HEART DISEASE PREDICTION USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

JOTHIKA MANGAI B [927621BAD018]

LIBERNA ASUWATHA A [927621BAD027]

SUPRIYA G [927621BAD055]

In partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

M.KUMARASAMY COLLEGE OF ENGINEERING, KARUR

ANNA UNIVERSITY: CHENNAI 600 025

NOV 2023

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report "**HEART DISEASE PREDICTION USING MACHINE LEARNING**" is the bonafide work of "JOTHIKAMANGAI B [927621BAD018], LIBERNA ASUWATHA A [927621BAD027], SUPRIYA G [927621BAD055]" who carried out the project work under my supervision.

SIGNATURE

Dr.R.Raja Guru,M.Tech.,Ph.D,

ASSOCIATE PROFESSOR AND HEAD

Department of Artificial Intelligence,

M.Kumarasamy College of Engineering,

Thalavapalayam,

Karur-639113

SIGNATURE

Dr.R.Raja Guru,M.Tech.,Ph.D,

ASSOCIATE PROFESSOR AND HEAD

Department of Artificial Intelligence,

M.Kumarasamy College of Engineering,

Thalavapalayam,

Karur-639113

TABLE OF CONTENTS

ABSTRACT

The diagnosis and prognosis of cardiovascular disease are crucial medical tasks to ensure correct classification, which helps cardiologists provide proper treatment to the patient. Machine learning applications in the medical niche have increased as they can recognize patterns from data. Using machine learning to classify cardiovascular disease occurrence can help diagnosticians reduce misdiagnosis. This research develops a model that can correctly predict cardiovascular diseases to reduce the fatality caused by cardiovascular diseases. This paper proposes a method of k-modes clustering with Huang starting that can improve classification accuracy. Models such as random forest (RF), decision tree classifier (DT), multilayer perceptron (MP), and XGBoost (XGB) are used. GridSearchCV was used to hypertune the parameters of the applied model to optimize the result. The proposed model is applied to a real-world dataset of 70,000 instances from Kaggle. Models were trained on data that were split in 80:20 and achieved accuracy as follows: decision tree: 86.37% (with cross-validation) and 86.53% (without cross-validation), XGBoost: 86.87% (with cross-validation) and 87.02% (without cross-validation), random forest: 87.05% (with cross-validation) and 86.92% (without cross-validation), multilayer perceptron: 87.28% (with cross-validation) and 86.94% (without cross-validation). The proposed models have AUC (area under the curve) values: decision tree: 0.94, XGBoost: 0.95, random forest: 0.95, multilayer perceptron: 0.95. The conclusion drawn from this underlying research is that multilayer perceptron with cross-validation has outperformed all other algorithms in terms of accuracy. It achieved the highest accuracy of 87.28%.

LIST OF FIGURES

LIST OF TABLES

ACRONYMS/LIST OF ABBREVIATIONS

Acronym	Abbreviations
BMI	Body Mass Index
MDVI	Multidimensional Voice Program
DFA	Direct Fluorescent Antibody Test

CHAPTER-1

INTRODUCTION

Heart failure is a complex and potentially life-threatening condition that significantly burdens healthcare systems worldwide

It is a pathophysiologic condition in which the heart's inability to pump blood at a rate sufficient to meet the needs of the body's metabolizing tissues results from faulty cardiac function [1,2].

It includes a number of heart-related illnesses, such as coronary artery disease, heart attacks, heart failure, arrhythmias, and several other cardiovascular ailments.

Heart disease is a leading cause of death globally [3], accounting for many premature deaths and posing a significant burden on healthcare systems.

Heart disease is a common and significant health issue in many parts of the world [4].

The American Heart Association says that heart failure is projected to increase dramatically [5].

Accurate prediction of heart failure can play a vital role in early detection and prevention of adverse outcomes, ultimately leading to improved patient outcomes and reduced healthcare costs.

Timely and accurate detection of heart failure is crucial for effective management and treatment [6].

Detecting heart failure early allows for prompt intervention and the implementation of appropriate medical strategies, which can help slow the progression of the disease, alleviate symptoms, and improve the patient's quality of life.

Early detection can also reduce the risk of complications and hospitalizations associated with advanced stages of heart failure. From 1989 [7] to now, there have been many approaches to finding the best methods for cardiac failure prediction.

In 2017, Simge et al. [8] used Matlab and WEKA to find the best way to detect heart failure disease and obtained a good accuracy of 67.7% for the ensemble subspace discriminant algorithm and the decision tree algorithm.

Then, in 2018, Ali et al. [9] utilized the Cleveland dataset [10] for their studies and obtained 84% accuracy for the Naive Bayes algorithm. Further, in 2019, Saba et al. [11] performed prediction for heart diseases and obtained 84.85% accuracy for the logistic regression (SVM) technique.

However, most of them used the same dataset from the UCI repository [12], which contains 300 records.

