

**CARDIO SENTINEL
PROGNOSTICATOR USING ML**

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering degree in Computer Science and
Engineering

by

RAVEEN S (Reg.No-40111045)

RAHUL M (Reg.No - 40111025)



**DEPARTMENT OF COMPUTER SCIENCE AND
ENGINEERING**

SCHOOL OF COMPUTING

SATHYABAMA

**INSTITUTE OF SCIENCE AND
TECHNOLOGY(DEEMED TO BE
UNIVERSITY)**

CATEGORY -1 UNIVERSITY BY UGC

**Accredited with Grade “A++” by NAAC | 12B Status by
UGC|Approved by AICTE JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119**

APRIL – 2024



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Category - I University by UGC

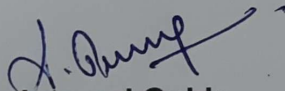
Accredited "A++" by NAAC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **RAVEEN S(40111045)** and **RAHUL M(40111025)** who carried out the Project entitled "**CADIO SENTINEL PROGONASTICATOR USING ML**" under our supervision from November 2023 to April 2024.

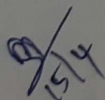

Internal Guide

Dr. ARUL PRAKASH A., M.Tech., Ph.D.

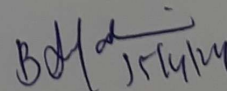

HEAD OF THE DEPARTMENT
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Dr. LAKSHMANAN M.E., Ph.D.
Jeppiaar Nagar, Rajiv Gandhi Sala,
Chennai - 600 119.

Submitted for Viva voce Examination held on

15/4/24


15/4

Internal Examiner


15/4/24

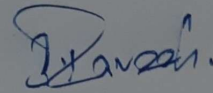
External Examiner

DECLARATION

I, **Rahul M** (Reg.No- 40111025), hereby declare that the Project Report entitled **CARDIO SENTINEL PROGNOSTICATOR USING ML**” done by me under the guidance of **Dr. ARUL PRAKASH A.,M.Tech.,Ph.D** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in **Computer Science and Engineering**.

DATE: 15/04/2024

PLACE: Chennai



SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph. D, Dean**, School of Computing, **Dr. L. Lakshmanan M.E., Ph.D.**, Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. ARUL PRAKASH A.,M.Tech.,Ph.D**, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

Myocardial Infarction and Brain attacks are responsible for the fatalities of individuals from cardiovascular diseases (CVDs), and especially the deaths occur before age 70. 17.9 million people are thought to pass away from CVDs annually. Accurate monitoring for each patient individually is not always possible, and clinicians cannot consult with patients every 24 hours due to the additional time and knowledge required. Using the patient's various cardiac characteristics and the machine learning approach of logistic regression on a publicly accessible dataset from Kaggle, we developed and examined models for predicting heart disease in this research. The main objective is to ascertain of acquiring coronary heart disease (CHD) up to 10 years of health risk. More than 4,000 records, 15 attributes, and patient data are included in the collection. To forecast outcomes, it makes predictions about a dependent variable based on one or more sets of independent variables. Both binary classification and multi-class classification can use it. This study aims to establish the most significant heart disease risk factors and estimate the overall risk using logistic regression. Machine Learning is used across many ranges around the world. The healthcare industry is no exclusion. Machine Learning can play an essential role in predicting presence/absence of locomotors disorders, heart diseases and more. Such information, if predicted well in advance, can provide important intuitions to doctors who can then adapt their diagnosis and dealing per patient basis. We work on predicting possible heart diseases in people using Machine Learning algorithms.

