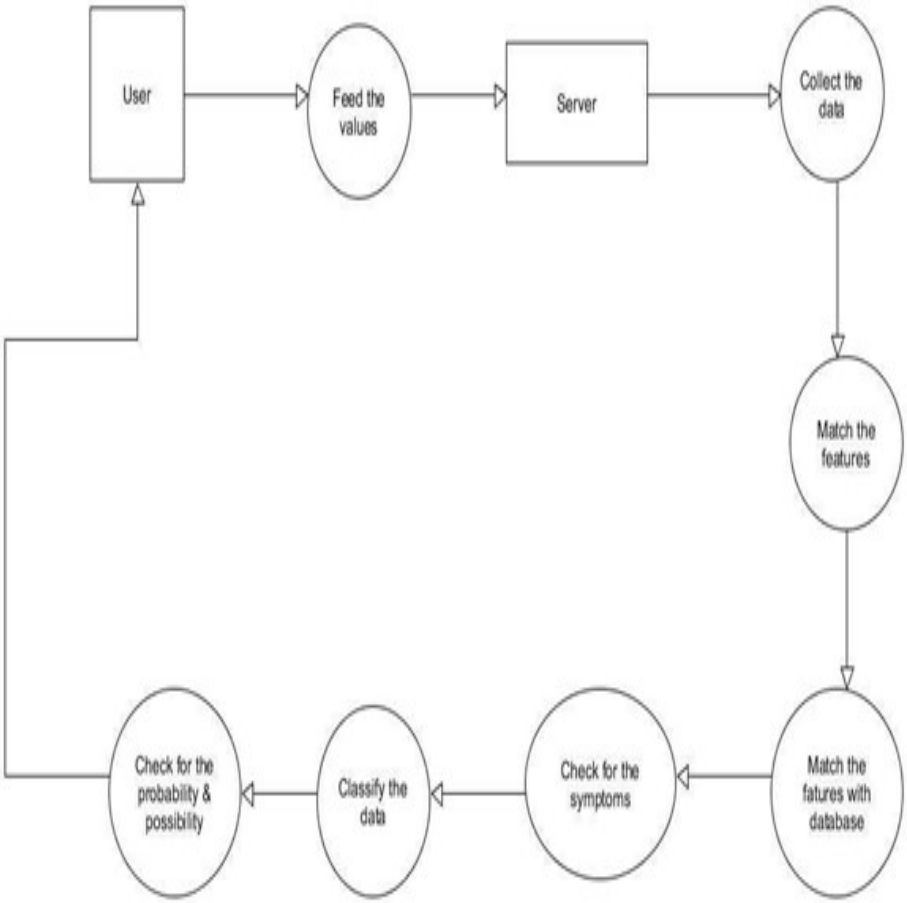


Figure 6.2: Data Flow Diagram

6.5 USE CASE DIAGRAM

A use case diagram is a type of UML diagram that shows the various ways a user or external system can interact with a system or software application. It is usually used to visualize the functional requirements of a system and identify the actors or users that interact with it.

Figure 6.3: Use Case Diagram

6.6 SEQUENCE DIAGRAM

A sequence diagram is a type of diagram that depicts the interaction of objects or components in a system over time. It shows the flow of messages and the order of exchange between different objects. Sequence diagram is useful to understand the behavior of the system, especially in the context of interaction between objects or components. It can improve system design by identifying potential problems and pointing out areas where communication or coordination can be improved.

A sequence diagram shows the interaction of objects in a time series. It represents the objects and classes involved in the scenario and the sequence of messages exchanged between the objects required for the scenario to work. Sequence diagrams are usually associated with use cases in the Logical model of the system being developed. A sequence diagram is sometimes called an event diagram or an event scenario. Sequence diagrams, often used by developers, use interactions between objects only once. It shows how different parts of the system interact with each other to perform functions and the sequence of interactions when specific use cases are executed. In this diagram users submit patient data to a heart disease prediction system to predict the likelihood of developing heart disease.

The system then checks the patient data and runs a heart disease prediction model. After the model processes the patient data, the system sends the prediction results back to the user. This sequence diagram shows how the heart disease prediction system interacts with the user to process patient data and provide prediction results. It shows the main steps in the process and the flow of messages between the user and the system.

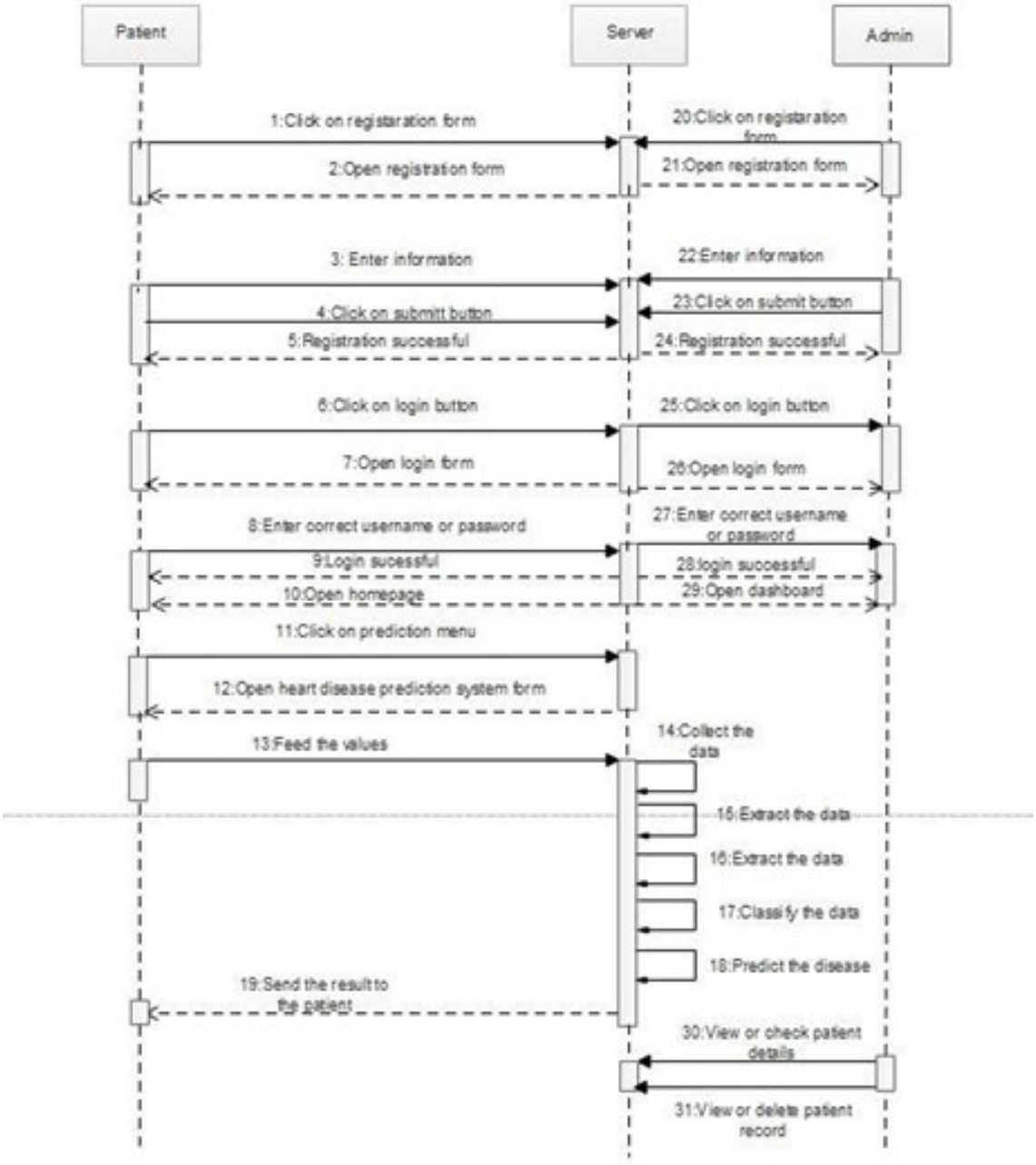


Figure 6.4: Sequence Diagram

6.7 ACTIVITY DIAGRAM

Activity Diagram A sequence diagram is a type of diagram that depicts the interaction of objects or components in a system over time. It shows the flow of messages and the order of exchange between different objects. Sequence diagram is useful to understand the behavior of the system, especially in the context of interaction between objects or components. It can improve system

design by identifying potential problems and pointing out areas where communication or coordination can be improved. A sequence diagram shows the interaction of objects in a time series. It represents the objects and classes involved in the scenario and the sequence of messages exchanged between the objects required for the scenario to work. Sequence diagrams are usually associated with use cases in the Logical model of the system being developed.

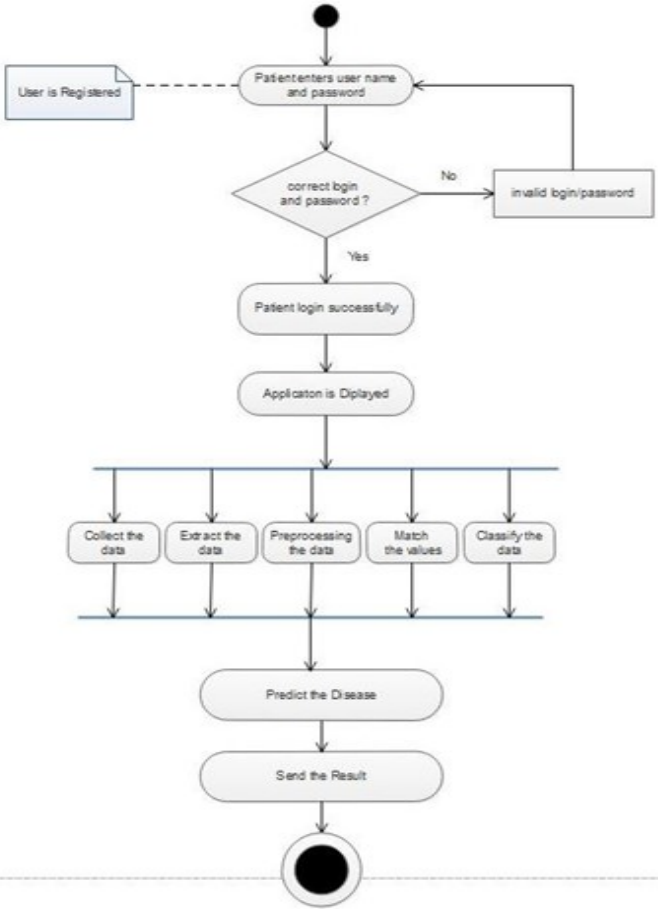


Figure 6.5: Activity Diagram

6.8 MODULES

6.8.1 Input dataset

This data set goes back to 1988 and consists of four databases: Cleveland, Hungary, Switzerland, and Long Beach V. The ”Target” field indicates that the patient has heart disease. The final number is 0 = no disease and 1 = predict disease. Contents:

1. age
2. sex
3. chest pain type (4 values)
4. resting blood pressure
5. serum cholesterol in mg/dl
6. fasting blood sugar ¿ 120 mg/dl
7. resting electrocardiographic results (values 0,1,2)
8. maximum heart rate achieved.
9. exercise induced angina.
10. old peak = ST depression induced by exercise relative to rest
11. the slope of the peak exercise ST segment
12. number of major vessels (0-3) colored by fluoroscopy.
13. . thal: 0 = normal; 1 = fixed defect; 2 = reversable defect
14. The names and social security numbers of the patients were recently removed from thedatabase, replaced with dummy values.

6.8.2 Data preprocessing

Data processing, part of data preparation, describes all operations performed on raw data to prepare the data for other processing procedures. It is usually an important initial step for data mining. More recently, data processing techniques have been adapted to trainand train machine learning models and AI models. Data processing transforms data into formats that are more easily and

efficiently processed for data mining, machine learning, and other data science tasks. This method isused in the early stages of machine learning and AI development pipelines to ensure accurate results. There are several different tools and methods used for preprocessing data, including the following:

- sampling, which selects a representative subset from a large population of data.
- transformation, which manipulates raw data to produce a single input.
- denoising, which removes noise from data.
- imputation, which synthesizes statistically relevant data for missing values.
- normalization, which organizes data for more efficient access; and
- feature extraction, which pulls out a relevant feature subset that is significant in a particular context.

  These tools and methods can be used on a variety of data sources, including data stored infiles or databases and streaming data. The steps used in data preprocessing include the following:

  1. Data profiling: Data processing is the process of checking, analyzing, and reviewing data to collect statistics about its quality. It begins by exploring the available data and its features. Data scientists identify data sets relevant to the problem at hand, record their important properties, and generate hypotheses about properties that may be relevant to a proposed analytics or machine learning problem. It also associates data sources with relevant business concepts and considers which processing libraries to use.

  2. Data cleansing: The goal here is to find the easiest way to fix quality problems, such as removing bad data, filling in missing data, or ensuring that the raw data is suitable for feature engineering.

  3. Data reduction: Raw data sets often contain data that characterizes phenomena in ways that are different or not specific to a particular ML, AI, or analytics problem. Data reduction uses techniques such as principal component analysis to transformraw data into simpler forms for specific use cases.

4. Data transformation: Here, data scientists think about how to organize different aspects of data to make them more meaningful for the purpose. This may include managing unstructured data, combining variables when meaningful, or identifying critical intervals to focus on.

5. Data enrichment: In this step, the data scientist applies various feature engineeringlibraries to the data to make the desired changes. The results should be informative as the new model is designed to achieve an optimal balance between the time requiredand the calculation required.

6. Data validation: At this stage, the data is divided into two sets. The first set is usedto train a machine learning or deep learning model. The second set is the test data, which is used to measure the accuracy and robustness of the resulting model. This second step helps to identify any problems in the hypothesis used in cleaning and engineering the data properties. If the data scientists are satisfied with the results, theycan send the preprocessing task to a data engineer, who will figure out how to scale it for production. If not, the data scientists can go back and make changes to the way they implemented the data cleaning and feature engineering steps.