This is a rather limited amount of data for machine learning training

1.1. BACKGROUND:

In the field of healthcare, there is a growing demand for more precise and efficient disease prediction methods. Traditional approaches often rely on manual analysis andcan be limited by human biases and resource constraints. Machine learning (ML), powered by vast datasets and advanced algorithms, has emerged as a promising solution to address these challenges.

Early Disease Detection:

One primary objective of using ML for heart disease prediction is the early detection of various medical conditions. ML models can analyze a diverse range of data sources, such as medical records, genetic information, and imaging scans, to identify subtle patterns and risk factors associated with diseases. Early detection enables timely interventions and treatment, significantly improving patient outcomes.

Personalized Medicine:

Another crucial goal is to personalize medical care based on individual patient profiles. ML models can tailor treatment plans to suit the specific needs and characteristics of each patient, optimizing therapeutic strategies and minimizingpotential side effects.

Resource Optimization:

ML-based disease prediction models allow healthcare providers to allocate resources more efficiently. They can identify high-risk individuals, schedule preventive screenings, and allocate resources where they are most needed, reducing unnecessary tests and treatments while optimizing hospital workflows.

Cost Savings:

The accurate prediction of diseases can result in substantial cost savings. Preventing diseases or diagnosing them at an early stage can alleviate the financial burden on both patients and healthcare systems by reducing the need for expensive, late-stage treatments.

Improved Patient Care:

The application of ML in disease prediction enhances the overall quality of patient care. Healthcare providers can make informed decisions based on data-driven insights,leading to better patient experiences and outcomes.

Research Opportunities:

The use of ML generates extensive medical data, which can serve as a valuable resource for medical research. Researchers can gain new insights into disease mechanisms, potential treatments, and epidemiological trends, further advancing the field.

Challenges and Ethical Considerations:

While the potential benefits are substantial, the deployment of ML in disease prediction also presents challenges. Ethical concerns related to data privacy, bias, and model interpretability must be carefully addressed to ensure responsible and equitableuse of these technologies

1.2 PROBLEM STATEMENT:

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive. The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease. Decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs. According to (Wurz & Takala, 2006) the opportunities to improve care and reduce costs concurrently could apply to as much as 30% of overall healthcare spending. The successful application of data mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors. Among these sectors just discovering is healthcare. The healthcare environment is still „information rich" but „knowledge poor". There is a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in the data for African genres.

1.3 OBJECTIVES:

The main objective of this research is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set Heart disease prediction system aims to exploit data mining techniques on medical data set to assist in the prediction of the heart diseases.

Provides new approach to concealed patterns in the data. Helps avoid human biasness. To implement Naïve Bayes Classifier that classifies the disease as per the input of the user. Reduce the cost of medical tests.

In addition to improving patient care, the application of machine learning in disease prediction also plays a crucial role in optimizing resource allocation within the healthcare system. By identifying high-risk individuals and allocating resources accordingly, healthcare facilities can operate more efficiently, ensuring that those who need care the most receive it promptly. This approach not only enhances the overall healthcare experience but also contributes to cost reduction, making healthcare more accessible and affordable.

Moreover, the generation of vast datasets and continuous research in the field of machine learning for disease prediction further advances our understanding of disease mechanisms and potential therapeutic approaches. This knowledge is invaluable for both medical researchers and healthcare practitioners, as it can pave the way for innovative treatments and interventions. Machine learning techniques can uncover hidden insights within the data, leading to breakthroughs in disease understanding andmore effective treatment strategies.

CHAPTER 2

LITERATURE REVIEW

Senthil Kumar Mohan et al

[1] proposed Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques in which strategy that objective is to finding critical includes by applying Machine Learning bringing about improving the exactness in the expectation of cardiovascular malady.

The expectation model is created with various blends of highlights and a few known arrangement strategies. We produce an improved exhibition level with a precision level of 88.7% through the prediction model for heart disease with hybrid random forest with a linear model

(HRFLM) they likewise educated about Diverse data mining approaches and expectation techniques, Such as, KNN, LR, SVM, NN, and Vote have been fairly famous of late to distinguish and predict heart disease.

Sonam Nikhar et al

[2] has built up the paper titled as Prediction of Heart Disease Using Machine Learning Algorithms by This exploration plans to give a point by point portrayal of Naive Bayes and decision tree classifier that are applied in our examination especially in the prediction of Heart Disease. Some analysis has been led to think about the execution of prescient data mining strategy on the equivalent dataset, and the result uncovers that Decision Tree beats over Bayesian classification system.

Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD),

[3] Prediction of Heart Disease Using Machine Learning, in this paper proposed system they used the neural network algorithm multi-layer

perceptron (MLP) to train and test the dataset. In this algorithm there will be multiple layers like one for input, second for output and one or more layers are hidden layers between these two input and output layers.

Each node in input layer is connected to output nodes through these hidden layers. This connection is assigned with some weights. There is another identity input called bias which is with weight b, which added to node to balance the perceptron. The connection between the nodes can be feedforwarded or feedback based on the requirement.

