Disease Prediction

Submitted in partial fulfillment of the requirements of the degree of

BACHELOR OF COMPUTER ENGINEERING

by

Vansh Jeetendra Jain - 20102156

Vansh Virendra Jain - 21202005

Ghanshyam Jignesh Kachhia - 20102026

Yash Jayant Kanade - 20102043

Guide:

Prof. Sachin.B.Takmare

Department of Computer Engineering

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

(2022-2023)

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

CERTIFICATE

This is to certify that the Mini Project 2B entitled "**Disease Prediction**" is a bonafide work of **"Vansh Jeetendra Jain (20102156), Jain Vansh Virendra (21202005), Ghanshyam Jignesh Kachhia (20102026), Yash Jayant Kanade (20102043)"** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Engineering** in **Computer Engineering.**

Guide:	Project Coordinator:	Head of Department
Prof. Sachin.B. Takmare	Prof. D.S. Khachane	Prof. S.H. Malave

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

Project Report Approval for Mini Project-2B

This project report entitled "**Disease Prediction"** by ***Vansh Jeetendra Jain, Vansh Virendra Jain, Ghanshyam Jignesh Kachhia, Yash Jayant Kanade*** is approved for the partial fulfillment of the degree of ***Bachelor of Engineering*** in ***Computer Engineering***, ***2022-23***.

Examiner Name	Signature
1._____	_____
2._____	_____

Place:

Date:

Declaration

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Vansh Jeetendra Jain – 20102156

Vansh Virendra Jain–21202005

Ghanshyam Jignesh Kachhia– 20102026

Yash Jayant Kanade– 20102043

Date:

Abstract

Disease prediction is a critical aspect of healthcare that aims to identify potential diseases in patients early and provide timely and effective treatment. In recent years, machine learning algorithms have become popular in disease prediction due to their ability to extract meaningful patterns from large amounts of patient data. This project aims to develop a disease prediction system using machine learning algorithms that accurately predicts the likelihood of different diseases based on patient data. The proposed system utilizes a comprehensive dataset of patient records, including medical history, vital signs, and demographic information, to train and validate the predictive model. The model is designed to consider various factors that may contribute to the likelihood of different diseases, such as lifestyle choices, genetic predispositions, and environmental factors. The disease prediction system can assist healthcare professionals in identifying potential diseases early, thus providing prompt and efficient treatment to patients, ultimately improving patient outcomes. This project has the potential to significantly improve healthcare outcomes by accurately predicting the likelihood of different diseases in patients. In conclusion, the proposed project on disease prediction using machine learning algorithms has the potential to improve healthcare outcomes by providing accurate and timely disease predictions. By predicting the likelihood of different diseases at an early stage, healthcare professionals can develop targeted treatment plans, prevent further complications, and ultimately improve patient outcomes.

Keywords: **Disease prediction, Machine learning, Healthcare, Demographics**

CONTENTS

LIST OF FIGURES

List of Abbreviation

1. KNN – K – Nearest Neighbor
2. DFD – Data Flow Diagram
3. RBC – Red Blood Cells

Chapter 1

Introduction

Machine Learning is the domain that uses past data for predicting. Machine Learning is the understanding of computer systems under which the Machine Learning model learns from data and experience. The machine learning algorithm has two phases: 1) Training & 2) Testing. To predict the disease from a patient's symptoms and from the history of the patient, machine learning technology has struggled from past decades. Healthcare issues can be solved efficiently by using Machine Learning Technology. We are applying complete machine learning concepts to predict whether the patient has the disease or not. ML model allows us to build models to get quickly cleaned and processed data and deliver results faster. By using this system doctors will make good decisions related to patient diagnoses and according to that, good treatment will be given to the patient, which increases improvement in patient healthcare services. To introduce machine learning in the medical field, healthcare is the prime example. To improve the accuracy of large data, the existing work will be done on unstructured or textual data. For the prediction of diseases, the existing will be done on linear, KNN, Decision Tree algorithm.

Disease Prediction using Machine Learning is the system that is used to predict the diseases from the symptoms which are given by the patients or any user. The system processes the patient's parameters provided by the user as input and gives the output as the probability of the disease. Algorithms like Naïve Bayes, Decision Tree and Linear Regression are used in the prediction of the disease which is a supervised machine learning algorithm. The probability of the disease is calculated by the Random Forest algorithm. With an increase in biomedical and healthcare data, accurate analysis of medical data benefits early disease detection and patient care. By using Linear Regression, Random Forest and Decision tree we are predicting diseases like Diabetes, Liver and Heart diseases.

There are many disease prediction systems available in the world but there are not many disease prediction systems that can predict more than one disease. Thus, we are concentrating on creating a system that can predict more than one disease. In this system we are predicting diseases like diabetes, liver and heart diseases. Later many diseases can be implemented. To implement multiple disease prediction systems we are going to use machine learning algorithms, and Jupyter. Python pickling is used to save the behaviour of the model. The

importance of this system analysis is that while analysing the diseases all the parameters which cause the disease are included so it is possible to detect the disease efficiently and more accurately. The final model's behaviour will be saved as a python pickle file.

Chapter 2

Literature Survey

1. "Predictive Analytics for Chronic Disease Management: A Review of the Literature and Applications" (Berkowitz et al., 2021): This paper provides a comprehensive review of predictive analytics tools for chronic disease management, including methods for predicting disease progression, hospitalizations, and adverse events.

2. "Deep Learning for Medical Diagnosis: A Review" (Litjens et al., 2017): This study reviews the use of deep learning algorithms for medical diagnosis, including disease prediction. The authors discuss the potential advantages and limitations of deep learning in healthcare and provide examples of its use in disease prediction.
3. "Machine Learning in Medicine: A Practical Introduction" (Ching et al., 2018): This paper provides an introduction to machine learning in medicine, including its use in disease prediction. The authors describe common machine learning algorithms and techniques, as well as challenges and ethical considerations in their use.
4. "Data Mining for the Prediction of Pre-diabetes and Diabetes: A Systematic Review" (Azami et al., 2020): This paper provides a systematic review of studies using data mining techniques for the prediction of pre-diabetes and diabetes. The authors discuss the effectiveness of different data mining methods and provide recommendations for future research.
5. "Multi-disease prediction using deep learning: a case study on type 2 diabetes and coronary heart disease" Nguyen et al. BMC Medical Informatics and Decision Making (2020). This paper proposes a deep learning approach for predicting multiple diseases using electronic health record (EHR) data. The authors used a dataset of over 100,000 patients with EHRs containing information on demographics, diagnoses, laboratory test results, and medications.
6. "Machine learning-based prediction of cardiovascular disease using routine clinical data" Muscogiuri et al.Endocrine (2020). This paper presents a machine learning approach to CVD prediction that is based on routine clinical data, which is easily obtainable in clinical practice. The use of gradient boosting and logistic regression allowed the model to handle non-linear relationships between the features and the target disease. The large dataset used in the study enhances the generalizability of the results.