6.8.3 Feature extraction

- Manual feature extraction requires identifying and describing features relevant toa particular problem and applying methods to extract those features. In many cases, a good understanding of the background or domain can help you make informed decisions about what features might be useful. Over decades of research, engineers and scientists have developed methods for extracting features for images, signals, and text. An example of a simple feature is the definition of a window on a signal.

6.8.4 Model selection

Machine learning models determine the results you get after running machine learning algorithms on the collected data. It is important to choose the right model forthe job at hand. Over the years, scientists and engineers have developed new technologies such as word recognition, image recognition, prediction, and more. have developed different models suitable for different problems such as You should also choose whether your model is suitable for quantitative or categorical data.

6.8.5 Training the model

Learning is the most important step in machine learning. In training, you feed your data into a machine learning model to find patterns and make predictions. It learns data models in order to complete a set of tasks. Over time, with practice, the model gets better at predicting.

6.8.6 Evaluating the model

After training your model, you need to test how it performs. This is done by testing the performance of the model on previously unseen data. The invisible data used is atest set that divides our data beforehand. If the test is performed on the same data used for training, you will not get accurate measurements because the model has been applied to the data and found the same pattern as before. This will give you unparalleled high accuracy.

6.8.7 Make prediction

You can use your model on unseen data to make predictions accurately.

Chapter 7

SYSTEM BEHAVIOR DESCRIPTION

7.1 ER-DIAGRAM

An ER (Entity-Relationship) diagram is a visual representation of the entities in a database and their relationships to each other. Using symbols and lines to describe objects, properties, and relationships between them. ER Diagram shows the relationship of different modules. An Entity-relationship model is a result of using a systematic process to describe and define a subject area of business data . An ER Model is typically implemented as a database. In case of a relational database, which stores data in a table, every row of each table represents one instance of an entity The main components of an ER diagram include attributes and relationships:

Entities: An entity is an object or concept that exists independently and has unique characteristics. Examples of entities are customers, orders, products, employees, and departments.

Attributes: An attribute is a characteristic of an entity. They describe the quality or characteristics of an organization. For example, the customer's organization name, address, phone number, and email address may be included.

Relationships: Relationships describe relationships between entities. They represent associations between entities in a database. For example, one customer can place multiple orders and orders can only be placed by one customer. This is an example of a multi-person relationship. ER diagrams use symbols to represent these parts. Some common symbols used in ER diagrams are:

Entity symbol: Shown as a rectangle with rounded corners.

Attribute symbol: Attribute symbol: Symbolized by an oval or ellipse.

Relationship symbol: This is represented as a diamond shape. Relationships between entities and relationships are represented by lines, with each type of relationship having a specific line type. This ER diagram shows the structure of a heart disease prediction system, including the entities, attributes, and relationships between them.

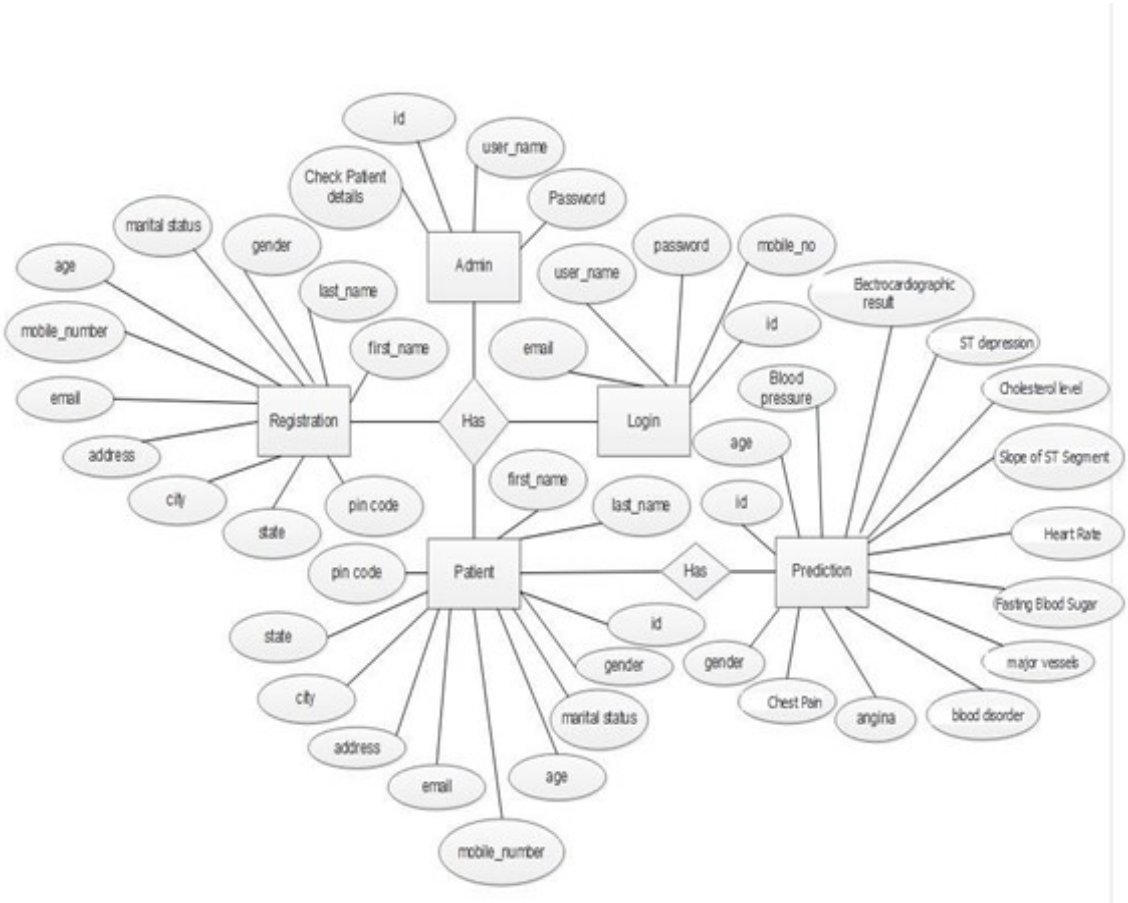


Figure 7.1: ER Diagram

Chapter 8

SOFTWARE INFORMATION

8.1 SOFTWARE INFORMATION

Python:

Python is an interpreted, high-level, and general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object- oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed, and garbage collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented, and functional programming. Python is often described as a" battery included" language due to its comprehensivestandard library.

Python was created in the late 1980s as a successor to the ABC language. Python 2.0, released in 2000, introduced features like list comprehension and a garbage collection system with reference counting. Python 3.0, released in 2008, was a major revision of the language that is not completely backward compatible, and much Python 2 code does not run unmodified on Python 3. The Python2 language was officially discontinued in.

Python was conceived in the late 1980s by Guido van Rossum at Centrum Wiskunde Informatica (CWI) in the Netherlands as a successor to the ABC language (itself inspired by SETL), capable of exception handling and interfacing with the Amoeba operating system. Its implementation began in December 1989. Van Rossum shouldered sole responsibility for the project, as the lead developer, until 12July 2018, when he announced his" permanent vacation" from his responsibilities as Python's Benevolent Dictator for Life, a title the Python community bestowed upon him to reflect his long-term commitment as the project's chief decision-maker. He now shares his leadership as a member of a five-person steering council. In January 2019, active Python core developers elected Brett Cannon, Nick Coghlan, Barry Warsaw, Carol Willing and Van Rossum to a five-member" Steering Council" to lead the project.

Anaconda:

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment.The distribution includes data-science packages suitable for Windows, Linux, and macOS. It is developed and maintained by Anaconda, Inc., which was founded by Peter Wang and Travis Oliphant in 2012.As an Anaconda, Inc. product, it is also known as Anaconda Distribution or Anaconda Individual Edition, while other products from the company are Anaconda Team Edition andAnaconda Enterprise Edition, both of which are not free.

Package versions in Anaconda are managed by the package management system conda. This package manager was spun out as a separate open-source package as it ended up being usefulon its own and for other things than Python. There is also a small, bootstrap version of Anacondacalled Mini conda, which includes only conda, Python, the packages they depend on, and a small number of other packages. Anaconda distribution comes with over

250 packages automatically installed, and over 7,500 additional open-source packages can be installed from PyPI as well as the conda package and virtual environment manager. It also includes a GUI, Anaconda Navigator,as a graphical alternative to the command line interface (CLI). The big difference between conda and the pip package manager is in how package dependencies are managed, which is a significant challenge for Python data science and the reason conda exists. When pip installs a package, it automatically installs any dependent Python packageswithout checking if these conflict with previously installed packages [citation needed]. It will install a package and any of its dependencies regardless of the state of the existing installation [citation needed].

Because of this, a user with a working installation of, for example, Google TensorFlow, can find that it stops working having used pip to install a different package that requires a different version of the dependent NumPy library than the one used by TensorFlow. In some cases, the package may appear to work but produces different results in detail. In contrast, conda analyses the current environment including everything currently installed, and, together with any version limitations specified (e.g., the user may wish to have TensorFlow version 2,0 or higher), works out how to install a compatible set of dependencies, and shows a warning if this cannot be done.

Open-source packages can be individually installed from the Anaconda repository, Anaconda Cloud (anaconda.org), or the user's own private repository or mirror, using the conda install command. Anaconda, Inc. compiles and builds the packages available in the Anaconda repositoryitself, and provides binaries for Windows 32/64bit, Linux 64 bit and MacOS 64-bit. Anything available on PyPI may be installed into a conda environment using pip, and conda will keep trackof what it has installed itself and what pip has installed. Custom packages can be made using the conda build command and can be shared with others by uploading them to Anaconda Cloud, PyPI or other repositories. .

Streamlit:

Streamlit is an open-source python framework for building machine learning and data science web applications. We can instantly develop web applications and easily deploy them with Streamlit. Streamlit allows you to write an application the same way you write python code. Streamlit allows you to seamlessly work on an interactive coding loop and view the results in a web application. If the python source code of the streamlit script changes, the application will display in the upper right corner whether to restart the application or not. You can also select the ”Always restart” option to restart every time the source script changes. This makes our development process much easier and every time you make some changes it will be reflected in your web application immediately. This loop between encoding and viewing live results allows you to work seamlessly with streamlit technology. Streamlit allows you to write an application the same way you write python code.

Streamlit is a popular open-source Python library used for building data applications. It provides an easy-to-use interface for building and deploying web applications in minutes. Streamlit is designed to simplify the process of creating web applications using Python scripts. It is an excellent choice for those who want to quickly create interactive data applications without having to deal with the complexities of web development. Streamlit is a high-level library that allows users to quickly create interactive data visualizations and web applications. It is built on popular Python data analysis libraries such as NumPy, Pandas and Matplotlib. Streamlit is easy to use and requires minimal code, making it a popular choice for data scientists, developers, and beginners.

Streamlit is designed to make it easy to build web applications in Python. It provides a simple and intuitive interface for building data applications. The library is built with the idea of ”building an app in minutes”. With Streamlit, you can easily create web applications that are interactive and engaging. One of the key features of Streamlit is its ability to create reactive web applications.

In terms of deployment, Streamlit makes it easy to deploy your web application to various platforms such as Heroku, Google Cloud, and Amazon Web Services. It provides a simple command-line interface to deploy your application to the cloud. In short, Streamlit is a powerful Python library that makes it easy to create interactive web applications and data visualizations. With its simple and intuitive interface, built-in components, and support for popular data analysis and visualization libraries, Streamlit is an excellent choice for creating web applications in Python. Streamlit is a versatile library that provides a variety of features for building interactive data applications. Here are some other features of Streamlit:

- **Customizable Themes:** Streamlit provides a variety of built-in themes that can be used to customize the look and feel of your app. You can also create your own themes using custom CSS.

- **Interactive Widgets:** Streamlit provides a wide variety of interactive widgets such as buttons, checkboxes, radio buttons and more. These widgets allow users to work with data and modify the application's behavior.

- **Sharing and Collaboration:** Streamlit makes it easy to share and collaborate on your apps with others. You can share your app's URL with others, and they can view and interact with your app in real time.

- **Session State Management:** Streamlit provides built-in support for session state management. This feature allows your application to store and retrieve user input and other data across multiple sessions.

- **Integration with machine learning libraries:** Streamlit provides integration with popular machine learning libraries such as TensorFlow, PyTorch, and Scikit-learn. This allows you to easily create interactive machine learning applications.

- **Real-time data streaming:** Streamlit provides a powerful feature for real-time data streaming to your application. This feature allows you to visualize and analyze streaming sources such as stock prices, weather data, and more.

- **Dynamic Caching:** Streamlit provides dynamic caching to help you optimize your application's performance. The caching mechanism automatically stores and retrieves frequently accessed data, improving the overall performance of your application.

- **Serverless deployment:** Streamlit can be easily deployed on serverless platforms such as AWS Lambda and Google Cloud Functions. This allows you to deploy your application without the need for dedicated servers, reducing costs and improving scalability.

Spyder:

Spyder is a powerful scientific environment written in Python for Python and designed by scientists, engineers, and data analysts. It represents a unique combination of advanced editing, analysis, debugging and profiling features of a comprehensive development tool with the data exploration, interactive execution, in-depth inspection, and beautiful visualization capabilities of a scientific package.

Spyder is a free, open-source scientific environment written in Python for Python and designed by scientists, engineers, and data analysts. It represents a unique combination of advanced editing, analysis, debugging and profiling features of a comprehensive development tool with the data exploration, interactive execution, in-depth inspection, and beautiful visualization capabilities of a scientific package.