Abhay Kishore et al

[4] developed Heart Attack Prediction Using Deep Learning in which This paper proposes a heart attack prediction system using Deep learning procedures, explicitly Recurrent Neural System to predict the probable prospects of heart related infections of the patient. Recurrent Neural Network is a very ground-breaking characterization calculation that utilizes Deep Learning approach in Artificial Neural Network. The paper talks about in detail the significant modules of the framework alongside the related hypothesis. The proposed model deep learning and data mining to give the precise outcomes least blunders. This paper gives a bearing and point of reference for the advancement of another type of heart attack prediction platform. Prediction stage.

Lakshmana Rao et al

[5] Machine Learning Techniques for Heart Disease Prediction in which the contributing elements for heart disease are more (circulatory strain, diabetes, current smoker, high cholesterol, etc..).

So, it is difficult to distinguish heart disease. Different systems in data mining and neural systems have been utilized to discover the seriousness of heart disease among people. The idea of CHD ailment is bewildering, in addition, in this manner, the disease must be dealt with warily. Not doing early identification, may impact the heart or cause sudden passing. The perspective of therapeutic science furthermore, data burrowing is used for finding various sorts of metabolic machine learning a procedure that causes the framework to gain from past information tests, models without being

expressly customized. Machine learning makes rationale dependent on chronicled information.

In recent years, the healthcare industry has seen a significant advancement in the field of data mining and machine learning. These techniques have been widely adopted and have demonstrated efficacy in various healthcare applications
particularly in the field of medical cardiology.

The rapid accumulation of medical data has presented researchers with an unprecedented opportunity to develop and test new algorithms in this field.

Heart disease remains a leading cause of mortality in developing nations and identifying risk factors and early signs of the disease has become an important area of research.
The utilization of data mining and machine learning techniques in this field can potentially aid in the early detection and prevention of heart disease.

The purpose of the study described by Narain et al. (2016) is to create an innovative machine-learning-based cardiovascular disease (CVD) prediction system in order to increase the precision of the widely
used Framingham risk score (FRS).

With the help of data from 689 individuals who had symptoms of CVD and a validation dataset from the Framingham research, the proposed system—which uses a quantum neural network to learn and recognize patterns of CVD—was experimentally validated and compared with the FRS.

The suggested system's accuracy in forecasting CVD risk was determined to be 98.57%, which is much greater than the FRS's accuracy of 19.22% and other existing techniques.

According to the study's findings, the suggested approach could be a useful tool for doctors in forecasting CVD risk

Assisting in the creation of better treatment plans, and facilitating early diagnosis.

In a study conducted by Shah et al. (2020) the authors aimed to develop a model for predicting cardiovascular disease using machine learning techniques.

The data used for this purpose were obtained from the Cleveland heart disease dataset, which consisted 303 instances and 17 attributes, and were sourced from the UCI machine learning repository.
The authors employed a variety of supervised classification methods, including

Bayes, decision tree, random forest, and k-nearest neighbour (KKN). The results of the study indicated KKN model exhibited the highest level of accuracy, at 90.8%.

The study highlights the potential utility of machine learning techniques in predicting cardiovascular disease, emphasizes the importance of selecting appropriate models and techniques to achieve optimal results. In a study by Drod et al. (2022) the objective was to use machine learning (ML) techniques to identify

CHAPTER-3

FEASIBILITY STUDY

3.1 DATA PREPROCESSING AND TRAINING:

Data preprocessing and training for heart disease prediction involves a comprehensive process of collecting and preparing healthcare data to build accurate predictive models.

This crucial phase encompasses tasks such as data cleaning, feature engineering, and splitting the dataset into training and testing sets. The objective is to transform raw medical data into a format suitable for machine learning algorithms, ensuring that these models can effectively predict the likelihood of various diseases based on individual health-related features. This process requires a combination of domain expertise, ethical considerations, and rigorous model development to revolutionize healthcare by enabling early disease detection and personalized healthcare management

3.1 TESTING ACCURACY MODULE:

The testing accuracy module for heart disease prediction plays a pivotal role in evaluating the performance of machine learning models developed for healthcare applications.

This module assesses the predictive accuracy, precision, recall, F1-score, and other relevant metrics to measure the model's ability to correctly identify disease risks. It provides a critical step in ensuring the reliability and effectiveness of the predictive models, thus contributing to early disease detection and improved healthcare management. Rigorous testing and validation are essential to confirm the model’s suitability for real-world healthcare scenarios, where accuracy an reliability are paramount for making informed decisions and personalized treatment.

3.2 CREATING WEB PAGE:

Creating a web page for heart disease prediction is a dynamic and multifaceted endeavor.

This webpage serves as a platform for users to input their health-related information, which is then processed through machine learning20 models to predict the risk of various diseases. The web page's design should be user-friendly, offering an

intuitive interface for data input, while the backend involves data preprocessing, model integration, and result presentation.

To ensure its effectiveness, the web page should prioritize data security, privacy, and accessibility, while also providing educational resources and insights to empower users in managing their health proactively.

Ultimately, the goal of this web page is to democratize healthcare by making advanced disease prediction accessible to a wide audience, thus contributing to early detection and improved healthcare outcomes

ABOUT PAGE MODULE

The "About" page module for heart disease prediction serves as a fundamental component of the project's user interface.