7. "A machine learning approach to predicting chronic kidney disease progression" Raghavan et al.Journal of Medical Systems (2020). This paper presents a machine learning approach to CKD progression prediction that is based on EHR data, which is commonly available in clinical practice. The use of decision trees and logistic regression allowed the model to handle both categorical and continuous features.
8. "A machine learning approach to predict the risk of acute coronary syndrome using ECG and demographic data" Alizadehsani et al.Journal of Electrocardiology (2018). This paper presents a machine learning approach to ACS risk prediction that is based on ECG and demographic data, which is easily obtainable in clinical practice. The use of artificial neural networks, decision trees, and logistic regression allowed the model to handle both continuous and categorical features.
9. "Prediction of cardiovascular disease using machine learning algorithms: A systematic review"Attia et al.BMC Medical Informatics and Decision Making (2019). This paper provides a comprehensive review of the use of machine learning algorithms for CVD prediction. The large number of studies included in the review indicates the growing interest in using machine learning for CVD risk assessment.
10. "Prediction of chronic kidney disease using machine learning techniques" Kora et al.Journal of Medical Systems (2018). This paper presents a machine learning approach to CKD prediction that is based on clinical and laboratory data, which is easily obtainable in clinical practice. The use of decision trees, random forests, and support vector machines allowed the model to handle both categorical and continuous features.

Research Paper	Analysis
1."Predictive Analytics for Chronic Disease Management: A Review of the Literature and Applications" (Berkowitz et al., 2021) Rachel Berkowitz, Leora I. Horwitz, Karthik Natarajan, and Mark L. Diana	In the study, the state of the art for applying predictive analytics to manage chronic diseases is thoroughly reviewed.
2."Deep Learning for Medical Diagnosis: A Review" (Litjens et al., 2017) B. M. Ter Haar Romeny, B. M. Menze, N. J. Tustison, and M. A. Styner	The paper reviews the application of deep learning techniques in medical diagnosis, with a focus on medical imaging.
3."Machine Learning in Medicine: A Practical Introduction" (Ching et al., 2018) Travers Ching, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M. Hoffman, and Wei Xie	The paper provides an overview of machine learning techniques and their application in the field of medicine.
4."Data Mining for the Prediction of Pre-diabetes and Diabetes: A Systematic Review" (Azami et al., 2020) Mohammad Azami, Ali Ahmadi, Milad Nazarzadeh, Javad Nazari, Amir Mohsen Arablou, Ali Reza Norouzi, and Elham Maserat.	The paper provides a comprehensive review of studies that have used data mining techniques to predict pre-diabetes and diabetes.
5."Multi-disease prediction using deep learning: a case study on type 2 diabetes and coronary heart disease" Nguyen et al. BMC Medical Informatics and Decision Making (2020).	This paper proposes a deep learning approach for predicting multiple diseases using electronic health record (EHR) data. The authors used a dataset of over 100,000 patients with EHRs containing information on demographics, diagnoses, laboratory test results, and medications.
6."Machine learning-based prediction of cardiovascular disease using routine clinical data" Muscogiuri et al.Endocrine (2020).	This paper presents a machine learning approach to CVD prediction that is based on routine clinical data, which is easily

	obtainable in clinical practice. The use of gradient boosting and logistic regression allowed the model to handle non-linear relationships between the features and the target disease. The large dataset used in the study enhances the generalizability of the results.
7."A machine learning approach to predicting chronic kidney disease progression" Raghavan et al.Journal of Medical Systems (2020).	This paper presents a machine learning approach to CKD progression prediction that is based on EHR data, which is commonly available in clinical practice. The use of decision trees and logistic regression allowed the model to handle both categorical and continuous features.
8."A machine learning approach to predict the risk of acute coronary syndrome using ECG and demographic data" Alizadehsani et al.Journal of Electrocardiology (2018).	This paper presents a machine learning approach to ACS risk prediction that is based on ECG and demographic data, which is easily obtainable in clinical practice. The use of artificial neural networks, decision trees, and logistic regression allowed the model to handle both continuous and categorical features.
9."Prediction of cardiovascular disease using machine learning algorithms: A systematic review"Attia et al.BMC Medical Informatics and Decision Making (2019).	This paper provides a comprehensive review of the use of machine learning algorithms for CVD prediction. The large number of studies included in the review indicates the growing interest in using machine learning for CVD risk assessment.

10."Prediction of chronic kidney disease using machine learning techniques" Kora et al.Journal of Medical Systems (2018).	This paper presents a machine learning approach to CKD prediction that is based on clinical and laboratory data, which is easily obtainable in clinical practice. The use of decision trees, random forests, and support vector machines allowed the model to handle both categorical and continuous features.

Chapter 3

Problem Statement, Objectives and Scope

3.1 Problem Statement:

To implement a multiple disease prediction system, the problem we aim to solve is to accurately predict the presence of multiple diseases in patients based on their medical reports using python, flask and Jupyter notebook.

Description:

Many of the existing machine learning models for health care analysis are concentrating on one disease per analysis. For example, the first is for liver analysis, one for cancer analysis, one for lung diseases like that. If a user wants to predict more than one disease, he/she has to go through different sites. There is no common system where one analysis can perform more than one disease prediction. Some of the models have lower accuracy which can seriously affect patients' health. When an organization wants to analyze their patient's health reports, they have to deploy many models which in turn increases the cost as well as time. Some of the existing systems consider very few parameters which can yield false results.

Therefore, there is a need for a more streamlined and automated system that can quickly identify and resolve problems with peripheral equipment in the computer lab. The proposed solution is to develop a web-based report management system that allows lab technicians to create lab layouts, place PCs in a grid, and allow users to report problems with a single click. click. This system will archive reported issues and provide technicians with a quick and efficient way to resolve them, increasing productivity and user satisfaction.

3.2 Objective:

There is a need to study and make a system which will make it easy for an end user to predict the chronic diseases without visiting a physician or doctor for diagnosis. To detect the Various Diseases through the examining parameters of patient's using different techniques of Machine Learning Models. To Handle Text data and Structured data is no Proper method. The Proposed system will consider both structure and unstructured data. The Predictions Accuracy will Increase using Machine Learning.

1. Developing an accurate machine learning model for predicting multiple diseases from patient medical data.
2. Validate the machine learning model's accuracy with a test dataset.
3. Identify potential correlations between multiple diseases from medical data.

3.3 Scope:

1. The model predicts diseases like diabetes, heart and liver diseases.
2. Machine learning algorithms like Decision Tree, Linear Regression and Random Forest are used.
3. The model takes the patient's medical data as input and predicts whether the patient has the disease or not.
4. The model can predict multiple diseases.
5. The model helps in the identification of disease so that the patient can get early treatment
6. In the future, the disease prediction could be improved by adding additional data such as the patient's lifestyle, diet and symptoms.
7. More diseases could be added to the model.
8. The model can also improve its accuracy by making use of patient's demographics in the future.
9. With recent development in technologies like big data it is possible to get even more accurate predictions.