Visual Studio Code (VS Code):

Visual Studio Code (VS Code) is a free and open-source source code editor developed by Microsoft. It was first released in 2015 and has since become one of the most popular code editors in the world. VS Code is a cross-platform application that runs on Windows, macOS, and Linux. It has a number of features that make it a popular choice among developers, including support for multiple programming languages, debugging, code navigation, and extensibility.

Features:

1. **Integrated Development Environment (IDE):**
   VS Code provides a complete IDE that supports a few programming languages, including JavaScript, TypeScript, Python, and many others. It has a built-in terminal that allows developers to run commands and run code directly from the editor. VS Code also supports debugging, making it easier for developers to identify and fix errors in their code.
2. **IntelliSense:**
   IntelliSense is a VS Code feature that provides intelligent code completion, code hinting, and syntax highlighting. IntelliSense helps developers write code faster and more accurately by suggesting code snippets, function names, and variable names as they type.
3. **Git Integration:** VS Code has excellent Git integration that makes it easy for developers to manage their code repositories. Git is a popular version control system that helps developers track changes to their code and collaborate with others. VS Code provides a built-in Git interface that allows developers to commit, download, and commit code changes directly from the editor.
4. **Extensions:** VS Code has an extensive library of extensions that allow developers to customize and extend the functionality of the editor. These extensions range from simple themes to complex tools that provide support for specific programming languages, frameworks, and technologies. Developers can also create their own extensions using the VS Code Extension API.
5. **Task Runner:** VS Code has a built-in task runner that allows developers to automate repetitive tasks such as compiling code or running tests. The job launcher can be configured to run jobs on save, build, or on demand.
6. **Debugging:** VS Code has a powerful debugging interface that allows developers to debug their code using breakpoints, trace expressions, and call stacks. The debugger supports a few programming languages, including JavaScript, TypeScript, Python, and many others.

7. **Live Share:** Live Share is a VS Code feature that allows developers to collaborate in real-time on the same code base. With Live Share, developers can share their editor, terminal, and debug sessions with others, making it easy to collaborate on complex code projects.
8. **Code Navigation:** VS Code has excellent code navigation features that make it easy for developers to find and navigate to specific code locations. These features include Go to Definition, Peek Definition and Find All References.
9. **Integrated Terminal:** VS Code has a built-in terminal that allows developers to run commands and run code directly from the editor. Terminal supports a variety of shells, including Bash, PowerShell, and Command Prompt.
10. **Multi-Language Support:**
    VS Code supports a wide variety of programming languages, including JavaScript, TypeScript, Python, Java, C++, and many more. The editor provides syntax highlighting, code completion, and debugging support for each of these languages.
11. **Customizable User Interface:** VS Code has a customizable user interface that allows developers to customize the appearance of the editor to their liking. Developers can choose from a few built-in themes or create their own themes using the VS Code Theme API.
12. **Performance:** VS Code is designed to be fast and responsive even when working with large codebases. The editor uses a combination of optimized rendering and background processing to ensure a smooth user experience.
13. **Accessibility:**

VS Code is designed to be accessible to everyone, regardless of ability. The editor provides keyboard shortcuts, screen reader support, and other accessibility features to ensure that the editor can be used by all developers.

Chapter 9

PROJECT PLAN

9.1 PROJECT PLANNING

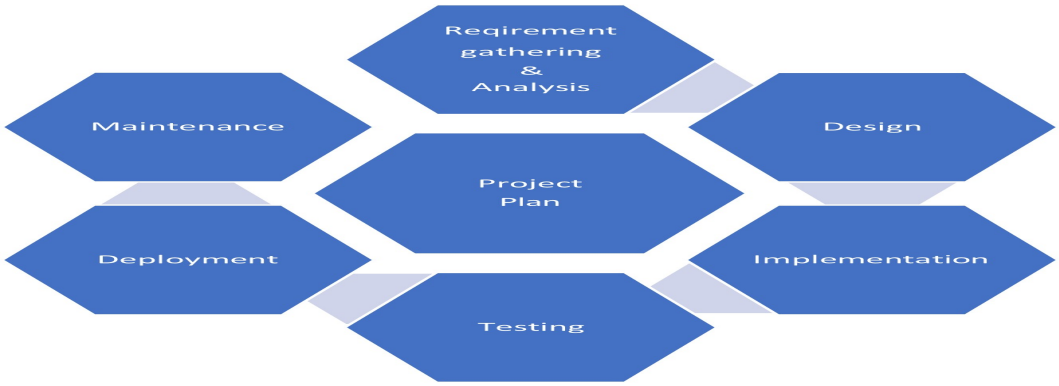


Figure 9.1: DFD Zero Diagrams

9.2 PLANNING:

Firstly, we are defining the project goals and objectives, identifying the project requirements, and creating a roadmap for achieving the desired outcomes. It includes complete estimation and scheduling and discusses the cost required for this project.

9.3 MODELING:

An activity diagram shows the complete flow of the system, and the algorithm is solved step by step. Also, we must use case diagram, sequence diagram ER diagram, data flow diagram. This all the diagram shows how heart disease prediction system works, flow, sequence.

9.4 IMPLEMENTATION:

9.4.1 Coding:

We design the heart disease prediction system using programming languages. We use different programming languages such as python, CSS, HTML, PHP, bootstrap. For designing we use HTML, CSS. For backend Python language is used.

9.4.2 Testing:

Testing is done through system analysis. We manually test for input and output errors such as interface errors, performance errors, data structure errors, and initialization errors.

9.5 DEPLOYMENT:

After testing in this phase, we have deployed our project.

9.6 MAINTENANCE:

Continuously improve the system by updating the model(s) with new data and incorporating new features and functionalities.

9.7 ADVANTAGES:

- Machine learning algorithms can analyze large amounts of patient data and identify patterns that human doctors might miss.
- The heart disease prediction system can provide more accurate and reliable predictions about the likelihood of developing heart disease.
- Heart disease prediction systems can help reduce healthcare costs by identifying patients at high risk for heart disease and providing early intervention.
- This allows doctors to take preventive measures and make lifestyle changes that can reduce the risk of heart disease and improve patient outcomes. By providing more accurate heart disease risk predictions and personalized care recommendations, heart disease prediction systems can help improve patient outcomes and reduce heart disease-related morbidity and mortality.

9.8 HARDWARE RESOURCES:

- System : Intel I3 Processor.
- Hard Disk : 256 GB.
- Monitor : 15 inches
- Ram : 4 GB

9.9 SOFTWARE RESOURCES:

- Operating system : Windows 10.
- Coding Language : Python, HTML, CSS
- IDE : Jupyter Notebook, VScode, Spyder

9.10 COST ESTIMATION OF PROJECT:

Equation for calculation of cost of project using COCOMO -2 model is:

C=D*C,

Where,

C = Cost of project

D = Duration in month

C = Cost incurred per person-month, C, Rs.5000/- (per person-month) (approx.) C=9*5000

=45000/-

Hence according COCOMO - 2 model the cost of project is 45000/-(approx.)

Chapter 10

PROJECT SCHEDULE

10.1 PROJECT SCHEDULE:

Major Tasks in the Project stages are:

Task 1: Literature Research:

Literature research involves studying and analyzing written works, such as books, articles, and other forms of literature, to gain knowledge and insights on a particular topic. It is a systematic process that includes searching for relevant sources, critically evaluating them, and synthesizing the information to support a research question or thesis. Literature research often forms the foundation of academic studies, allowing scholars to explore existing knowledge, identify gaps, and contribute new ideas to their field. It requires effective information retrieval skills, critical thinking, and the ability to organize and present findings in a coherent manner.

Task 2: System Analysis:

System analysis is a process of studying and understanding complex systems to identify their components, interactions, and behaviors to improve their efficiency, effectiveness, and functionality. It involves gathering information, defining requirements, and proposing solutions to meet specific goals. During system analysis, analysts examine the existing system or the problem at hand, collect data and requirements from stakeholders, and analyze the information

to identify areas of improvement. This may involve conducting interviews, surveys, and observations to gain a comprehensive understanding of the system and its users.

The analysis phase typically includes the following key steps:

1. Problem Identification: Clearly define the problem or the objectives of the system that need to be analyzed.
2. Requirements Gathering: Collecting information from stakeholders, users, and other relevant sources to understand their needs, expectations, and constraints.
3. Process Modeling: Creating models and diagrams to represent the flow of data, activities, and interactions within the system. This helps in visualizing and understanding the system's processes.
4. Data Modeling: Analyzing the data requirements of the system and creating data models that define how data is structured, stored, and utilized within the system.
5. Feasibility Analysis: Assessing the practicality and viability of proposed solutions by considering factors such as cost, time, resources, and technical constraints.
6. Solution Proposal: Presenting recommendations and alternative solutions to address the identified problems or achieve the desired goals of the system.
7. Documentation: Documenting the findings, requirements, models, and proposed solutions for future reference and communication with stakeholders.

System analysis plays a crucial role in the development and improvement of various systems, including software applications, business processes, and organizational structures. It helps in identifying inefficiencies, bottlenecks, and opportunities for optimization, leading to enhanced performance and productivity.

Task 3: Design Planning and Dataset:

Design Planning: Design planning refers to the process of creating a blueprint or a roadmap for the development of a system, product, or project. It involves determining the overall structure, components, and functionality of the design based on the requirements and objectives identified during the analysis phase. In design planning, the following key aspects are typically considered:

1. System Architecture: Defining the high-level structure and organization of the system, including its modules, components, and their interactions.
2. User Interface Design: Creating a user-friendly and intuitive interface that facilitates effective interaction between users and the system.
3. Data Design: Designing the data structures and databases required to store and manage the system's data efficiently.
4. Algorithm and Logic Design: Designing algorithms and logical processes that drive the system's functionality and operations.
5. Integration Planning: Planning how different components, modules, or subsystems will be integrated to ensure smooth operation and compatibility.
6. Security and Performance Considerations: Incorporating security measures and optimizing system performance through design decisions.

Design planning is crucial as it establishes the foundation for the development process, guiding the implementation phase and ensuring that the final product meets the desired requirements and goals.

Dataset:

A dataset is a collection of organized and structured data that is used for analysis, research, or training machine learning models. It can be in various forms, such as a spreadsheet, database, text files, images, or videos. Datasets are essential in data-driven fields as they provide the raw material for analysis, modeling, and decision-making. They serve as the basis for extracting insights, discovering patterns, and making predictions or classifications. Datasets can

be generated through various methods, including data collection, data synthesis, or data aggregation from multiple sources.

Key characteristics of a dataset include:

1. Size: The amount of data contained within the dataset, ranging from small-scale datasets to large-scale ones, depending on the specific application.
2. Structure: The organization and format of the data, such as tabular, hierarchical, or unstructured data.
3. Variables and Attributes: The different features or attributes associated with each data point, which provide the context and information for analysis.
4. Data Quality: The accuracy, completeness, and consistency of the data, which affect the reliability of any analysis or modeling performed on the dataset.

Datasets are used in a wide range of domains, including scientific research, business analytics, healthcare, finance, and artificial intelligence. They enable researchers, analysts, and developers to gain insights, make informed decisions, and train models to perform various tasks, such as image recognition, natural language processing, or recommendation systems.

Task 4: Learning Required Technologies:

Learning required technologies refers to the process of acquiring knowledge, skills, and proficiency in specific technologies that are necessary for a particular field or job role..

Learning required technologies is an ongoing process throughout your professional career. Embracing a proactive approach to learning and adapting to emerging technologies will help you remain competitive, broaden your skill set, and open up new opportunities in the ever-changing job market.

Task 5: Implementation:

Implementation refers to the stage in a project or process where the planned activities and designs are put into action or executed. It involves the actual development, construction, or deployment of a system, product, or solution based

on the specifications and requirements established during the earlier phases of planning and design. The implementation phase is a critical step in turning plans and designs into tangible results. It requires coordination, attention to detail, and effective project management to ensure that the implemented solution aligns with the intended goals and meets the defined requirements.

Task 6: System Testing:

System testing is a crucial phase in the software development lifecycle that involves evaluating a system or software application to ensure that it functions correctly and meets the specified requirements. It is performed after the implementation phase and before the system is deployed to end-users or customers. System testing involves various types of testing, including functional testing, performance testing, usability testing, security testing, and compatibility testing.

Task 7: Initial Report :

An initial report refers to a preliminary document that provides an overview or summary of a specific topic, project, or situation. It serves as an introductory or initial assessment of the subject matter, presenting key information and findings in a concise format. The main objective of system testing is to identify defects, errors, or discrepancies in the system's behavior, functionality, and performance. It aims to ensure that the system operates as expected and meets the needs of its intended users.

Task 8: Final Report:

An initial report sets the stage for further exploration and decision-making. It provides stakeholders with a snapshot of the subject matter, highlights key areas of interest or concern, and helps define the direction for subsequent actions, investigations, or reporting. An initial report generally follows a structured format.

Chapter 11

PROJECT IMPLEMENTATION

11.1 INTRODUCTION:

We must implement a heart disease prediction system using machine learning algorithms. The implementation of the heart disease prediction system requires expertise in machine learning, software, and healthcare. It is important to follow data privacy, security and ethical best practices to ensure that patient information is protected and used appropriately. Collaboration with healthcare providers and domain experts can ensure that the system is designed and deployed in a way that meets the needs of patients and healthcare providers. Collection of large datasets of patient health information, including demographics, medical history, laboratory results, and imaging studies.

11.2 LIST OF MODULE AND FUNCTIONALITY:

- **Data Collection Module:** Collects patient information such as demographics, medical history, laboratory results, and imaging studies.
- **Data Processing Module:** Transforms, normalizes and transforms collected data into a suitable format for machine learning algorithms.
- **Feature selection:** Module selects relevant features from the database that are more predictive of heart disease.