It offers an overview of the project's mission, goals, and team, providing essential context to users and stakeholders.

This section typically outlines the project's objectives, emphasizing the significance of early disease prediction, personalized healthcare, and data driven insights.

It may also introduce the core team members responsible for the project's development, including data scientists, healthcare professionals, and domain experts.

The "About" page is a pivotal point of reference, communicating the project's purpose and the collective effort to harness machine learning for improving healthcare through early disease detection and tailored healthcare management.

CHAPTER-4

PROJECT METHODOLOGY

4.1 DESCRIPTION OF THE WORKING FLOW OF PROPOSAL SYSTEM:

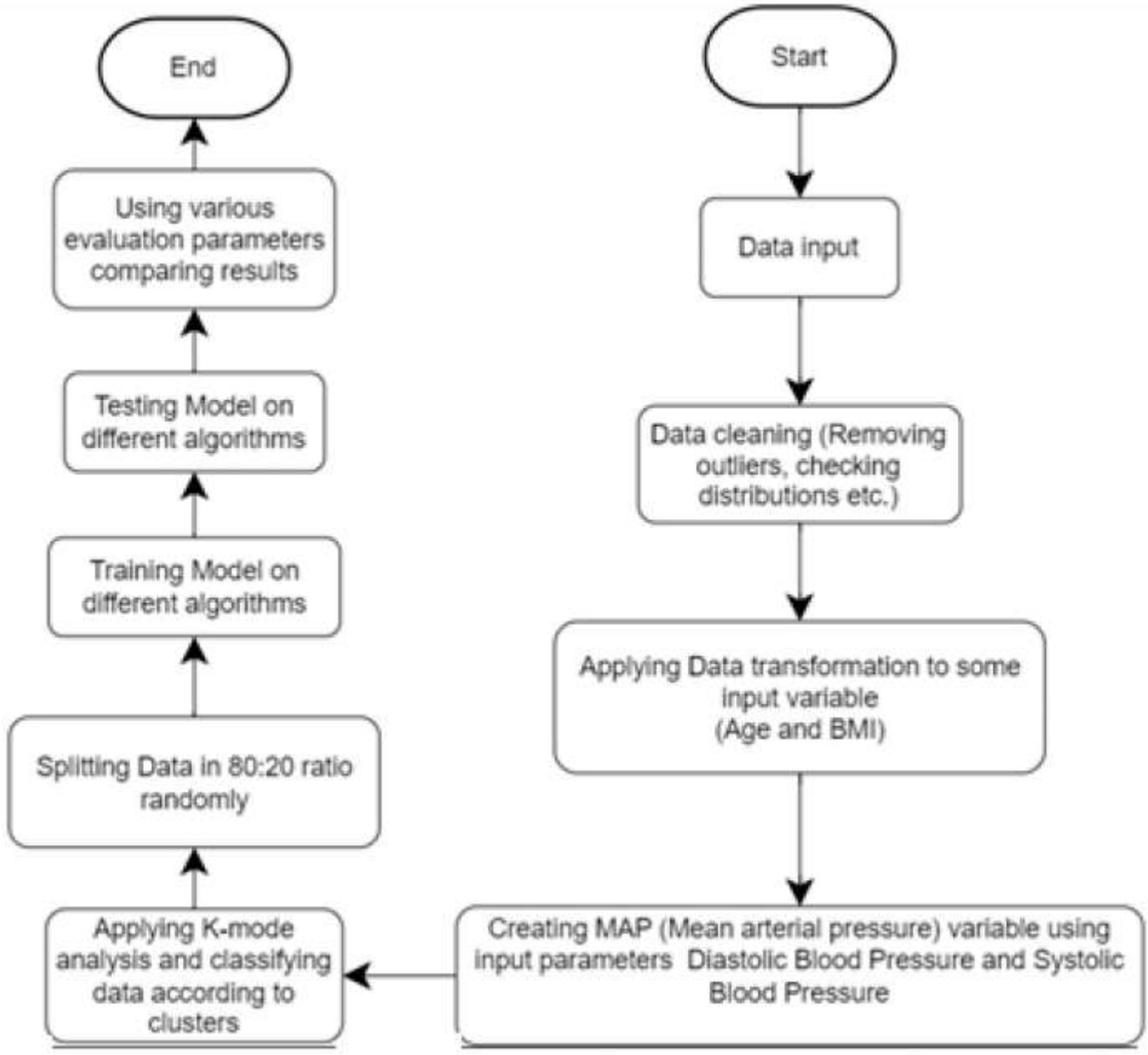

FIG.4.2: WORKING FLOW OF PROPOSAL SYSTEM

4.2 Data Collection:

An Organized Dataset of individuals had been selected Keeping in mind their history of heart problems and in accordance with other medical conditions [2]. Heart diseases are the diverse conditions by which the heart is affected. According to World Health Organization (WHO), the greatest numbers of deaths in middle aged people are due to cardiovascular diseases. We take a data source which is comprised of medical history of 304 different patients of different age groups. This dataset gives us the much-needed information i.e. the medical attributes such as age, resting blood pressure, fasting sugar level etc. of the patient that helps us in detecting the patient that is diagnosed with any heart disease or not. This dataset contains 13 medical attributes of 304 patients that helps us detecting if the patient is at risk of getting a heart disease or not and it helps us classify patients that are at risk of having a heart disease and that who are not at risk. This Heart Disease dataset is taken from the UCI repository. According to this dataset, the pattern which leads to the detection of patient prone to getting a heart disease is extracted. These records are split into two parts: Training and Testing.