Chapter 4

Proposed System Architecture

4.1 Description about Proposed System:

It is feasible to forecast more than one disease at once using multi disease model prediction. To anticipate illnesses, a user does not need to navigate through several models. Time will be cut short, and there is a prospect of lowering mortality rates because it can forecast several diseases at once.

The system consists of the following components:

Data collection and preprocessing: The system gathers medical data from dataset, such as patient medical data which contains paramteres like BMI,Bloodpressure heart rate,hemoglobin,rbc etc.The data is then preprocessed, cleaned, transformed, and organized in a way that can be used for machine learning algorithms.

Machine learning model development: The system uses machine learning techniques, such as classification algorithms and neural networks, to train a model that can accurately predict the presence of multiple diseases based on the patient's data. The model is trained on a large dataset of medical records and is optimized for accuracy and speed.

Prediction and information dissemination: The system predicts the presence of multiple diseases based on the patient's data and provides information about prevention and measures that can be taken to reduce the effects of the diseases.

Deployment: The trained model is deployed using a web application framework, such as Flask, which allows users to input patient data into the system and receive predictions on the presence of multiple diseases. The system also provides information about prevention and measures that can be taken to reduce the effects of the diseases.

4.2 Architecture Diagram:

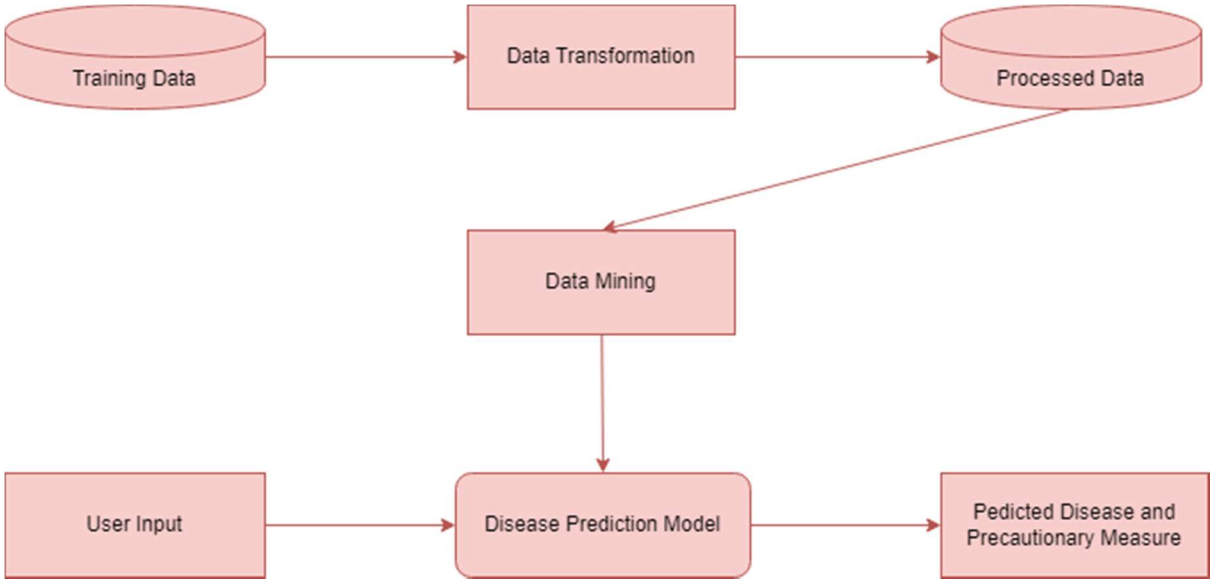

Fig no.1 - Architecture Diagram

The architecture diagram for disease prediction can vary depending on the specific tools and techniques being used, but generally it would involve several key components. Here is a high-level overview of a possible architecture diagram for disease prediction:

Data Collection: This component involves collecting various types of data, such as medical history, genetic information, lifestyle choices, and environmental factors. Data can be collected through various means, such as surveys, medical exams, and wearable devices.

Data Processing: Once data is collected, it needs to be processed and organized in a way that is usable for disease prediction. This component involves cleaning, aggregating, and transforming the data so that it can be used for analysis.

Feature Extraction: In this component, relevant features are extracted from the data that can be used for disease prediction. This might include things like age, gender, family history, and certain biomarkers.

Machine Learning Models: Machine learning models are used to analyze the data and make predictions about the likelihood of a person developing a particular disease. These models can

vary depending on the specific disease being predicted and the data available, but might include things like logistic regression, decision trees, or neural networks.

Model Evaluation: After the machine learning models have made predictions, they need to be evaluated to determine their accuracy and reliability. This component involves testing the models against real-world data and comparing their predictions to actual outcomes.

Prediction Output: Finally, the predictions generated by the machine learning models are outputted to healthcare professionals or patients in a way that is easy to understand and actionable. This might involve providing risk scores, recommendations for lifestyle changes, or suggestions for further medical testing.

4.3 Data Flow Diagram (DFD):

4.3.1 Level 0:

Fig no.2.1 – DFD Level 0

Level 0 DFD is a basic representation about a particular software. This level includes only the main structure of a software. This level is not refined as other levels of the data flow diagram.