- **Machine Learning Model Training Module:** Trains a machine learning algorithm on pre-processed data to predict the likelihood of a heart attack.
- **Machine Learning Model Evaluation Model:** Evaluate the performance of machine learning models using various metrics such as accuracy, sensitivity, and specificity.
- **User Interface Module:** Provides an interface for users to enter patient information and view cardiac prognosis.
- **Risk Assessment Module:** Calculates heart disease risk based on patient input data and trained machine learning models.

11.3 ALGORITHMS / METHODOLOGIES:

11.3.1 Extra Trees Classifier:

Extra tree classification (ETC) is an ensemble learning algorithm used for classification problems in machine learning. This is an extension of the random forest algorithm, where multiple decision trees are constructed and predicted collectively. ETCs introduce randomness into the tree construction process to reduce redundancy and improve generalization performance. Extra Trees Classifier (ETC) is an ensemble learning method based on decision trees. It is similar to the Random Forest Classifier, but with a few key differences. In ETC, decision trees are trained using random subsets of the input features, and the splits in the trees are determined randomly, rather than based on an optimal split. This approach of using random splits and subsets is intended to increase diversity among the trees in the ensemble, which can help reduce switching and improve generalization performance.

ETC can be used for both binary and multi-class classification tasks, as well as regression tasks using the Extra Trees Regressor. It is often used as a powerful and efficient classification algorithm, especially when dealing with large datasets with many features.

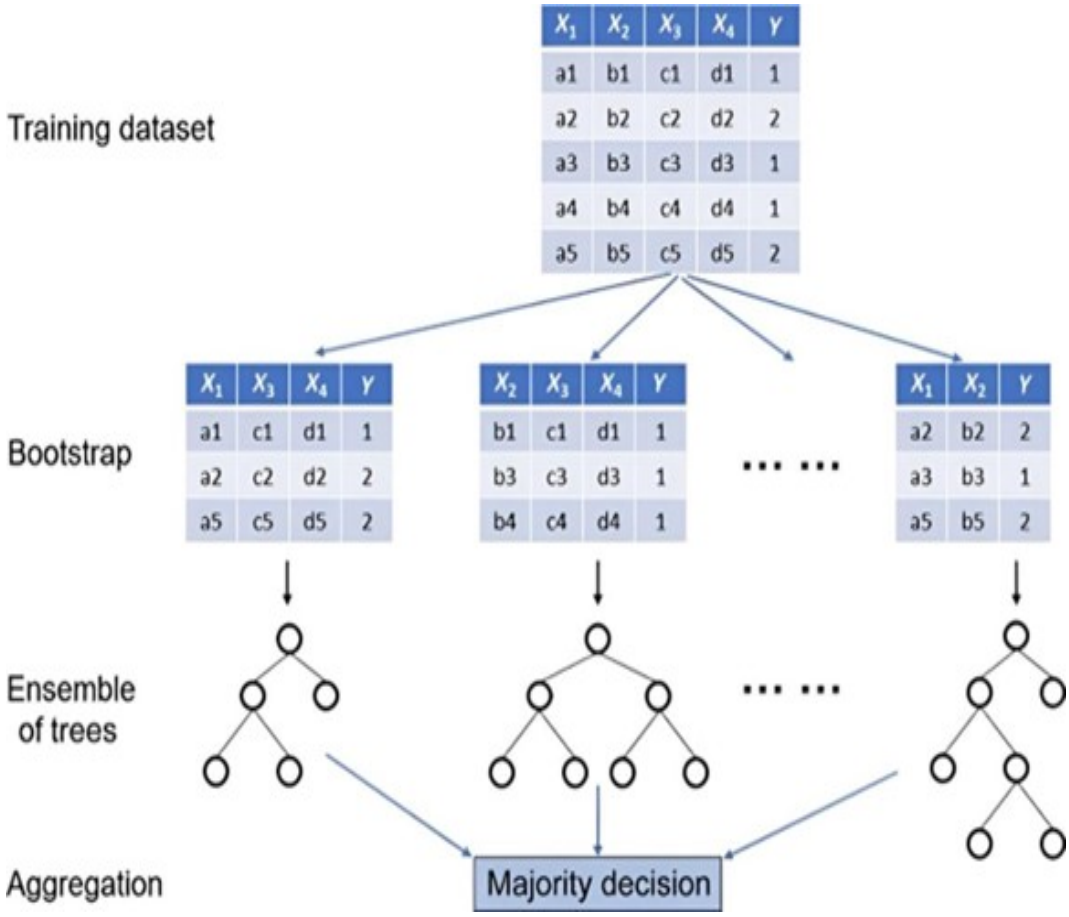


Figure 11.1: Extra Trees Classifier

11.3.2 Logistic Regression:

Logistic regression is a statistical method to analyze data where one or more independent variables determine the outcome. It is usually used for binary classification problems, the resulting variable is a binary variable, which means it needs one of two possible values. In logistic regression, the output of the model is the probability that the variable is 1 (or positive class).

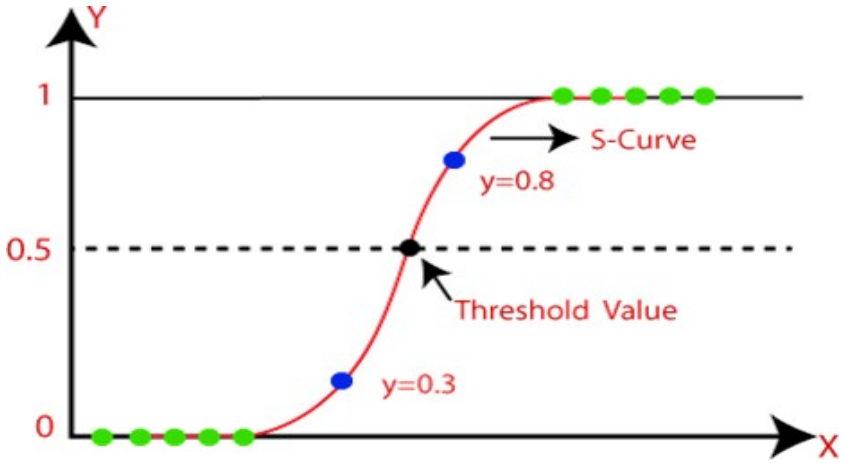


Figure 11.2: Logistic Regression Curve

$$y = \frac{e^{(b_0 + b_1 X)}}{1 + e^{(b_0 + b_1 X)}}$$

Figure 11.3: Logistic Regression Formula

Steps in the Logistic Regression algorithm:

1. Data Processing: Data is cleaned, pre-processed and missing values are controlled. Attribute scaling: Attributes are standardized, and their values are brought into the same range. This step helps the model to be closer.
2. Model Training: The model is trained on pre-processed data using a probability-based estimation method. The parameters of the model are studied in such a way that the probability of the observed data is increased.
3. Model Evaluation: The trained model is evaluated on the validation database. The evaluation criteria used are precision, accuracy, recall, F1 score and receiver operating characteristic (ROC) curve.
4. Prediction: After the model is trained and evaluated, it is ready to make

predictions about new data.

The logistic regression algorithm has several advantages:

- It is simple to implement and can be easily understood.
- It can handle both binary and multiclass classification problems.
- It is robust to noise in the data and can work well with small datasets.
- The output of the model is the probability of the outcome variable, which can be useful in many applications.

11.3.3 K- Nearest Neighbors Classifiers:

The k-Nearest Neighbors (k-NN) classifier is a popular machine learning algorithm used for both classification and regression tasks. It is a type of lazy learning algorithm, which means that it does not learn the model from the training data, but instead remembers the training data. In the k-NN algorithm, ”k” represents the number of nearest neighbors that will be used to classify a new data point. The classification of a new data point is determined by most votes of its nearest neighbors in the training set

The distance between two data points is usually measured using the Euclidean distance, although other distance metrics can be used. One of the advantages of k-NN is its simplicity and interpretability. It can be easily implemented and understood even by those without strong machine learning experience. However, K-NN can suffer from the curse of dimensionality, which means that its performance can degrade as the number of features in the dataset increases. Moreover, k-NN does not handle unbalanced data well because it tends to favor the majority class in the dataset. Therefore, it is often used in conjunction with other algorithms or preprocessing techniques to solve these problems.

Chapter 12

SOFTWARE TESTING

Types of Testing:

12.0.1 Unit Testing

Unit testing involves the design of test cases that verify that the internal program the logic works correctly, and the program inputs produce valid outputs. All decision branches and internal code flow should be verified. It's testing individuals' software units of the application .is performed after the completion of each unit before integration. This is structural testing that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at the component level and test a specific business process, application and/or system configuration. Unit tests ensure that each unique business process path is executed exactly as documented specification and contains clearly defined inputs and expected results.

12.0.2 White Box Testing

White Box Testing is testing in which the software tester has knowledge of the inner workings, structure, and language of the software, or at least its purpose. It is the purpose. It is used to test areas that cannot be reached from the black box level.

12.0.3 Black Box Testing

The shortcoming of Black Box Testing is testing software without any knowledge of the inner workings, structure, or language of the module under test. Black box tests, like most other kinds of tests, must be written from a definitive source document, such as a specification or requirements document, such as a specification or requirements document. It is testing in which the software under test is considered a black box that you cannot see go on. A test provides inputs and responds to outputs regardless of how the software works.

12.0.4 System Testing

System testing ensures that the entire integrated software system meets the requirements. It tests the configuration to ensure known and predictable results. Example system testing is a configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-controlled process linkages and integration points.

12.0.5 Integration Testing

Integration tests are designed to test integrated software components and determine whether it runs as one program. Testing is event driven and more involved the basic result of screens or fields. Integration tests show that although components were individually satisfied as demonstrated by successful unit tests the combination of components is correct and consistent. Integration testing is specifically aimed at detecting problems that result from the combination of components.

Chapter 13

Result

13.1 OUTCOME

The success of the machine learning heart disease prediction system depends on several factors, including the quality of the database used to train the model, the accuracy of the machine learning algorithm used, and the evaluation criteria used to evaluate the model's performance. If successful, a machine-learning heart disease prediction system could help healthcare providers identify patients at high risk of heart disease, allowing for early intervention and treatment. This can improve patient outcomes and reduce healthcare costs.

Overall, the results of machine-based heart disease prediction systems could have a significant impact on the healthcare industry and potentially improve patient care and outcomes. Logistic regression, KNN classifier, extra trees classifier, this proposed machine learning algorithms will work effectively and accurately predict the heart disease using the efficient data inputs. From all of these logistic regression gives the 86

13.2 HOME PAGE

Once the user fills out the required information, the prediction page utilizes algorithms or risk assessment models to analyze the data and generate a risk

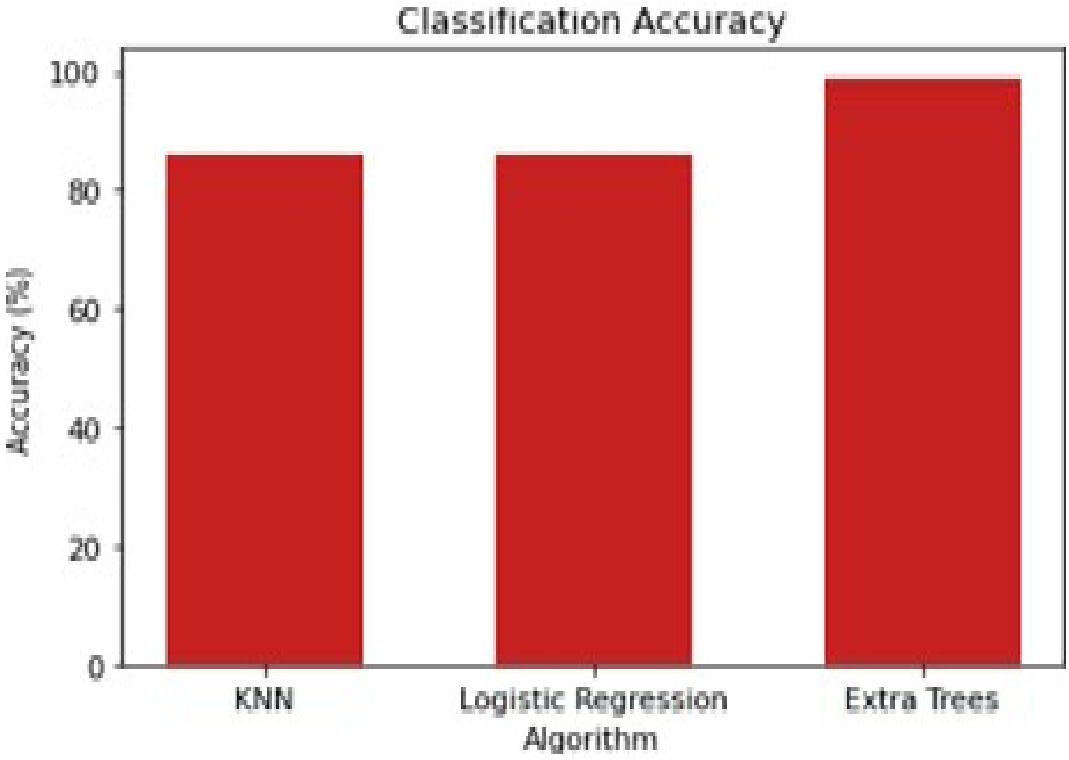


Figure 13.1: Result of proposed System

score or probability of developing heart disease. The output may indicate the user's risk category (low, moderate, high) or provide a numerical score.