4.3 Data Preprocessing:

- Describe the data cleaning process, which involves handling missing values, correcting errors, and addressing inconsistencies within the dataset.
- Explain the transformation steps, such as converting categorical variables to numerical values or standardizing units of measurement.
- Discuss normalization techniques to scale the data, making it suitable for analysis and ensuring that no particular feature dominates the prediction process due to its magnitude.
- Mention how outliers were identified and treated, either by removing or transforming them to minimize their impact on the models.
- Highlight the importance of maintaining data integrity throughout preprocessing and validating the final dataset's quality

4.4 Feature Selection and Engineering:

- Feature selection in heart disease prediction systems identifies the most important features of cardiac diseases. Feature selection can include filter, wrapper, embedded, and hybrid methods. The filter method is fast, while the wrapper and embedded methods consider feature dependencies and interact with classifiers.

- Heart disease prediction systems use machine learning algorithms to classify patients as likely to have heart disease. These systems use 14 medical attributes of a patient, trained under three algorithms: Logistic regression, KNN, and Random Forest Classifier.

4.5 Model Development:

- Specify the machine learning or AI models chosen for disease prediction, detailing the rationale behind their selection, such as logistic regression, randomforests, or neural networks.
- Describe the training process, including the partitioning of the dataset into training, validation, and test sets to evaluate model performance.
- Explain how performance metrics, like accuracy, sensitivity, specificity, and F1score, were used to assess the models and their respective advantages in disease prediction.
- Discuss the application of cross validation techniques to validate the model's robustness and reduce the risk of overfitting.
- Mention the process of hyperparameter tuning to optimize model performance and generalization.

4.6 Results and Discussion:

- Present the key findings, including the model's accuracy, sensitivity, specificity,and other relevant metrics, to measure its success in disease prediction.
- Interpret the results, highlighting any insights gained from the analysis, such asthe most influential features or factors in the prediction process.
- Discuss the clinical implications of the project, emphasizing how early diseasediagnosis can positively impact patient care and treatment outcomes.
- Address any limitations or challenges encountered during the project, such asdata availability or model performance constraints.
- Provide recommendations for further research and improvements to enhance thepredictive accuracy and real world applicability of the disease prediction system.

CHAPTER 5

RESULT AND DISCUSSION

The application of machine learning for heart disease prediction has the potential to revolutionize healthcare in several ways. By harnessing data driven insights, this transformative approach enhances patient care by developing accurate predictivemodels that identify disease risk factors and patterns at an early stage. This, in turn, allows for proactive interventions and personalized treatment strategies, leading to improved patient outcomes and a reduced burden of illness and healthcare costs.

Furthermore, the optimization of resource allocation within the healthcare system ensures that high risk individuals receive timely care, improving the overall healthcare experience while making it more accessible and cost effective. This dual benefit not only improves patient wellbeing but also eases the financial strain on healthcare providers and systems.

Additionally, the continuous generation of vast datasets and ongoing research in thefield of machine learning for disease prediction advances our understanding of disease mechanisms and potential therapeutic approaches. This knowledge is invaluable for both medical researchers and healthcare practitioners, leading to innovative treatmentsand interventions that can positively impact patient care.

In summary, the integration of machine learning in disease prediction has the potential to drive substantial improvements in healthcare delivery, patient outcomes, and our collective understanding of diseases. It represents a promising step toward a more proactive, personalized, and efficient healthcare system that benefits individualsand society as a whole.

The potential of machine learning in disease prediction extends beyond the immediate benefits mentioned. It also paves the way for a shift from a reactive healthcare model to a proactive one. By identifying disease risk factors early on, healthcare providers can implement preventive measures, lifestyle interventions,

and timely screenings, ultimately reducing the occurrence of diseases. This shift towards prevention not only saves lives but also reduces the strain on healthcare systems by lowering the demand for costly treatments and hospitalizations.

Moreover, machine learning driven disease prediction can support patient engagement and empowerment. When individuals are aware of their risk factors and have personalized strategies for disease prevention, they become active participants in their own healthcare. This can lead to healthier lifestyles, better adherence to treatmentplans, and improved overall wellbeing.

Ethical considerations and responsible data handling are integral to this transformative process. Ensuring data privacy and mitigating biases in machine learning models are vital to maintain trust and equity in healthcare. Transparency in how predictions are made and the use of patient data is crucial.