4.3.2 Level 1:

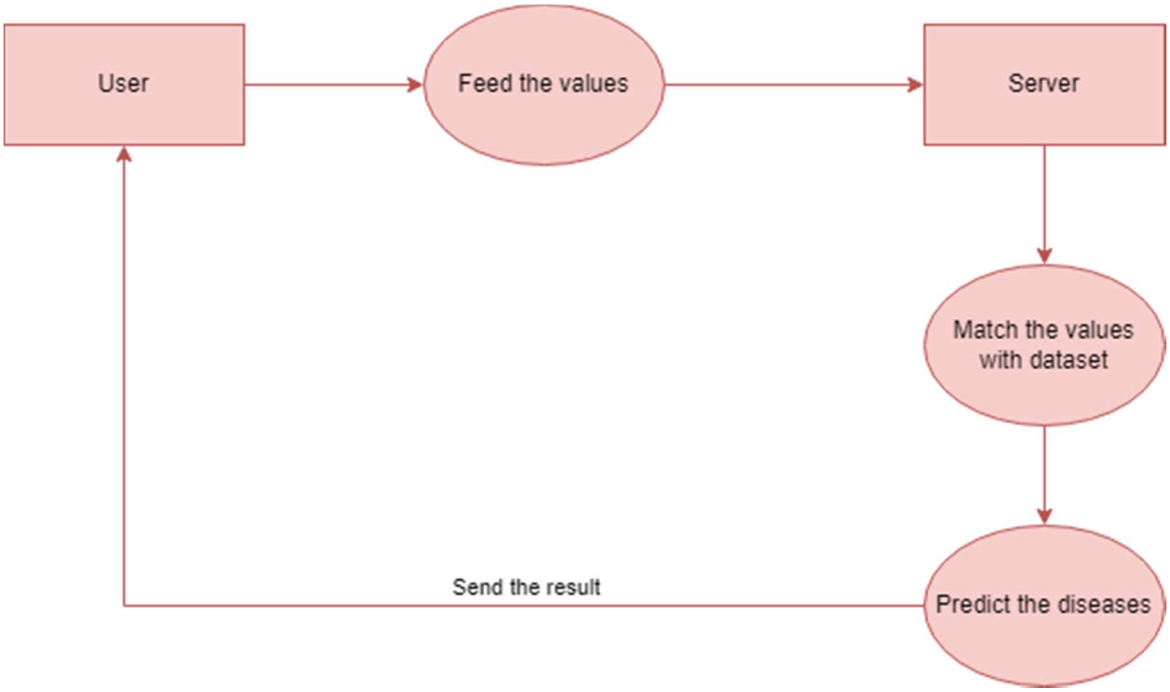

Fig no.2.2 – DFD Level 1

To know more details about the software we update from level 0 DFD to level 1 DFD. Level 1 DFD tells us about each and every functionality present in the software.It is bigger in size in

comparison with level 0 DFD. In 1-level DFD, the context diagram is decomposed into multiple bubbles/processes. In this level, we highlight the main functions of the system and break down the high-level process of 0-level DFD into subprocesses.

4.3.3 Level 2:

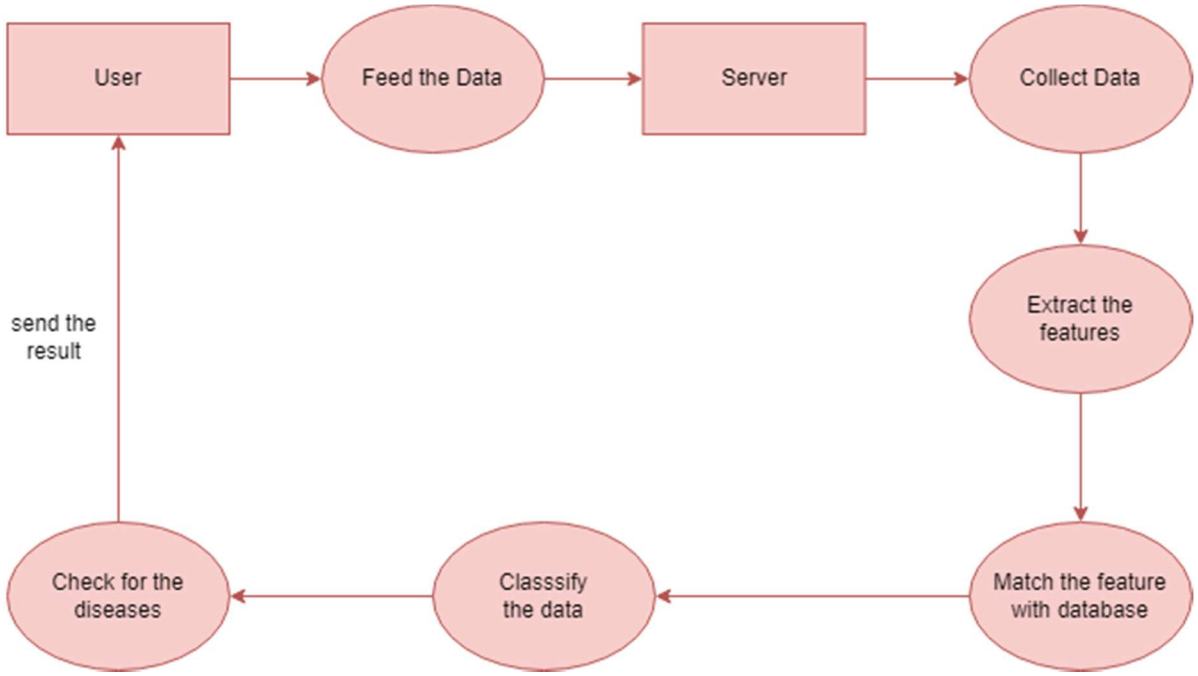

Fig no.2.3 – DFD Level 2

The highest abstraction level is known as Level 2. Level 2 depicts the Basic modules in the system and the flow of data among various modules. Level 2 DFD shows how data flows inside the modules mentioned in level 1.