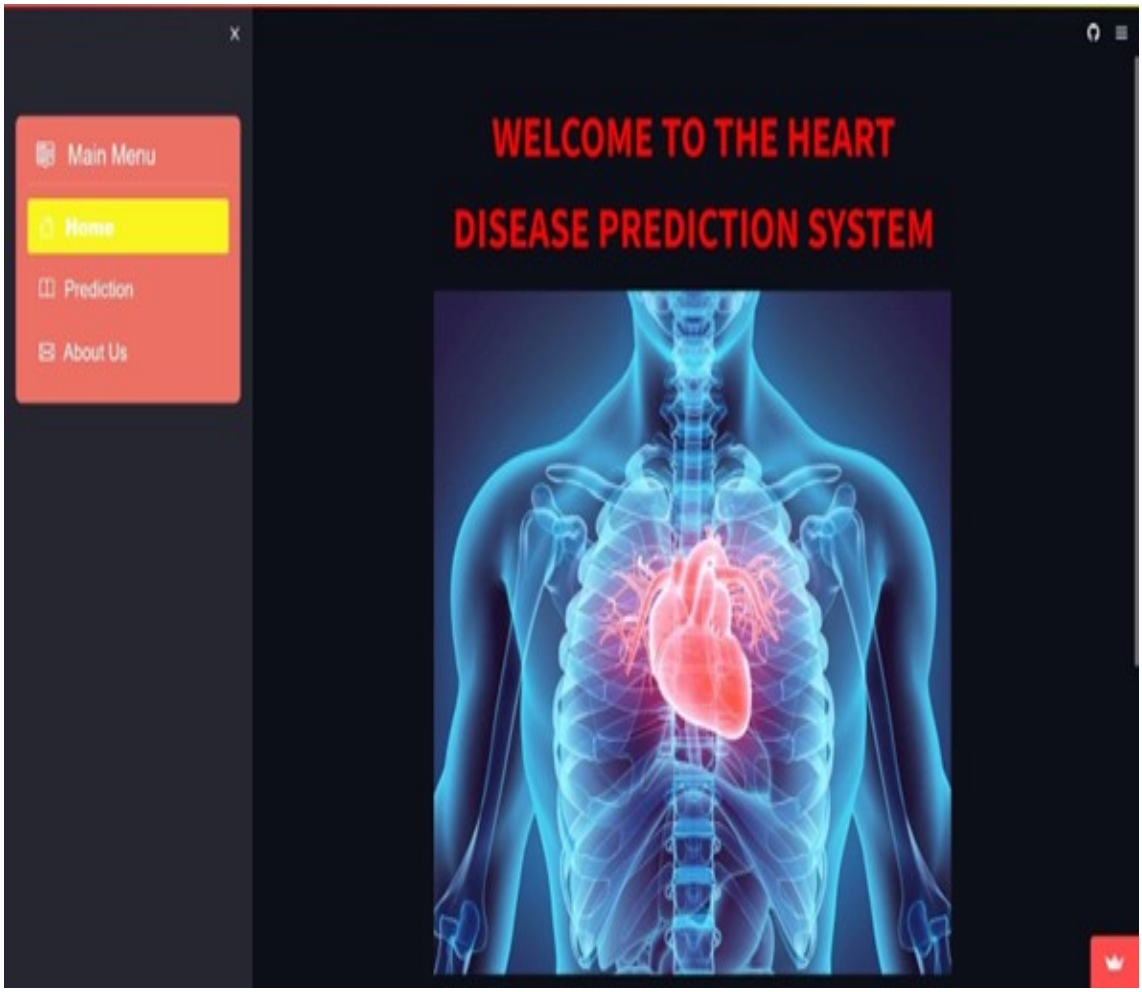


Figure 13.2: Home Page

A heart prediction form is an online or physical form that collects specific information from individuals to assess their risk of developing heart conditions. The form typically includes fields related to personal details such as age, gender, weight, height, lifestyle habits (e.g., smoking, exercise), medical history (e.g., previous heart conditions, family history), and symptoms (e.g., chest pain, shortness of breath). Additionally, the form may ask for specific measurements like blood pressure and cholesterol levels. The purpose of the heart prediction form is to gather relevant information that healthcare professionals or algorithms can use to evaluate an individual's risk of heart conditions. This risk assessment can help identify individuals who may require further medical evaluation, lifestyle modifications, or preventive measures to mitigate the risk of developing heart problems. The form serves as a tool to

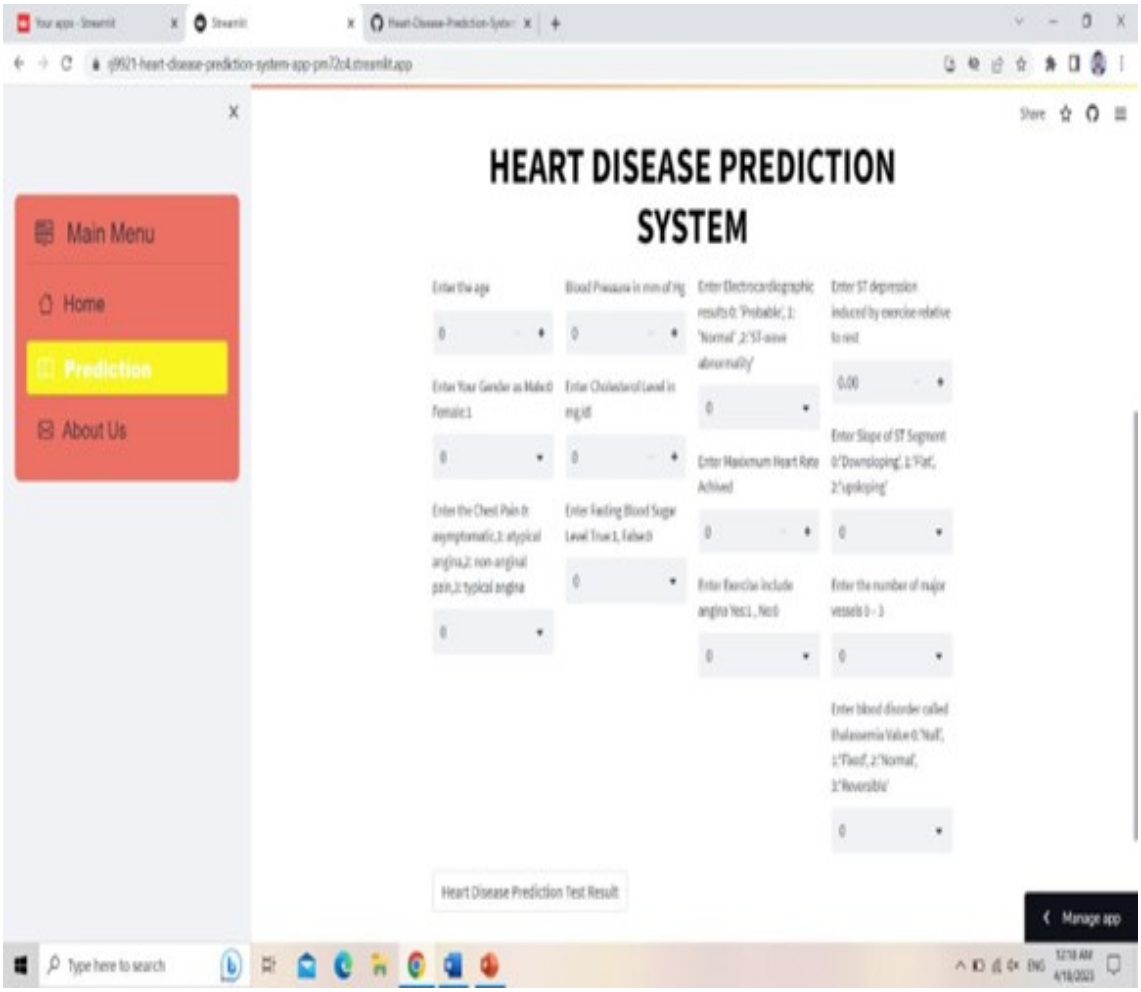


Figure 13.3: Prediction Page

gather important data and facilitate a comprehensive evaluation of an individual's heart health.

Chapter 14

CONCLUSION AND FUTURE SCOPE

Conclusion : Conclusion :

Heart disease is a major killer in India and worldwide, and the use of promising technologies like machine learning for early diagnosis of heart disease will have a huge impact on society. Early diagnosis of heart disease can help make decisions about lifestyle changes in patients at risk, and then reduce complications, which can be an important milestone in the medical field. The number of people suffering from heart disease is increasing every year. This suggests early diagnosis and treatment. The use of appropriate technology support in this area can be very beneficial for the medical fraternity and patients.

The heart is one of the basic and vital organs of the human body, and the prediction of heart diseases is also important interest in human beings, so algorithm accuracy is one parameter for performance analysis algorithms. The accuracy of algorithms in machine learning depends on the data set that was used for training and test purpose. When we analyze the algorithms based on the confusion matrix, we find Extra Trees classifier is the best. For the future scope, the machine learning approach will be used more for the best heart disease analysis.

Logistic regression, KNN classifier, extra trees classifier, this proposed machine learning algorithms will work effectively and accurately predict the heart disease using the efficient data inputs.

Feature Scope:

The future scope of heart disease prediction systems using machine learning is promising. Machine learning algorithms can analyze large amounts of data and recognize patterns that are invisible to human experts. The algorithm can be used to predict a patient's likelihood of developing heart disease by examining various risk factors, including age, gender, blood pressure, cholesterol levels and family history. Overall, the prospect of predicting heart disease through machine learning is exciting, and we can expect significant progress in this field in the coming years.

Chapter 15

References

1. Sunil S. Khatal, Dr. Yogesh Kumar Sharma “Analyzing the role of Heart DiseasePrediction System using IoT and Machine Learning” 2020
2. Chipara, Octav, Chenyang Lu, Thomas C. Bailey, and Gruia-Catalin Roman, “Reliable clinical monitoring using wireless sensor networks: experiences in a step-down hospital unit,” In Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, ACM, pp. 155–168, 2010.
3. Khambete, N. D, and A. Murray, “National efforts to improve healthcare technology managementand medical device safety in India,” Appropriate Healthcare Technologies for Developing Countries, 7th International Conference on, IET, pp. 1–5, 2012.
4. Priyan Malarvizhi Kumar and Usha Devi Gandhi,” A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases”, J. Computers and Electrical Engineering, pp. 222-235, 2018.
5. Mingyu Park and Younghoon Song,JAewonLee,JeongyeupPaek,” Design and implementation of a smart chair system for IOT”, IEEE,2016.
6. P.M. Kumar, S. Lokesh, R. Varatharajan, C. Gokulnath, P. Parthasarathy,” Cloud and IoT baseddisease prediction and diagnosis system for healthcare using Fuzzy neural classifier”, Future Generation Computer Systems,

2018.

7. Trevor. (2018). Enterprise Personal Analytics: A Research Agenda. 10.13025/S88H04.
8. Minerva, R., Biru, A., Rotondi, D. (2015),"IEEE- Towards a Definition of the Internet of Things(IoT)". 9. Bilal Afzal, Muhammad Umair, Ghalib Asadullah Shah, Ejaz Ahmed, "Enabling IoT platforms for social IoT applications: Vision, feature mapping, and challenges", Future Generation Computer Systems, 2017.
9. Luminoso, L. (2017). Creative engineering. Design Engineering (Canada), 63(1), 30-31. [11]Holler, J., T siats is, V., Mulligan, C., Avesand, S., Karnouskos, S., Boyle, D. (2014). From
10. Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence".
11. Koppar, Anant R, and Venu gopalachar Sridhar, "A workflow solution for electronic health records to improve healthcare delivery efficiency in rural India," In eHealth, Telemedicine, and Social Medicine, 2009. eTELEMED'09. International Conference on, pp. 227–232, IEEE, 2009.
12. Sharma YK, Khatal Sunil S. Health Care Patient Monitoring using IoT and Machine Learning,in IOSR Journal of Engineering (IOSR JEN) National Conference on "Recent Innovations in Engineering and Technology" MOMENTUM-19
13. Jha, R. K., Henge, S. K., Sharma, A. (2020). Optimal machine learning classifiers for predictionof heart disease. International Journal of Control and Automation, 13(1 Special Issue), 31-37.
14. Patil PR, Sharma Y, Kshirsagar M. U2R Attack Detection Using Machine Learning. Science[ETEBMS-2016]. 2016 Mar;5:6.
15. Sharma YK, Khan V. A Research on Automatic Handwritten Devnagari Text Generation inDifferent Styles Using Recurrent Neural Network (Deep Learning) Especially for Marathi Script.
16. Monika D.Rokade ,Dr.Yogesh Kumar Sharma,"Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced net-

work traffic." IOSR Journal of Engineering(IOSR JEN), ISSN (e): 2250-3021, ISSN (p): 2278-8719

17. Monika D.Rokade, Dr. Yogesh Kumar Sharma. (2020). Identification of Malicious Activity for Network Packet using Deep Learning. International Journal of Advanced Science and Technology, 29(9s), 2324 - 2331.
18. Monika D.Rokade ,Dr.Yogesh Kumar Sharma"MLIDS: A Machine Learning Approach forIntrusion Detection for Real Time Network Dataset", 2021 International Conference on Emerging Smart Computing and Informatics (ESCI)

Chapter 16

Laboratory assignments on Project Analysis of Algorithmic Design

Aim: To develop the problem under consideration and justify feasibility using concepts of knowledge canvas and IDEA Matrix.

Learning has two aspects:

Knowledge Canvas:-

Knowledge is a familiarity, awareness or understanding of someone or something, such as facts, information, descriptions, or skills, which is acquired through experience or education by perceiving, discovering, or learning. The canvas is com-

Delivery Channels	Inquiring Channels	Knowledge Mining	Knowledge Promotion
Necessary Knowledge	Knowledge Customers	Knowledge Owner/Container	Available Knowledge
Expected Benefits	Knowledge Rating	Sharing Rewards	Costs

Figure 15.1: Knowledge Map

posed by 12 blocks. The right side describes the available knowledge, the left side describes the needed knowledge. From top to bottom you can read aspects of knowledge transmission (communication and technology), then who needs/owns what and finally costs and benefits of knowledge for the organization and people, in few words you can read the how, what and why.