In essence, the adoption of machine learning for heart disease prediction is not just a technological advancement but a fundamental shift in healthcare philosophy. It empowers individuals, reduces healthcare costs, and advances our understanding of diseases, ushering in an era of more effective, patient centric, and data driven healthcare. The journey toward this healthcare revolution is ongoing, with ongoing research and collaboration between healthcare professionals, data scientists, and policymakers driving the field forward

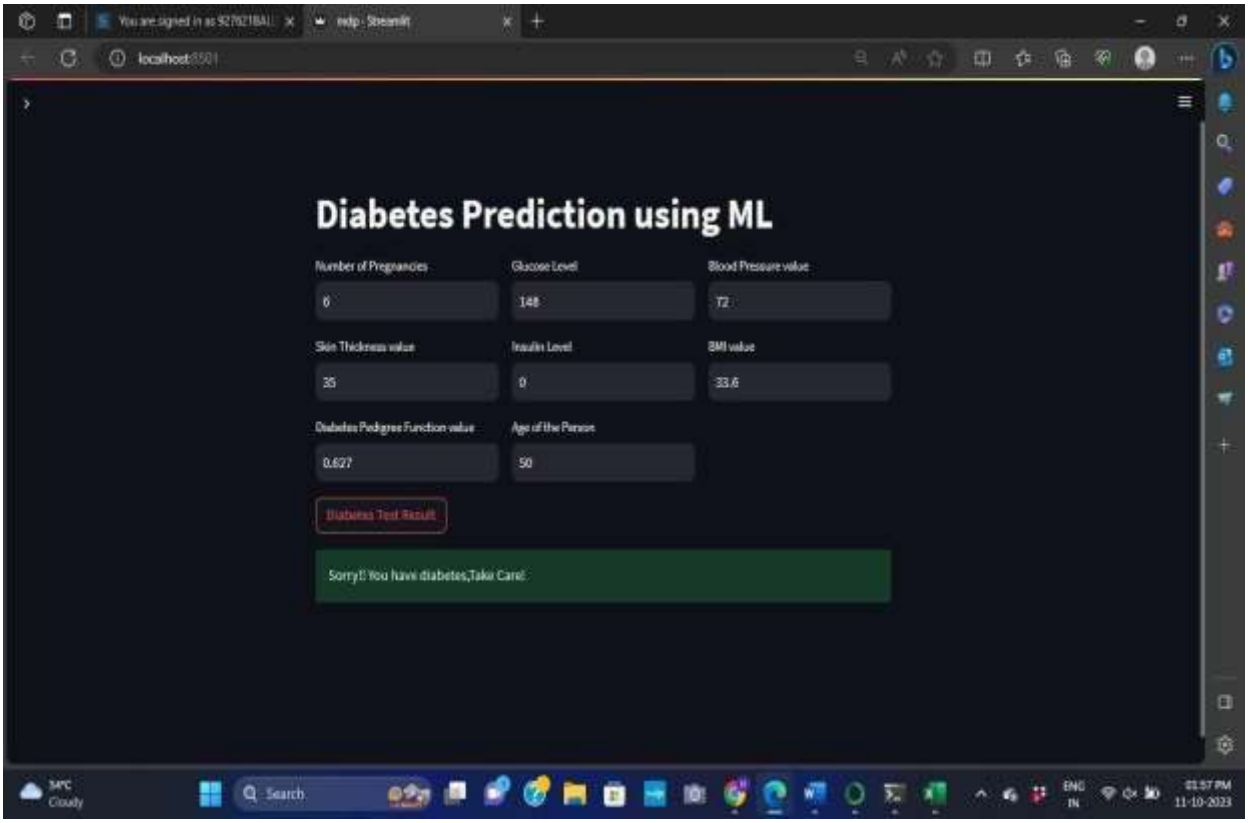

FIG.5.1: DIABETES DISEASE PREDICTION

Predicting the result of diabetes disease can be a critical medical endeavor. By analyzing a patient's medical history, lifestyle factors, and genetic predisposition, healthcare professionals can use advanced algorithms to estimate the likelihood of diabetes development. Early detection and intervention based on these predictions can significantly improve patient outcomes and reduce the risk of complications associated with diabetes. However, it's important to remember that predictions are not always 100% accurate, and regular monitoring and follow up are essential to confirm diagnosesand adjust treatment plans as needed. Ultimately, diabetes prediction plays a vital role in preventive healthcare and personalized medicine, aiding in the management of thisprevalent chronic condition.