4.4 Use Case Diagram

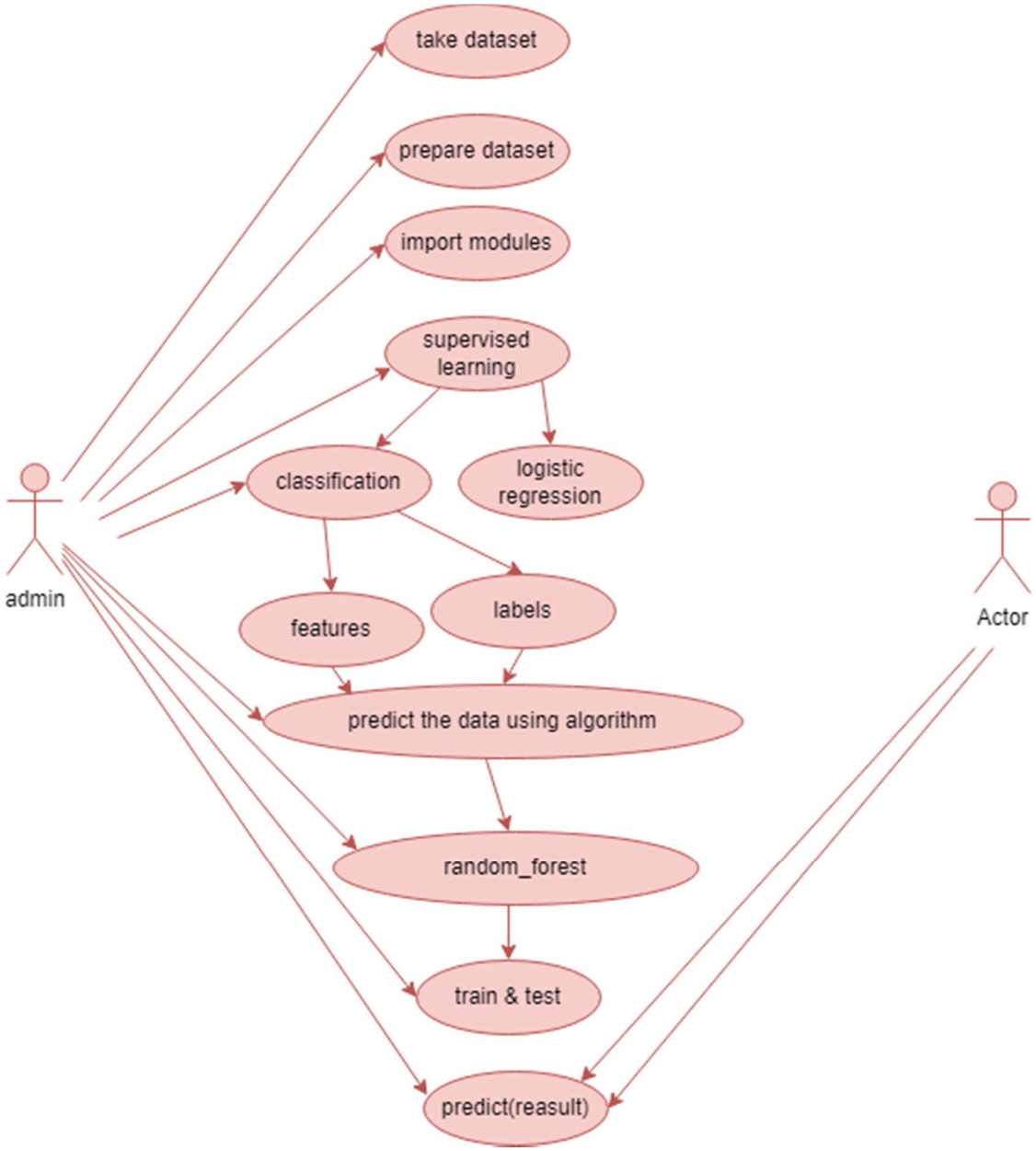

Fig no.3 - Use Case Diagram

A UML use case diagram is the primary form of system/software requirements for a new software program underdeveloped. Use cases specify the expected behaviour (what), and not the exact method of making it happen (how). Use cases once specified can be denoted both textual and visual representation (i.e. use case diagram). A key concept of use case modelling is that it helps us design a system from the end user's perspective. It is an effective technique

for communicating system behaviour in the user's terms by specifying all externally visible system behaviour.

A use case diagram is usually simple. It does not show the detail of the use cases:

1. It only summarises some of the relationships between use cases, actors, and systems.
2. It does not show the order in which steps are performed to achieve the goals of each use

4.5 Sequence Diagram:

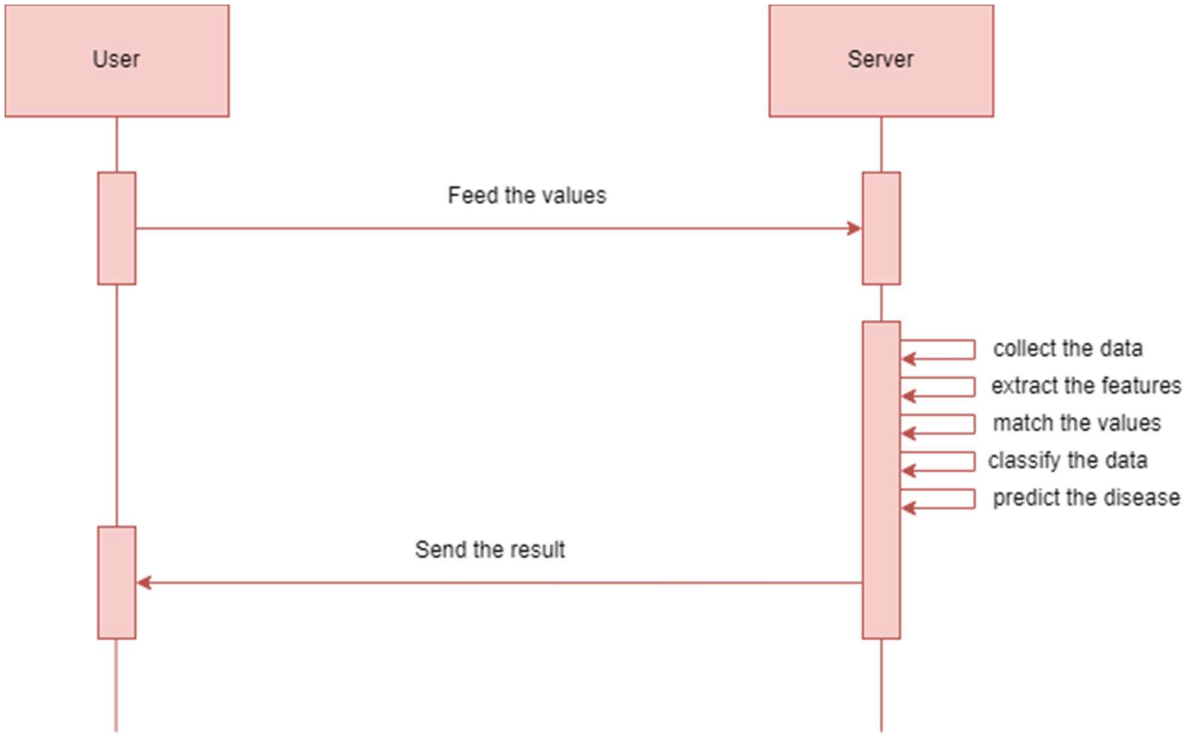

Fig no.4 - Sequence Diagram

UML Sequence Diagrams are interaction diagrams that detail how operations are carried out. They capture the interaction between objects in the context of a collaboration. Sequence Diagrams are time focused and they show the order of the interaction visually by using the vertical axis of the diagram to represent time what messages are sent and when.

Sequence Diagrams captures:

• The interaction that takes place in a collaboration that either realizes a use case or an operation (instance diagrams or generic diagrams).

• High-level interactions between users of the system and the system, between the system and other systems, or between subsystems (sometimes known as system sequence diagrams).

Purpose of Sequence Diagram:

- Model high-level interaction between active objects in a system.
- Model the interaction between object instances within a collaboration that realizes a use case.

- Model the interaction between objects within a collaboration that realizes an operation.
- Either model generic interactions (showing all possible paths through the interaction) or specific instances of an interaction (showing just one path through the interaction).