IDEA Matrix:

It improves and increases capabilities that are necessary while it decreases and deletes pointers that are unnecessary. To measure these two aspects in systematic way, we have introduced Learning IDEA Matrix. A solid framework for learning opportunity evaluation and knowledge innovation is required otherwise locating ML opportunity would become more difficult. To build and practice new paradigms

of machine learning with Knowledge Innovation IDEA Matrix framework is developed. This framework is applied to more than one dozen successful ML projects in different domains to locate ML opportunities. This includes industries from financial advisories, agriculture, teaching learning, and health care. This section introduces this learning IDEA framework for systemic evaluation of problem to identify and evaluate ML opportunities. This framework focuses on highest leverage points of knowledge building and knowledge flow optimization to locate and evaluate machine learning opportunities. The framework is based on flexibility and simplicity while applying the knowledge concepts. Learning IDEA matrix is based on learning experiments. Here idea is that one should identify opportunities to evaluate and improve learnability and effectively the over all performance. Idea matrix helps to identify learning problems. Figure depicts IDEA framework for machine learning and systemic knowledge innovation. It looks for different parameters marked by I, D, E, and A depicting impact and need of machine learning.

Conclusion:-

We justify feasibility using concepts of knowledge canvas and IDEA Matrix.The problem gives feasible solution.

I	D	E	A
INCREASE: Effectiveness and efficiency for authentication system as per location.	**DRIVE:** Calculate efficiency of visual authentication system at farmer side.	**EDUCATE:** Identify the challenges of security of various field like dataset, disease prediction system as per crop.	**ACCELERATE:** Low energy and bandwidth consumption, efficient storage and parallel authentication.
IMPROVE: High data trustworthiness with the help of system.	**DELIVER:** Efficient techniques for farmers to improve his farming activity and increase maximum income.	**EVALUATE:** Incorporate DataStream (dataset) binding.	**ASSOCIATE:** Only authorized parties i.e. BS can process and check integrity for system.
IGNORE: Different types of fake users and its data.	**DECREASE:** Time and Money.	**ELIMINATE:** Unauthorized parties to check integrity for system and dataset security.	**AVOID:** Number of diseases as per weather conditions & Prevention and Protection (of crop) related problems.

Table 15.1: IDEA Matrix

Chapter 17

Laboratory assignments on Project Quality and Reliability Testing of Project Design

Aim:-

Use of divide and conquer strategies to exploit distributed/parallel/concurrent processing of the above to identify object, morphisms, overloading in functions (if any), and functional relations and any other dependencies (as per requirements). It can include Venn diagram, state diagram, function relations, i/o relations; use this to derive objects, morphism, overloading

Divide and conquer:-

In computer science, divide and conquer (DC) is an algorithm design paradigm based on multi-branched recursion. A divide and conquer algorithm works by recursively breaking down a problem into two or more sub-problems of the same or related type, until these become simple enough to be solved directly. The solutions to the sub-problems are then combined to give a solution to the original problem.
This divide and conquer technique is the basis of efficient algorithms for all kinds of problems, such as sorting (e.g., quicksort, merge sort), multiplying large numbers (e.g. the Karatsuba algorithm), finding the closest pair of points, syntactic analysis (e.g., top-down parsers), and computing the discrete Fourier transform (FFTs).

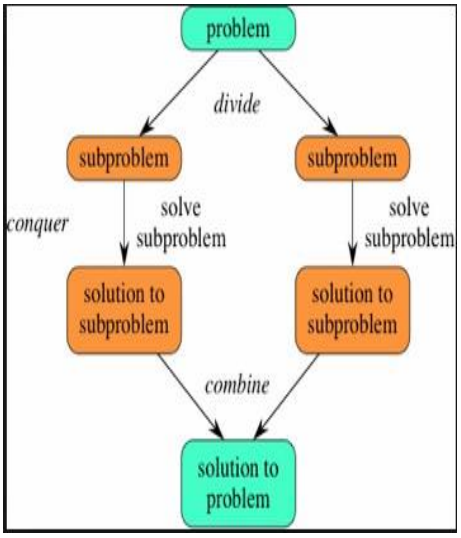


Figure 16.1: Divide and conquer diagram

It Divides Into Following Modules:-

[1] Bluring

In Bluring, we simple blur an image. An image looks more sharp or more detailed if we are able to perceive all the objects and their shapes correctly in it. For example An image with a face, looks clear when we are able to identify eyes,ears,nose,lips,forehead e.t.c. very clear. This shape of an object is due to its edges.So in bluring, we simple reduce the edge content and makes the transition form one color to the other very smooth.

[2] Thresholding

Thresholding is an simplest method of the image segmentation. From a grayscale image,thresholding can be used to create binary images.

[3] Blob Detection

In computer vision, blob detection methods are aimed at detecting regions in a digital image that differ in properties, such as brightness or color,compare to surrounding regions.blob is regions of an image in which some properties are constant or approximately constant.

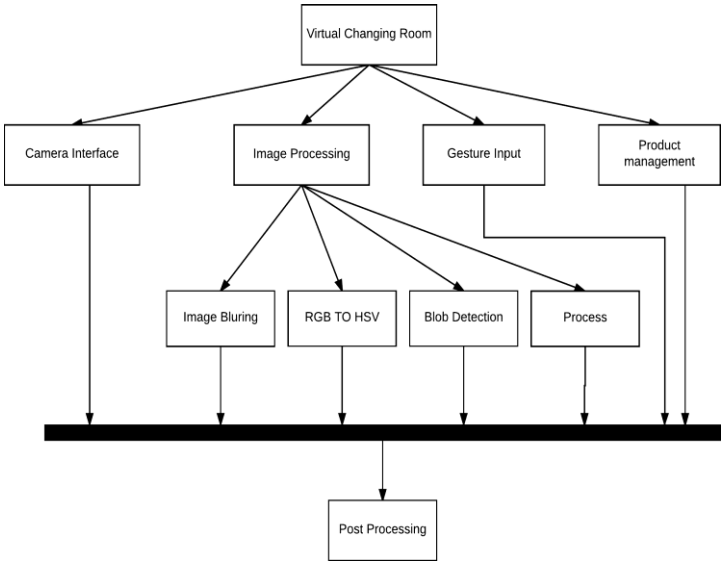


Figure 16.2: Divide and conquer strategy

[4] **R**GB to HSV Conversion:-

The images generally follows the RGB model (Red, Green Blue) but this model does not provide the higher level of accuracy that we want in system so there is need to convert the RGB to HSV(Hue, Saturation, Value) model as it provides higher level of accuracy.

16.1 COMPUTATIONAL COMPLEXITY CLASSES THEORY

When solving problems we have to decide the difficulty level of our problem. There are three types of classes provided for that. These are as follows:

- P Class
- NP Complete Class
- NP hard Class

P Class:-

In computational complexity theory, P, also known as PTIME or DTIME(nO(1)), is a fundamental complexity class. It contains all decision problems that can be solved by a deterministic Turing machine using a polynomial amount of computation time, or polynomial time.P class problems are deterministic problems i.e. P class problems can be solve by deterministic Turing Machine.

NP class:-

In computational complexity theory Equivalently, the formal definition of NP is the set of decision problems solvable in polynomial time by a theoretical non-deterministic Turing machine. This second definition is the basis for the abbreviation NP, which stands for ”non deterministic,polynomial time.” However, the verifier-based definition tends to be more intuitive and practical in common applications compared to the formal machine definition. The two definitions are equivalent because the algorithm for the machine definition consists of two phases, the first of which consists of a guess about the solution, which is generated in a non-deterministic way, while the second phase consists of a deterministic algorithm that

verifies or rejects the guess as a valid solution to the problem.

16.2 WHY PROJECT SHOULD BE A P CLASS ?

A Problem is assigned to the P(Polynomial time) class if there exists at least one algorithm to solve the problem. Such that the number of steps of the algorithm is bounded by a polynomial in N. where N is the length of the input. In our project obstacle detection is avoided and also.

Computational complexity of our project is O(N$\hat{2}$). Where N is the Number of pixels of the screen or image. Since our Problem has a solution and is solvable in polynomial time hence its a P problem.

Chapter 18

PLAGIARISM REPORT

Heart Disease Prediction System.pdf

ORIGINALITY REPORT

29%	22%	9%	20%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Savitribai Phule Pune University Student Paper	7%
2	www.ijitee.org Internet Source	2%
3	Palak Rai, Reeya Gupta, Vinicia Dsouza, Dipti Jadhav. "Virtual Canvas for Interactive Learning using OpenCV", 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), 2022 Publication	2%
4	www.scilit.net Internet Source	1%
5	happydaycards.org Internet Source	1%
6	Submitted to University of North Georgia Student Paper	1%
7	www.researchgate.net Internet Source	1%

Chapter 19

APPENDIX

19.1 PUBLISHED REVIEW PAPER

- 1. International Journal of Innovative Reasearch. (IJIRCCE)

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205257 |

Heart Disease Prediction System Using a Machine Learning

Moin H. Khan, Ayaan S. Shakil, Samruddhi S. Kanade, Prof. Kapil Dere

UG Student, Department of C.S, Jaihind College of Engineering, Kuran, Pune, India

Assistant Professor, Department of C.S. Jaihind college of Engineering, Kuran, Pune, India

ABSTRACT: Heart disease prediction is very essential in today's environment, various researches has already done to predict heart disease from large dataset. IoT environment basically generate data from different sensors and predict the disease possibility accordingly. Various synthetic data sets content different body parameters which are extracted by specific sensor values, the major role played by machine learning algorithm. In this research we propose heart disease prediction with the combination of IoT and machine learning approach, the IoT environment has established to extract the data from real-time Body Sensor Network (BSN) with intermediate sensing System and store data in the cloud server adequately. Such audit data has considered synthetic information which is basically used to predict heart disease possibility. In this research, we illustrate various machine learning algorithms as well as some deep learning algorithms to achieve drastic supervision for diseas e prediction. The experimental analysis shows the effectiveness of proposed deep learning classification algorithms over the classical machine learning algorithms.

KEYWORDS: Disease prediction system, IoT, machine Learning, Supervised learning, NLP, Heart Disease.

I. INTRODUCTION

The field of "telemedicine" refers to the fast evolution of a services in recent years that enable wearable technologies to facilitate wireless communication between a physician and a patient .As of late, diabetes is the leading cause of death worldwide .In the year 2000, 171 million It was projected that there would be up to 642 million people on the planet by 2040. This rise in the number demands that this illness be taken seriously. Billionaires are spent on diabetes treatment by numerous hospitals worldwide. Type 1 diabetes, pre-diabetes, Type 2 diabetes, and gestational diabetes are the four categories of diabetes patients .Type 1 diabetes was brought on by adults' and children's insufficient insulin. Pre-diabetic refers to the stage that precedes Type 2 and gestational diabetes in pregnant women.

Furthermore, there is an infinite list of disorders that are directly linked to the heart; according to the International Cardiac Society, there are over 15 different kinds of heart-related ailments. These illnesses just need minimal historical data to be directly traced. However, conditions like diabetes, cancer, and TB, The term "indirectly related to heart diseases" refers to these. Careful historical monitoring and observational pattern analysis of the ECG waveforms are necessary for these disorders. Usually, the actions listed below are taken to complete this task:

There are also some existed models for diabetes prediction. El Jerjawi et al [2] established a neural network model for diabetes prediction. They used some attributes such as PG Concentration (Plasma glucose at 2 hours in an oral glucose tolerance test), Diastolic BP (Diastolic Blood Pressure (mm Hg)). Most of them need some professional medical test, so it is not accessible for every person. Also, the final accuracy of the prediction is around 87 percent . Peter W.F et al [3] also uses the regression models to make predictions for diabetes mellitus. However, in its samples, 99 percent of them are white and non-Hispanic which does not include other races. Hence, we believe the model is not representative.

We selected characteristics like abrupt weight loss and obesity, which are easier to comprehend and use, to build our prediction model in this study. The patients are not need to perform certain medical tests, which improves the readability and applicability of our approach. Building upon the six machine learning models outlined above, we develop diabetes prediction models. In terms of the testing error, we also compare how well they performed. As a result, we discover that neural networks and boosting have the best accuracy, at 95.5% and 96.1%, respectively.

II. LITERATURE SURVEY

According to Sunil S. Khatal & Dr. Yogesh Kumar Sharma in “Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning.’’ Prediction of heart disease is very important in today's environment; various studies

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205257 |

have been done to predict heart disease from large database. The IoT environment basically generates data from various sensors and predicts the probability of disease accordingly. A variety of synthetic data, the main role of machine learning algorithms, is a variety of body parameters obtained by certain sensor values. In this study, we propose a combination of IoT and machine learning approaches to predict heart disease, an IoT environment designed to extract data from real-time body sensor networks (BSN) and remote sensing systems and store the data efficiently in a cloud server. Such studies are considered synthetic data used to predict the likelihood of heart disease. In this study, we present a variety of machine learning algorithms as well as some deep learning algorithms to achieve better control over the prediction of various diseases. Experimental analysis shows that deep learning classification algorithms are more effective than classical machine learning algorithms.

According to Chipara et. all in "Reliable clinical monitoring using wireless sensor networks: experiencesin a step-down hospital unit" presents the design, deployment, and empirical investigation of a wireless clinical monitoring system that collects pulse and oxygen saturation readings from patients. The main contribution of this paper is an indepth clinical trial evaluating the potential of a wireless sensor network for patient monitoring in a public hospital. We present a detailed analysis of system reliability in a seven-month long-term hospital setting involving 41 patients in a step-up cardiology unit. The network achieves high reliability (average 99.68%, range 95.21% - 100%). The overall reliability of the system was similar to pulse oximeters (mean 80.85%, range 0.46% to 97.69%). Sensing errors usually occur in short bursts, but there are longer durations due to sensor cutoff. We show that sensitivity can be significantly increased by implementing a connectivity alarm system that maximizes reliability and minimizes intervention costs. Retrospective data analysis showed that the system provided adequate temporal resolution to ensure transfer to the intensive care unit in three patients who experienced clinically significant events. These results show the potential and promise of using wireless sensor networks for continuous patient monitoring and detection of clinical deterioration in general hospital wards.