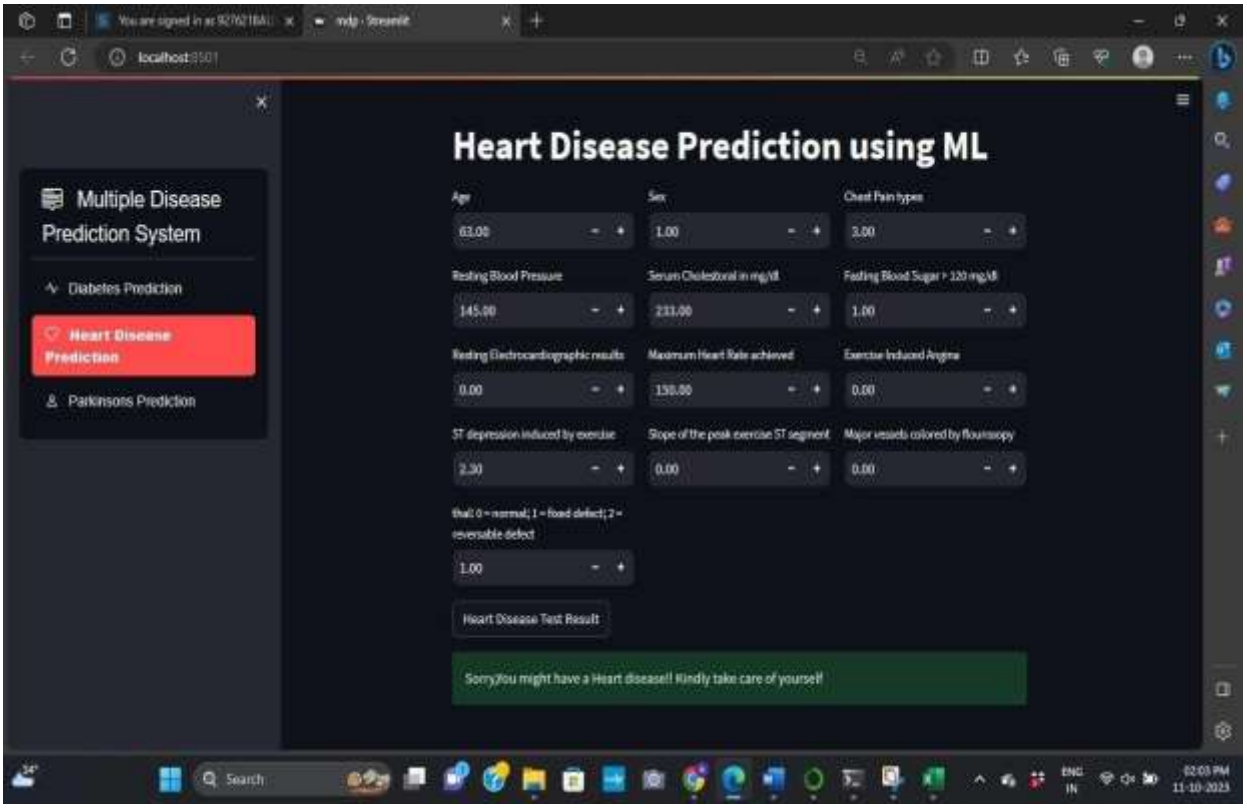

FIG.5.2: HEART DISEASE PREDICTION

Predicting the risk of heart disease involves analyzing various factors, such as age, lifestyle, and medical history. By leveraging advanced machine learning algorithms, we can accurately assess an individual's likelihood of developing heart disease. This predictive model empowers healthcare professionals to provide personalized recommendations for prevention and early intervention, reducing the overall burden of cardiovascular issues. Early detection allows for timely interventions, potentiallysaving lives and improving the quality of life for patients. These predictions contributeto a proactive approach to heart health, fostering better patient outcomes and healthcare resource allocation. In summary, heart disease prediction models offer a powerful tool to enhance preventive care and reduce the impact of this prevalent and serious medicalcondition.