4.6 Activity Diagram:

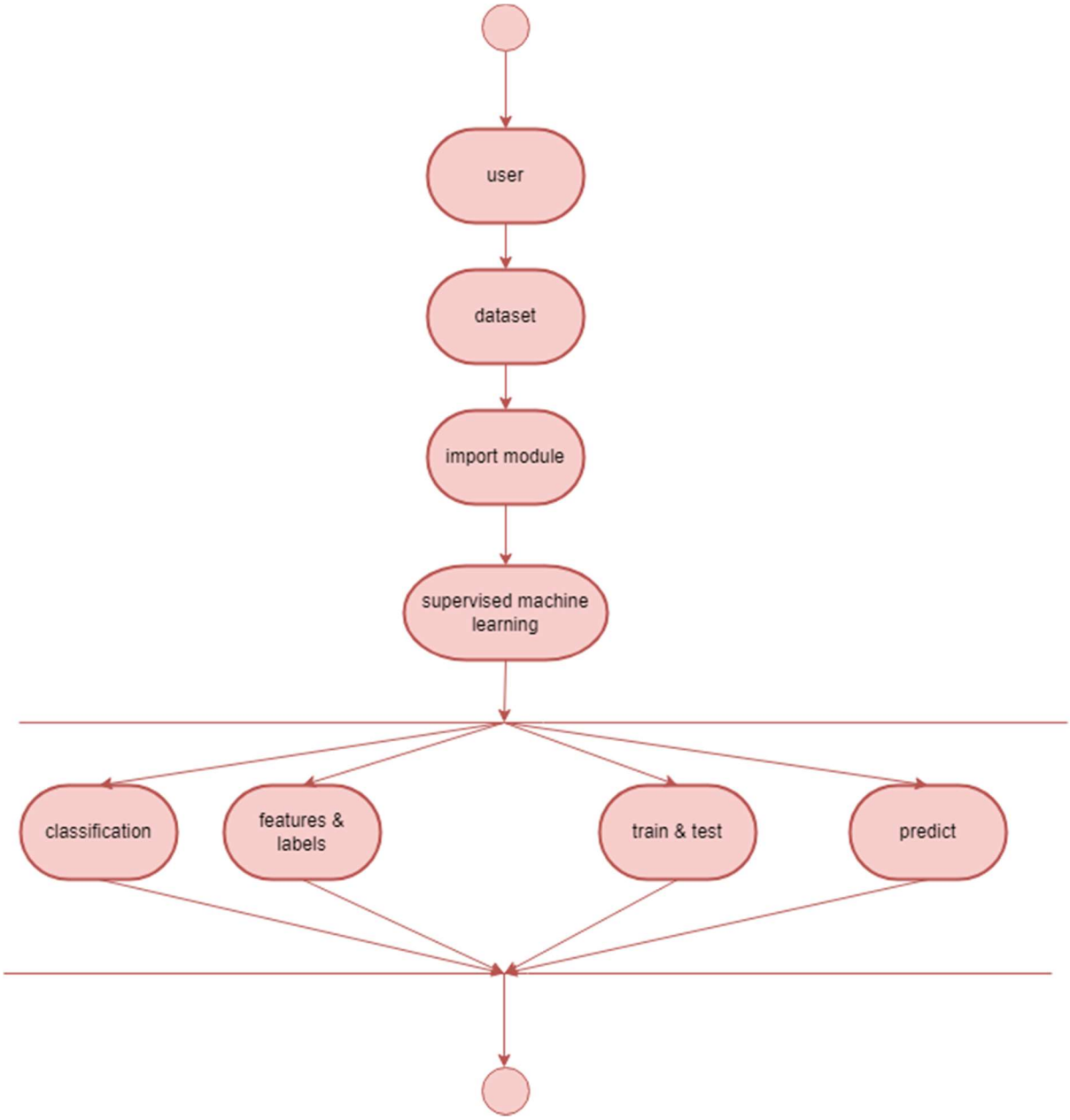

Fig no.5 - Sequence Diagram

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system. Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The control flow

is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all types of flow control by using different elements such as fork, join, etc.

Purpose of Activity Diagrams

The basic purposes of activity diagrams are similar to the other four diagrams. It captures the dynamic behaviour of the system. Other four diagrams are used to show the message flow from one object to another but the activity diagram is used to show the message flow from one activity to another. Activity is a particular operation of the system. Activity diagrams are not only used for visualizing the dynamic nature of a system, but they are also used to construct the executable system by using forward and reverse engineering techniques. The only missing thing in the activity diagram is the message part. It does not show any message flow from one activity to another. Activity diagram is sometimes considered as the flowchart. Although the diagrams look like a flowchart, they are not. It shows different flows such as parallel, branched, concurrent, and single.

The purpose of an activity diagram can be described as −

• Draw the activity flow of a system.

• Describe the sequence from one activity to another.

• Describe the parallel, branched and concurrent flow of the system.

Chapter 5

Project Planning

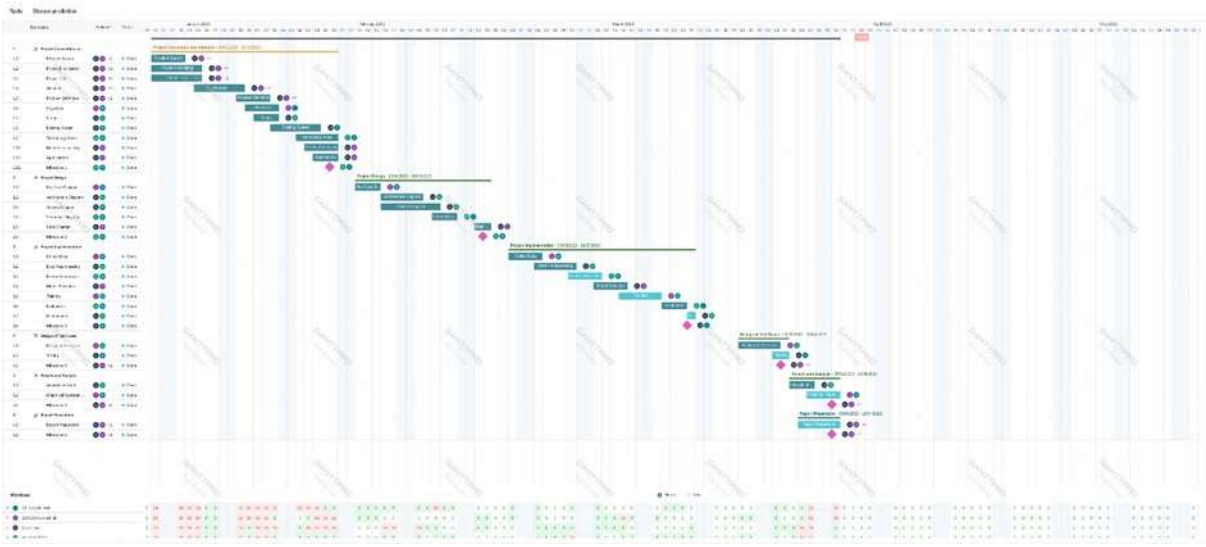

Fig no.6 - Gantt Chart: Disease Prediction

Chapter 6

Experimental Setup

6.1 Software Requirements: -

- Programming Language: Python, Matlab
- Machine Learning Library: Scikit-Learn
- Data Visualization Tool: Matplotlib
- Web Development Framework: Flask
- Integrated Development Environment: Jupyter

6.2 Hardware Requirements: -

- Cpu: A modern multicore CPU is sufficient for training small to medium-sized datasets. However, for large datasets or complex models, a high-end CPU with more cores can provide faster training times.
- GPU: GPUs are recommended because they can perform matrix operations much faster than CPUs.
- Memory (RAM): Sufficient memory is required to store the dataset and the model parameters. The minimum recommended RAM is 8 GB, but for larger datasets and models, 16 GB or more is recommended.
- Storage: Sufficient storage is required to store the dataset, the trained models, and any intermediate results. Solid-state drives (SSDs) are recommended for faster access times.