According to Khambete, N. D & A. Murray in "National efforts to improve healthcare technology management and medical device safety in India" In the practice of contemporary modern medicine, effective and safe use of healthcare technology is acknowledged worldwide as essential for any healthcare system. Achieving these goals can be particularly challenging in developing countries such as India, where an estimated 75% of medical technology is imported and studies have shownthat almost 30% of medical equipment is out of service. Furthermore, concerns regarding medical equipment safety have been raised in newspaper reports and also reported by a pilot study. However, recently, substantial efforts are being made to introduce changes in the health care system that will help in improving this situation. Discussions on these issues were initiated at two `International Clinical Engineering Workshops' (Trivandrum 2009 and Pune 2011) and two `Regional Clinical Engineering Workshops' (Latur and Mumbai 2011). A clear consensus emerged from these Workshops that urgent action was essential to initiate effective Healthcare Technology Management (HTM) practices in all health care sectors and actively promote medical device and equipment safety in India. Subsequently ,in February 2012, a round table meeting of experts was held, which focused on confronting medical and healthcare management staff with the problems to be solved, while at the same time helping to develop an action plan to bring about the necessary changes. These consultations identified existing gaps and underlying reasons, thus leading to development of an action plan. This paper reviews all these efforts and highlights the outcomes.

III. MODELS

In this section, we use logistic regression, support vector machine, decision tree, random forest, boosting and neural network to establish diabetes prediction models. We evaluate the model by calculation the train error and test error.

Logistic Regression:
Logistic regression models the likelihoods of potential outcomes using logistic functions. The dependent variable in a binary logistic regression model is divided into two groups. It determines the linear relationship by using the logit function to transform the probability, which is between 0 and 1, to any real integer .We must first ascertain the logistic coefficients. The table shows that there are positive correlations between diabetes and polyuria, polydipsia, sudden weight loss, weakness, polyphagia, genital thrush, visual blurring, irritability, and partial paresis; on the other hand, there are negative correlations between diabetes and age, itching, delayed healing, muscle stiffness, alopecia, and obesity. Table 4 provides an overview of the confusion matrix. There appears to be no over fitting because the train accuracy of 87.6% and the test accuracy of 87.2% are close to each other.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205257 |

Decision Tree:
A decision tree is a tree-like structure that is constructed by dividing the source set in order to develop a model that makes predictions depending on input variables. Examples are categorized by a collection of dividing rules according to the characteristics of classification. One of the most often used machine learning algorithms is decision trees, which have the advantages of simplicity and comprehensibility.

Random Forest:
Because a single decision tree is unstable, ensemble learning—which mixes several models to increase overall prediction accuracy and decrease the variation. A common ensemble learning technique is random forest. Since the complexity of the model is determined by the number of decision trees, we experiment with different numbers of trees and plot the resulting mistakes in Figure 4. In contrast to SVM and decision trees, random forests' testing errors are unlikely to increase as the number of trees increases, suggesting that they are less prone to overfit. It is unlikely to be an overfitting model, comparatively, even while the training error keeps decreasing and the testing error remains stable after declining. The confusion matrix is Compared with single decision tree, the testing accuracy of random forest, which is 94.9%, increased near 1.7 percent.

Boosting:
Another traditional ensemble learning technique is called "boosting," which lowers variance by reweighting the samples for each decision tree while it is being trained. similar to the arbitrary Moreover, we also utilize the quantity of decision trees to ascertain the model's and the outcome's complexity. Boosting is slightly more likely to be overfitting than random forest. The train error first decreases and then straightens out. Similar to other models, the test error first decreases and subsequently increases. The image indicates that we selected 700 trees for our boosting model, and Table 8 displays the confusion matrix. The fact that the train correctness is nearly 100% indicates that overfitting is more likely to be the cause of the boosting. Compared to the random forest,its accuracy of boosting is further improved by around 0.7 percent.

Support Vector Machine:
In this method, The regularization parameter C, which regulates the model's complexity, has to be changed. Greater C indicates a harsher punishment for the misclassification, which causes More likely, the model is overfitting. In order to compute the training and testing errors for each model with a different C, we employ several values of C, ranging from 0.1 to 4.6. The results are presented in Figure 2. The train error decreases as the model gets more complicated, while the test error initially decreases and subsequently increases. As a result, we determine that 2.1 is the ideal parameter for our model, and the matching confusion matrix that results is shown in Table 5. Consequently, the testing accuracy of the support vector machine is higher than that of the logistic regression (87.2%) at 90.4%.

Neural Network:
Neural network is made up of layers and neurons, and it learns by processing samples from first layer to the last layer. Once enough samples have been processed, the algorithm can construct the right model. Each of the three layers we built has 64, 16 or 8 neurons in it. Additionally, we select the logistic activation function and the 0.01 learning rate. Next, we obtain the Table 9 confusion matrix. As we can see, the neural network outperforms all other models with an accuracy of 96.2, owing to its larger parameter set that allows for better data fitting.

IV. CONCLUSION AND FUTURE WORK

Physicians can identify patients more accurately and treat them more quickly if they have access to an effective diabetes prediction model. We use descriptive statistics to assess the risk of diabetes.predictive dataset to look into the factors that affect diabetes. Table 10 lists the train and test errors of the six machine learning models—logistic regression, support vector machines, decision trees, random forests, boosting, and neural networks—on which we base our diabetes prediction models.Compared to the final three models, which are more sophisticated, the first three models—logistic regression, support vector machines, and decision trees—are simpler, intuitive, and have lesser accuracy. As we can see, test accuracy increases with model complexity. In the future, we can experiment with models that are more capable of learning and adapting, and we can employ a wider range of datasets to increase forecast accuracy.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205257 |

REFERENCES

1. Sunil S. Khatal, Dr. Yogesh Kumar Sharma "Analyzing the role of Heart DiseasePrediction System using IoT and Machine Learning" 2020
2. Chipara, Octav, Chenyang Lu, Thomas C. Bailey, and Gruia-Catalin Roman, "Reliable clinical monitoring using wireless sensor networks: experiences in a step-down hospital unit," In Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, ACM, pp. 155–168, 2010.
3. Khambete, N. D, and A. Murray, "National efforts to improve healthcare technology managementand medical device safety in India," Appropriate Healthcare Technologies for Developing Countries, 7th International Conference on, IET, pp. 1–5, 2012.
4. Priyan Malarvizhi Kumar and Usha Devi Gandhi," A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases", J. Computers and Electrical Engineering, pp. 222-235, 2018.
5. Mingyu Park and Younghoon Song,JAewonLee,JeongyeupPaek," Design and implementation of a smart chair system for IOT", IEEE,2016.
6. P.M. Kumar, S. Lokesh, R. Varatharajan, C. Gokulnath, P. Parthasarathy," Cloud and IoT baseddisease prediction and diagnosis system for healthcare using Fuzzy neural classifier", Future Generation Computer Systems, 2018.
7. Koppar, Anant R, and Venu gopalachar Sridhar, "A workflow solution for electronic health records to improve healthcare delivery efficiency in rural India," In eHealth, Telemedicine, and Social Medicine, 2009. eTELEMED'09. International Conference on, pp. 227–232, IEEE, 2009
8. Sharma YK, Khatal Sunil S. Health Care Patient Monitoring using IoT and Machine Learning,in IOSR Journal of Engineering (IOSR JEN) National Conference on "Recent Innovations in Engineering and Technology" MOMENTUM-19
9. Jha, R. K., Henge, S. K., & Sharma, A. (2020). Optimal machine learning classifiers for predictionof heart disease. International Journal of Control and Automation, 13(1 Special Issue), 31-37.J. Y.

19.2 CERTIFICATES

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Referred, Scholarly Indexed, Open Access Journal Since 2013)

IJIRCCE

Impact Factor 8.379

CERTIFICATE

OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

AYAAN S. SHAKIL

UG Student, Dept. of C.S, Jaihind College of Engineering, Kuran, Pune, India

in Recognition of Publication of the Research Paper Entitled

"Heart Disease Prediction System Using a Machine Learning"

in IJIRCCE, Volume 12, Issue 5, May 2024

Google scholar Crossref Mendeley doi INNO SPACE SJIF Scientific Journal Impact Factor

e-ISSN: 2320-9801
p-ISSN: 2320-9798

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

www.ijircce.com ijircce@gmail.com

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Referred, Scholarly Indexed, Open Access Journal Since 2013)

IJIRCCE

Impact Factor 8.379

CERTIFICATE

OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

MOIN H. KHAN

UG Student, Dept. of C.S, Jaihind College of Engineering, Kuran, Pune, India

in Recognition of Publication of the Research Paper Entitled

"Heart Disease Prediction System Using a Machine Learning"

in IJIRCCE, Volume 12, Issue 5, May 2024

Google scholar | Crossref | Mendeley | doi | INNO SPACE SJIF Scientific Journal Impact Factor

e-ISSN: 2320-9801
p-ISSN: 2320-9798

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

www.ijircce.com ijircce@gmail.com


International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Referred, Scholarly Indexed, Open Access Journal Since 2013)
IJIRCCE
Impact Factor 8.379
CERTIFICATE
OF PUBLICATION
The Board of IJIRCCE is hereby awarding this certificate to
SAMRUDDHI S. KANADE
UG Student, Dept. of C.S, Jaihind College of Engineering, Kuran, Pune, India
in Recognition of Publication of the Research Paper Entitled
"Heart Disease Prediction System Using a Machine Learning"
in IJIRCCE, Volume 12, Issue 5, May 2024
Google scholar
Crossref
Mendeley
doi
INNO SPACE
SJIF Scientific Journal Impact Factor
e-ISSN: 2320-9801
p-ISSN: 2320-9798
ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA
Editor-in-Chief
www.ijircce.com
ijircce@gmail.com


International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Referred, Scholarly Indexed, Open Access Journal Since 2013)
IJIRCCE
Impact Factor 8.379
CERTIFICATE
OF PUBLICATION
The Board of IJIRCCE is hereby awarding this certificate to
PROF. KAPIL DERE
Assistant Professor, Dept. of C.S, Jaihind College of Engineering, Kuran, Pune, India
in Recognition of Publication of the Research Paper Entitled
"Heart Disease Prediction System Using a Machine Learning"
in IJIRCCE, Volume 12, Issue 5, May 2024
Google scholar
Crossref
Mendeley
doi
INNO SPACE
SJIF Scientific Journal Impact Factor
e-ISSN: 2320-9801
p-ISSN: 2320-9798
ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA
Editor-in-Chief
www.ijircce.com
ijircce@gmail.com

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Heart Disease Prediction System using Machine Learning

Khan Moin, Shaikh Ayaan, Kanade Samruddhi,Prof.K.D.Dere

Jaihind College of Engineering, Kuran
Jaihind College of Engineering, Kuran
Jaihind College of Engineering, Kuran
Jaihind College of Engineering, Kuran

Abstract - Predicting cardiac disease is crucial in today's world, and several studies have already been conducted using sizable datasets. In essence, an IoT ecosystem gathers data from many sensors and uses that data to forecast the likelihood of illness. A machine learning algorithm plays a vital role in extracting distinct body characteristics from a variety of synthetic datasets that contain diverse sensor values. In this study, we suggest using an Internet of Things (IoT) and machine learning technique to forecast cardiac disease. The IoT environment has been set up to collect data from real-time Body Sensor Networks (BSNs) via intermediary sensing systems and appropriately store it on cloud servers. Such audit data has considered synthetic information, which is primarily utilized to forecast the likelihood of heart disease. In this study, we demonstrate a number of to accomplish extreme supervision for design and prediction, machine learning techniques and certain deep learning algorithms are used. The comparative study of experimental data demonstrates the superiority of the suggested deep learning classification algorithms over the traditional machine learning techniques.

Key Words: NLP, supervised learning, machine learning, IoT, disease prediction system, and heart disease.

1.INTRODUCTION

"Telemedicine" is the rapidly developing sector of services that allow wearable technology to permit wireless contact between patient and a physician in recent years. Nowadays, diabetes is the world's biggest cause of mortality. In 2000, there were 171million By 2040, it was predicted that there might be as manyas642 million people on Earth. The increase in numbers necessitates a serious approach to this ailment. Hospitals all across the world treat diabetes with billions of dollars. The four types of diabetic patients include type 1 diabetes, pre-diabetes, Type 2 diabetes, and gestational diabetes. Insulin deficiency in both adults and children caused type 1 diabetes. The temper- diabetic" describes the condition that exists in pregnant women before Type 2 and gestational diabetes. Moreover, the number of conditions that are directly related to the heart is endless; the International Cardiac Society lists over 15 distinct types of heart-related illnesses. For certain disorders to be directly traced, very little historical data is required. Nonetheless, the phrase "indirectly related to heart diseases "applies to ailments including diabetes, cancer, and tuberculosis. For these conditions, careful history monitoring and observational pattern analysis of the ECG waveforms are required. Typically, to do this work, the following steps are taken: Additionally, certain diabetes prediction algorithms already exist. A neural network model was developed by El Jarawa et aliform the prediction of diabetes. They made use of several characteristics, including diastolic blood pressure (mmHg) and plasma glucose concentration (PG Concentration, or plasma glucose at two hours in an oral glucose tolerance test). Since most of them require a professional medical examination, not everyone can access them. Furthermore, the prediction's ultimate accuracy is around 87% accurate. Regression models are also used by Peter W.F. et al.to forecast diabetes mellitus. Its samples, however, do not contain any members of other races; 99 percent of them are white and non-Hispanic. Asa result, we think the model is not typical.