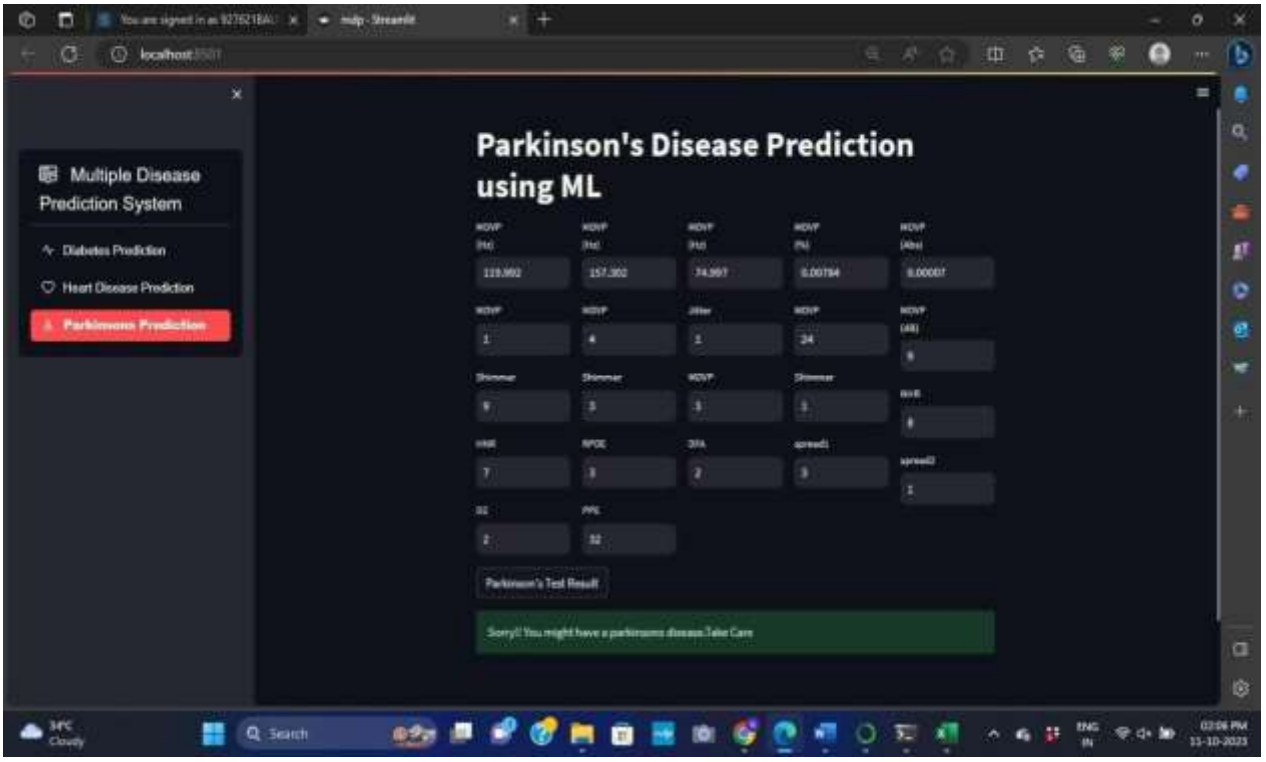

FIG.5.3: PARKINSON'S DISEASE PREDICTION

Predicting the results of Parkinson's disease diagnosis is a complex and multifaceted task. It typically involves a combination of clinical assessments, neurological examinations, and specialized tests. Early detection and accurate prediction are crucialfor optimizing treatment and improving the patient's quality of life. Machine learning algorithms and AI models have shown promise in assisting medical professionals in making more accurate predictions, utilizing data such as genetic markers, brain imaging, and motor function assessments. While these predictive tools are advancing, they still require further refinement and validation to enhance their reliability. The ultimate goal is to provide timely and precise diagnoses, enabling proactive interventions for individuals affected by Parkinson's disease.

CHAPTER 6

CONCLUSION

The primary objective of this study was to classify heart disease using different models and a real-world dataset. The k-modes clustering algorithm was applied to a dataset of patients with heart disease to predict the presence of the disease. The dataset was pre-processed by converting the age attribute to years and dividing it into bins of 5-year intervals, as well as dividing the diastolic and systolic blood pressure data into bins of 10 intervals. The dataset was also split on the basis of gender to take into account the unique characteristics and progression of heart disease in men and women.

The elbow curve method was utilized to determine the optimal number of clusters for both the male and female datasets. The results indicated that the MLP model had the highest accuracy of 87.23%. These findings demonstrate the potential of k-modes clustering to accurately predict heart disease and suggest that the algorithm could be a valuable tool in the development of targeted diagnostic and treatment strategies for the disease. The study utilized the Kaggle cardiovascular disease dataset with 70,000 instances, and all algorithms were implemented on Google Colab. The accuracies of all algorithms were above 86% with the lowest accuracy of 86.37% given by decision trees and the highest accuracy given by multilayer perceptron, as previously mentioned.

Limitations. Despite the promising results, there are several limitations that should be noted. First, the study was based on a single dataset and may not be generalizable to other populations or patient groups. Furthermore, the study only considered a limited set of demographic and clinical variables and did not take into account other potential

CHAPTER-7

REFERENCE

[1] Bo Jin ,Chao Che, Zhen Liu, Shulong Zhang, Xiaomeng Yin, And Xiaopeng Wei, "Predicting the Risk of Heart Failure With EHR Sequential Data Modeling" ,IEEE Access 2018.

[2] Aakash Chauhan , Aditya Jain , Purushottam Sharma , Vikas Deep, "Heart Disease Prediction using Evolutionary Rule Learning", "International Conference on "Computational Intelligence and Communication Technology" (CICT 2018).

[3] Ashir Javeed, Shijie Zhou, Liao Yongjian, Iqbal Qasim, Adeeb Noor, Redhwan Nour4, Samad Wali And Abdul Basit , "An Intelligent Learning System based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection" , IEEE Access 2017.

[4] Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access 2019.

[5] K.Prasanna Lakshmi, Dr. C.R.K.Reddy, "Fast Rule-Based Heart Disease Prediction using Associative Classification Mining", IEEE International Conference on Computer, Communication and Control (IC4-2015).

[6] M.Satish, D Sridhar, "Prediction of Heart Disease in Data Mining Technique", International Journal of Computer Trends & Technology (IJCTT), 2015.

[7] Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik, "An Intelligent Decision Support System for Cardiac Disease Detection", IJCTA, International Press 2015.

[8] Boshra Bahrami, Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", Journal of Multidisciplinary Engineering Science and Technology