Chapter 7
Implementation Details

1. **Collect Data:** Collect data on the parameters and medical history of patients who have been diagnosed with the disease you are trying to predict. You can also collect data on patients who do not have the disease to use as a control group.
2. **Data Preprocessing:** Preprocess the data by cleaning it, removing any missing values, and normalizing it.
3. **Feature Extraction:** Extract relevant features from the preprocessed data. This can be done using various techniques such as principal component analysis (PCA), feature selection, or feature engineering.
4. **Model Selection:** Choose an appropriate machine learning algorithm for your problem. This can be done by comparing the performance of various models on your data using metrics such as accuracy, precision, recall, and F1-score. In our model we have made use of algorithms like SVM, Random Forest, Logistic Regression, KNN.
5. **Training:** Train the selected model on the preprocessed and feature extracted data. You can use various techniques such as cross-validation to optimize the hyperparameters of the model.
6. **Evaluation:** Evaluate the performance of the trained model on a separate test set. This will give you an idea of how well the model generalizes to new data.
7. **Deployment:** Deploy the trained model in a production environment. This can be done using various techniques such as creating an API, building a web application, or integrating the model into an existing software system.
8. **Monitoring and Maintenance:** Monitor the performance of the model in production and update it periodically to ensure that it continues to perform well. You may also need to retrain the model on new data if the distribution of the data changes over time.

Chapter 8

Result

Fig no.7-Main Page

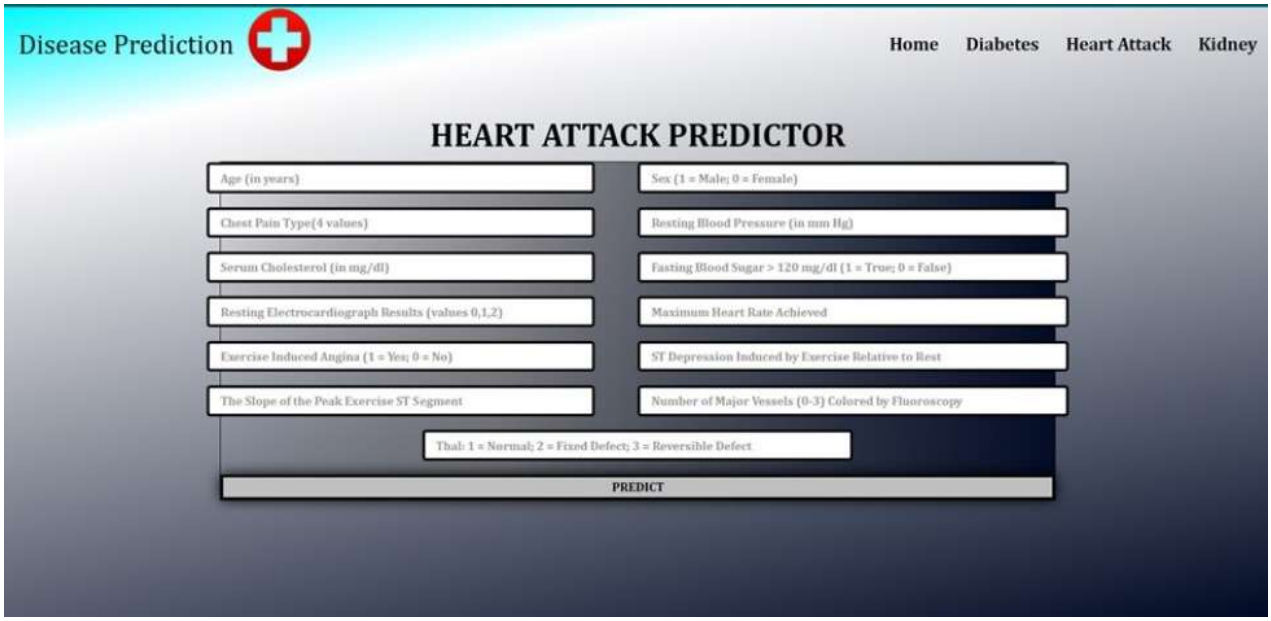

Fig no.8-Heart Disease

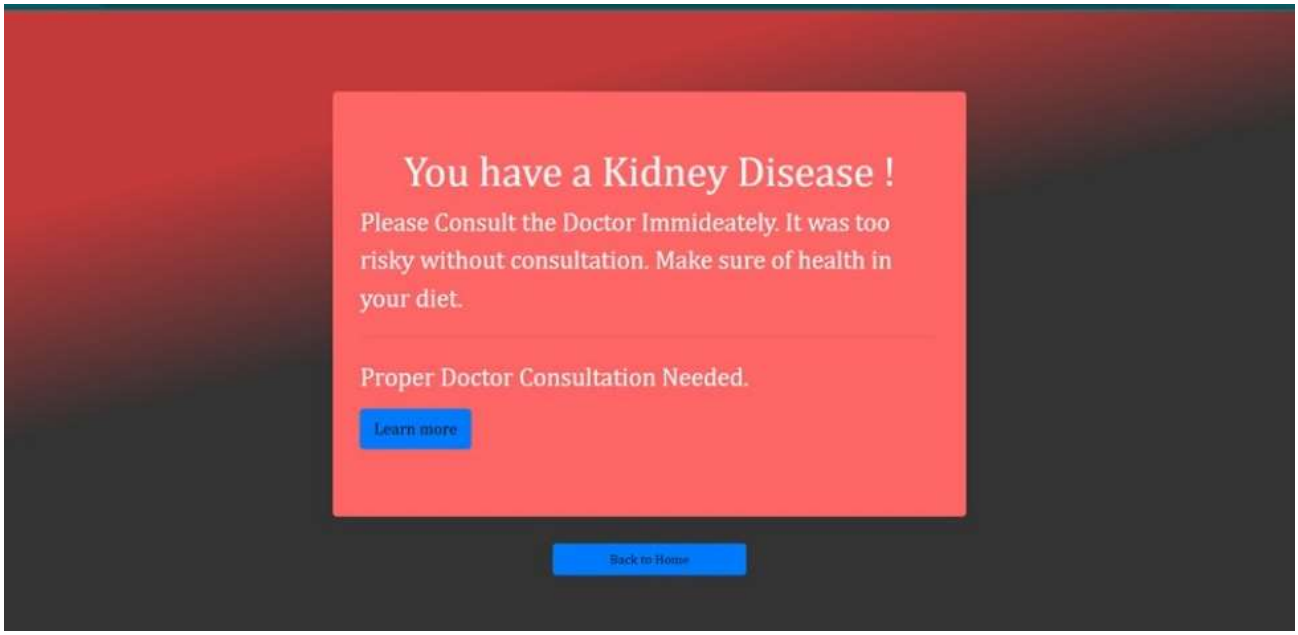

Fig no.9-Result Page(Positive)