In this study, we built our prediction model using attributes that are easy to understand and use, such as obesity and sudden weight loss. The readability and application of our technique enhanced by the fact that the patients are not required to undergo certain medical testing. We create diabetes prediction models by expanding on the six machine learning models that were previously mentioned. We also compare their performance in terms of testing error. Consequently, we find that, at 95.5% and 96.1%, respectively, neural networks and boosting have the greatest accuracy.

2. Literature Survey

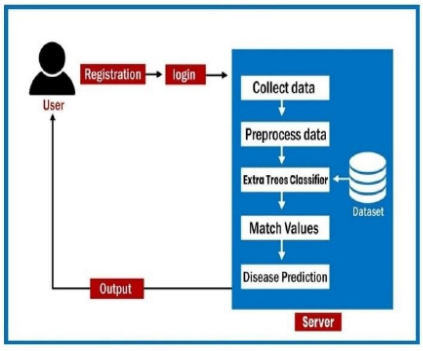
In "Analyzing the role of Heart Disease Prediction System using It and Machine Learning," Sunil S. Khalat and Dr. Yogesh Kumar Sharma state. In the modern world, heart disease prediction is crucial. Numerous studies have been conducted using massive databases to predict heart disease. In essence, the Internet of Things

ecosystem gathers data from a variety of sensors and uses that data to forecast the likelihood of sickness. Different body characteristics that are collected by specific sensor readings constitute a variety of synthetic data, which is the primary function of machine learning algorithms. In this work, we present an IoT environment built to efficiently store and extract data from real-time body sensor networks (BSN) and distant sensing devices in order to forecast cardiac disease using mix of IoT and machine learning methodologies. Studies of this kind are seen as using artificial intelligence to forecast the risk of heart disease. To get more control over the prediction of different illnesses, we provide a range of machine learning methods in this work along with a few deep learning techniques. Deep learning classification methods outperform traditional machine learning algorithms, according to experimental study.

The design, implementation, and empirical investigation of wireless clinical monitoring system that gathers patient pulse and oxygen saturation readings are presented by Chipara et al. in "Reliable clinical monitoring using wireless sensor networks: experiences in a step-down hospital unit." This paper's primary contribution is a comprehensive clinical experiment assessing wireless sensor network's potential for patient monitoring in public hospital. We offer a thorough examination of the system dependability in a long-term hospital context spanning seven months, with 41 patients in a step-up cardiac unit. High dependability is attained by the network (average 99.68%, range95.21% - 100%). The system's overall dependability (mean80.85%, range 0.46% to 97.69%) was comparable to that of pulse oximeters. Sensor cutoff causes extended periods of time when there are sensing failures, which often happen in brief bursts. We demonstrate that the implementation of a connection alert system that optimizes dependability and reduces intervention costs may lead to a large improvement in sensitivity. An review of retrospective data revealed that for three patients who had clinically relevant episodes, the system offered sufficient temporal precision to guarantee transfer to the intensive care unit. These findings highlight the promise and potential of employing wireless sensor networks in regular hospital wards for ongoing patient monitoring and clinical deterioration identification.

In —National initiatives to promote healthcare technology management and medical device safety in Indial, Khambete, N. D. & A. Murray state It is widely known in the practice of modern medicine that any healthcare system must have an efficient and safe use of healthcare technology. Reachingtheseobjectives can be especially difficult in developing nations like India, where studies have revealed that about 30%of medical equipment is out of service and where an estimated 75%of medical technology is imported. Moreover, a pilot study and newspaper publications have also expressed worries about the safety of medical equipment. On the other hand, significant attempts have been undertaken recently to implement adjustments in the healthcare system that will aid in improving this circumstance. Two "International Clinical Engineering Workshops" (Pune, 2011) and two "Regional Clinical Engineering Workshops" (Latur, Mumbai, 2011) served as forums for discussion of these topics.

Strong consensus was reached regarding the necessity of taking immediate action to introduce Healthcare Technology Management (HTM) methods that work across all healthcare sectors and actively promote equipment and medical device safety in India. A round table discussion of specialists was subsequently arranged for February 2012, with the aim of presenting the issues to be resolved to medical and healthcare management personnel and assisting in the creation of anatropal to implement the required adjustments. The fundamental causes and current gaps were determined by these conversations. This essay examines each of these initiatives and emphasizes their results.



3. Model

Logistic Regression:

In this part, we create diabetes prediction models using logistic regression, support vector machines, decision trees, random forests, boosting, and neural networks. By computing the train and test errors, we assess the model.

Logistic Regression: This statistical technique uses logistic function to describe the probabilities of possible outcomes. In binary logistic regression model, the dependent variable is split into two categories. The logit function is used to convert the probability, which is between 0 and 1, to any real number in order to calculate the linear connection. First, the logistic coefficients need to be determined. The table indicates that polyuria, polydipsia, abrupt weight loss, weakness, polyphagia, vaginal thrush, visual blurring, irritability, and partial paresis are positively correlated with diabetes; Conversely, diabetes has been found to have adverse associations with ageing, baldness, delayed healing, itching, and muscular stiffness. An overview of the confusion matrix is given in Table 4. Given the proximity of the test accuracy of 87.2%and the train accuracy of 87.6%, there does not seem to be any overfitting.

Decision Tree:

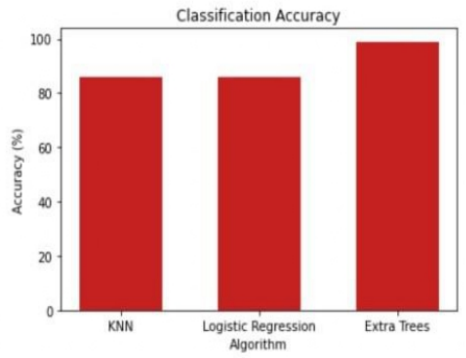
To create a model that generates predictions based on input variables, the source set is divided to create a decision tree, which is a structure like a tree. A set of division rules is used to classify examples based on the features of categorization. Decision trees are among the most popular machine learning algorithms due to their ease of use and readability. Random Forest: Ensemble learning, which combines many models to improve overall prediction accuracy and lower variance, is used since a single decision tree is unstable. Random forest is a popular approach for ensemble learning. We experiment with various numbers of decision trees since the number of trees determines the model's complexity.

Random forests:

It appears to be less prone to overfitting than SVM and decision trees since their testing errors are not expected to rise with the number of trees. Comparatively speaking, even if the training error continues to decline and the testing error stabilizes after dropping, the model is unlikely to be overfit. The matrix of confusion is The testing accuracy of random forests, whichis94.9%, rose by around 1.7% when compared to single decision trees. Increasing by reweighting the data for each decision tree during training, "boosting," another well-known ensemble learning strategy, reduces variance. comparable to the arbitrary In addition, we use the number of decision trees to determine the complexity of the model and the result. Increasing is a little bit more likely to be overfitting than haphazard. The train error straightens up after first declining. The test error grows after initially decreasing, much like previous models. As you can see from the figure, we chose 700 trees for our boosting model. It seems more likely that overfitting is the source of the boosting because the train accuracy is almost 100%. Its boosting accuracy is further enhanced by approximately 0.7 percent as compared to the random forest.

Helping Vector Computer:

This approach requires you to adjust the regularization parameter C, which controls the complexity of the model. Greater C denotes a more severe penalty for the incorrect categorization, which results in The model is most certainly overfitting. Firstly, to calculate the For training and testing errors, we use several values of C, ranging from 0.1 to 4.6, for each model. Presents the findings. As the model becomes more complex, the trainer or drops while the test error first falls and then increases. We thus conclude that 2.1 is the optimal value for our model, andTable5displays the corresponding confusion matrix that is produced. As a result, at 90.4%, the support vector machine's testing accuracy surpasses that of the logistic regression (87.2%). Neural Network: Samples are processed from the to player to the bottom layer by neurons and layers that work together to form a neural network. After processing enough samples. The appropriate model may be built using the algorithm. Thereare64, 16, or 8 neurons in each of the three levels that we constructed. We also choose the 0.01 learning rate anthologist activation function. The Table 9 confusion matraite obtained. With an accuracy of 96.2, the neural network performs better than all other models, as we can see, because of its bigger parameter set, which enables better data fitting.



4. Conclusion

If doctors have access to an efficient diabetes prediction model, they can diagnose patients more precisely and treat them faster. To evaluate the risk of diabetes, we employ descriptive statistics. Predictive dataset to investigate the variables influencing diabetes. The six machine learning models—logistic regression, support vector machines, decision trees, random forests, boosting, and neural networks—that serve as the foundation for our diabetes prediction models are included in along with their respective train and test errors. The first three models—decision trees, logistic regression, and support vector machines—are less accurate, easier to use, and less complex than the final three models. As we can see, the complexity of the model grows with test accuracy. We can test more advanced models in the future. The ability to pick up new skills and adjust, and we can use a larger variety of datasets to improve forecast precision.

5. References

1. Ha, R. K., Henge, S. K., & Sharma, A. (2020). Optimal machine learning classifiers for prediction of heart disease. International Journal of Control and Automation, 13(1Special Issue), 31-37. J. Y.

2. Sunil S. Khalat, Dr. Yogesh Kumar Sharma ―Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning‖ 2020.

3. Priyan Malarvizhi Kumar and Usha Devi Gandhi, ‖ A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases‖, J. Computers and Electrical Engineering, pp. 222-235, 2018.

4.P.M. Kumar, S. Lokesh, R. Varatharajan, C. Gokul Nath, P. Parthasarathy, ‖ Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier‖, Future Generation Computer Systems, 2018.

5.Mingyu Park and Younghood Song, Jawbone, Jongleur Paek,‖ Design and implementation of a smart chair system for IOT‖, IEEE,2016.

6.Khambete, N. D, and A. Murray, ―National efforts to improve healthcare technology management and medical device safety in India, ‖ Appropriate Healthcare Technologies for Developing Countries, 7th International Conference on, IET, pp. 1–5, 2012.

7.Chipara, Octav, Chen yang Lu, Thomas C. Bailey, and Gruia-Catalin Roman, ―Reliable clinical monitoring using wireless sensor networks: experiences in a step-down hospital unit, ‖ In Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, ACM, pp. 155–168, 2010.

8.Koppar, Anant R, and Venu Sridhar, ―A workflow solution for electronic health records to improve healthcare delivery efficiency in rural India,‖ IneHealth, Telemedicine, and Social Medicine, 2009. eTELEMED'09. International Conference on, pp. 227–232, IEEE, 2009

9. Sharma YK, Khalat Sunil S. Health Care Patient Monitoring using IoT and Machine Learning, in IOSR Journal of Engineering (IOSR JEN) National Conference on ―Recent Innovations in Engineering and Technology‖ MOMENTUM-19.

DOI: 10.55041/IJSREM27940

ISSN: 2582-3930

Impact Factor: 8.176

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT

An Open Access Scholarly Journal || Index in major Databases & Metadata

CERTIFICATE OF PUBLICATION

International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to

Shaikh Ayaan

in recognization to the publication of paper titled

Heart Disease Prediction System using Machine Learning

published in IJSREM Journal on Volume 08 issue on 29 May 2024

www.ijsrem.com

Editor-in-Chief
IJSREM Journal

ijsremjournal@gmail.com

DOI: 10.55041/IJSREM27940

ISSN: 2582-3930

Impact Factor: 8.176

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT

An Open Access Scholarly Journal || Index in major Databases & Metadata

CERTIFICATE OF PUBLICATION

International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to

Khan Moin

in recognization to the publication of paper titled

Heart Disease Prediction System using Machine Learning

published in IJSREM Journal on Volume 08 issue on 29 May 2024

www.ijsrem.com

Editor-in-Chief
IJSREM Journal

ijsremjournal@gmail.com

DOI: 10.55041/IJSREM27940

ISSN: 2582-3930

Impact Factor: 8.176

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT

An Open Access Scholarly Journal || Index in major Databases & Metadata

CERTIFICATE OF PUBLICATION

International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to

Kanade Samruddhi

in recognization to the publication of paper titled

Heart Disease Prediction System using Machine Learning

published in IJSREM Journal on Volume 08 issue on 29 May 2024

Editor-in-Chief
IJSREM Journal

www.ijsrem.com

ijsremjournal@gmail.com

DOI: 10.55041/IJSREM27940

ISSN: 2582-3930

Impact Factor: 8.176

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT

An Open Access Scholarly Journal || Index in major Databases & Metadata

CERTIFICATE OF PUBLICATION

International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to

Prof.K.D.Dere

in recognization to the publication of paper titled

Heart Disease Prediction System using Machine Learning

published in IJSREM Journal on Volume 08 issue on 29 May 2024

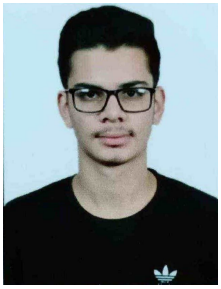
Editor-in-Chief
IJSREM Journal

www.ijsrem.com

ijsremjournal@gmail.com

Chapter 20

INFORMATION OF PROJECT GROUP MEMBERS



1. Name : Shaikh Ayaan Shakil
2. Date of Birth : 04/10/2002
3. Gender : Male
4. Permanent Address : A/P : Junnar Tal-Junnar, Dist-Pune
5. E-Mail : shaikhayaan78660@gmail.com
6. Contact Number : 9325758811

1. Name : Kanade Samruddhi Santosh
2. Date of Birth : 27/01/2003
3. Gender : Male
4. Permanent Address : A/P: Kalamb , Tal-Ambegaon, Dist-Pune
5. E-Mail : samukanade27@gmail.com
6. Contact Number : 9172818082

1. Name : Khan Moin Hayat
2. Date of Birth : 01/01/2002
3. Gender : Male
4. Permanent Address : A/P - Alephata Tal-Junnar, Dist-Pune
5. E-Mail : mkpmoinkhan@gmail.com
6. Contact Number : 7620946947