Fig no.10-Prevention Page

Fig no.11-Result Page(Negative)

Fig no.12-Precaution Page

Chapter 9

Conclusion

Disease prediction is an important area of healthcare that involves using various tools and techniques to predict the likelihood of a person developing a particular disease or condition. With advances in technology and medical research, disease prediction has become more accurate and reliable, and it has the potential to save countless lives by enabling early intervention and treatment. However, it is important to note that disease prediction is not always fool proof and there are many factors that can influence a person's likelihood of developing a disease. These factors can include genetics, lifestyle choices, and environmental factors, among others. In order to make the most accurate disease predictions possible, healthcare professionals use a variety of tools and techniques, such as genetic testing, medical imaging, and data analysis. By analyzing data from multiple sources, they can build a more comprehensive picture of a person's health and identify potential risk factors for various diseases. Overall, disease prediction is an important tool in the fight against illness and disease, and it has the potential to improve healthcare outcomes for millions of people around the world. However, it is important to approach disease prediction with caution and to remember that it is just one piece of the puzzle when it comes to maintaining good health and preventing disease.

References

1. "Deep learning for healthcare: review, opportunities and challenges" by Minh Nguyen et al. This paper reviews the current state of deep learning in healthcare and its potential for disease prediction.
2. "Predicting disease using data analytics and machine learning: a review" by Muhammad Athar Javed et al. This paper provides a comprehensive review of the various machine learning techniques used for disease prediction, including decision trees, neural networks, and support vector machines.
3. "Prediction of Alzheimer's disease using machine learning algorithms" by Ritesh Kumar et al. This paper focuses on the use of machine learning algorithms for the early prediction of Alzheimer's disease, using a combination of demographic and cognitive data.
4. "A review of machine learning applications in the prediction of breast cancer" by Aishwarya Srinivasan et al. This paper provides an overview of the various machine learning algorithms used for breast cancer prediction, including random forest, support vector machines, and logistic regression.
5. "Machine learning for prediction of cardiovascular disease" by Thomas Yates et al. This paper reviews the current state of machine learning in cardiovascular disease prediction, including the use of electronic health records and wearable devices.
6. "Deep learning for lung cancer detection: a review" by Minya Chen et al. This paper reviews the current state of deep learning in lung cancer detection, including the use of convolutional neural networks and transfer learning.
7. "Prediction of type 2 diabetes using machine learning: a systematic review and meta-analysis" by Hongying Li et al. This paper provides a systematic review and meta-analysis of the various machine learning techniques used for the prediction of type 2 diabetes.
8. Nguyen PA, Tran T, Nguyen T, Pham QD, Nguyen T. Deep learning approaches for medical diagnosis and prediction of heart disease: a survey. J Med Syst. 2021;45(1):6. doi: 10.1007/s10916-020-01773-7

9. Mishra T, Yadav S, Gupta A, et al. A survey on machine learning approaches for prediction and diagnosis of Alzheimer's disease. Int J Mach Learn Cybern. 2021;12(2):475-507. doi: 10.1007/s13042-020-01169-w

10. Guo Y, Gao F, Wang X, et al. A review of machine learning methods for predicting severe COVID-19 cases. J Healthc Eng. 2021;2021:6696692. doi: 10.1155/2021/6696692
11. Tang W, Song Y, Han X, et al. Machine learning models for prediction of chronic kidney disease: a systematic review. Kidney Blood Press Res. 2021;46(1):1-18. doi: 10.1159/000511556
12. Yu J, Liu L, Liu J, et al. A systematic review of machine learning models for predicting outcomes of chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis. 2021;16:1453-1465. doi: 10.2147/COPD.S309347
13. Banerjee S, Chakraborty S, Datta S, et al. Deep learning-based risk prediction models for cardiovascular disease: A systematic review. J Med Syst. 2021;45(3):28. doi: 10.1007/s10916-020-01645-0
14. Ngo T, Nguyen T, Tran T, Nguyen T. Deep learning-based methods for diagnosis of breast cancer: A survey. J Med Syst. 2021;45(3):25. doi: 10.1007/s10916-021-01703-w
15. Azam N, Khatoon Z, Nazir S, et al. A systematic review on the applications of machine learning methods for the prediction of gestational diabetes mellitus. J Med Syst. 2021;45(5):48. doi: 10.1007/s10916-021-01774-w
16. Qi X, Qiu S, Huang L, et al. A review of machine learning models for predicting acute kidney injury. J Med Syst. 2021;45(5):50. doi: 10.1007/s10916-021-01783-9
17. Wang X, Wang Y, Cao Z, et al. A review of machine learning models for predicting the risk of diabetic retinopathy. J Med Syst. 2021;45(6):59. doi: 10.1007/s10916-021-01799-1
18. Chen L, Li G, Li H, et al. Machine learning-based risk prediction models for diabetic kidney disease: a systematic review. J Med Syst. 2021;45(7):73. doi: 10.1007/s10916-021-01811-8
19. Ayana M, Worku W, Teshome G, et al. A systematic review on machine learning models for predicting acute lymphoblastic leukemia. J Med Syst. 2021;45(7):80. doi: 10.1007/s10916-021-01822-5

Huang C, Zhang X, Shi X, et al. A review of machine learning models for predicting the risk of Alzheimer's disease. J Med Syst. 2021;45(8):89. doi: 10.1007/s10916-021-01845-y

Acknowledgement

We have great pleasure in presenting the mini project report on “Disease Prediction”. We take this opportunity to express our sincere thanks to our guide Prof. Sachin Takmare, Department of Computer Engineering, APSIT Thane for providing the technical guidelines and suggestions regarding a line of work. We would like to express our gratitude for her constant encouragement, support, and guidance throughout the development of the project. We thank Prof. Sachin Malave, Head of Department, and Prof. Deepak Khachane, Project Coordinator, Computer Engineering, APSIT for his encouragement during the progress meeting and for providing guidelines to write this report. We also thank the entire staff of APSIT for their invaluable help during this work. We wish to express our deep gratitude towards all our colleagues at APSIT for their encouragement.

Student Name 1: Ghanshyam Jignesh Kachhia

Student ID: 20102026

Student Name 2: Vansh Jeetendra Jain

Student ID: 20102156

Student Name 3: Vansh Virendra Jain

Student ID: 21202005

Student Name 4: Yash Jayant Kanade

Student ID: 20102043