TABLE OF CONTENTS

Chapter No.	TITLE	Page No.
	ABSTRACT	v
	LIST OF FIGURES	viii
1	INTRODUCTION	1
2	LITERATURE SURVEY	3
	2.1 Inferences from Literature Survey	3
3	REQUIREMENTS ANALYSIS	6
	3.1 Machine Learning Algorithms	6
	3.2 Software & Hardware Requirements	10
	3.3 Functional & Non-Functional Requirements	10
4	DESCRIPTION OF PROPOSED SYSTEM	12
	4.1 Existing System	12
	4.2 Proposed System	13
	4.3 Use Case Diagram	13
	4.4 Class Diagram	15
5	IMPLEMENTATION DETAILS	17
	5.1 Collection of Dataset	22
	5.2 Selection of Attributes	23
	5.3 Pre-Processing of Data	24
	5.4 Balancing of Data	25
	5.5 Prediction of Disease	26
6	RESULT AND DISCUSSION	27
7	CONCLUSION AND FUTURE WORK	29
8	REFERENCES	33

APPENDIX	37
A. SOURCE CODE	37
B. SCREENSHORTS	43
C.RESEARCH PAPER	48

LIST OF FIGURES

FIGURE NO.	FIGURE NAME	PAGE NO
1.1	Machine Learning vs Traditional Programming	2
3.1	Machine Learning Algorithms	6
3.2	Logistic Regression	7
4.1	Use Case Diagram	14
4.2	Class Diagram	16
5.1	Agile Cycle Model	20
5.2	Collection of Dataset	23
5.3	Correlation of Matrix	23
5.4	Data Pre-Processing	24
5.5	Balancing of Data	25
5.6	Prediction of Disease	26

CHAPTER 1

INTRODUCTION

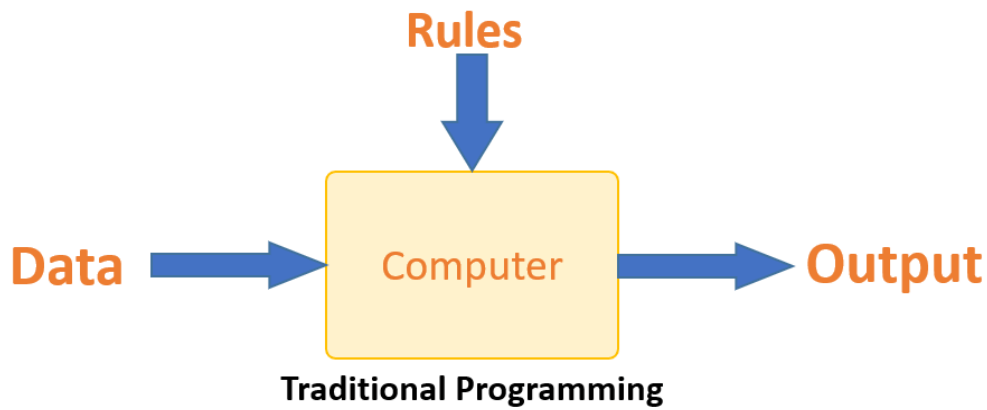
What is Machine Learning?

Machine Learning is a system of computer algorithms that can learn from example through self-improvement without being explicitly coded by a programmer. Machine learning is a part of artificial Intelligence which combines data with statistical tools to predict an output which can be used to make actionable insights. A typical machine learning tasks are to provide a recommendation. For those who have a Netflix account, all recommendations of movies or series are based on the user's historical data. Tech companies are using unsupervised learning to improve the user experience with personalizing recommendation. Machine learning is also used for a variety of tasks like fraud detection, predictive maintenance, portfolio optimization, automatize task and so on.

Machine Learning vs. Traditional Programming

Traditional programming differs significantly from machine learning. In traditional programming, a programmer code all the rules in consultation with an expert in the industry for which software is being developed. Each rule is based on a logical foundation; the machine will execute an output following the logical statement. When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain.

on a logical foundation; the machine will execute an output following the logical statement. When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain.



Traditional Programming

Machine learning is supposed to overcome this issue. The machine learns how the input and output data are correlated and it writes a rule. The programmers do not need to write new rules each time there is new data. The algorithms adapt in response to new data and experiences to improve efficacy over time.

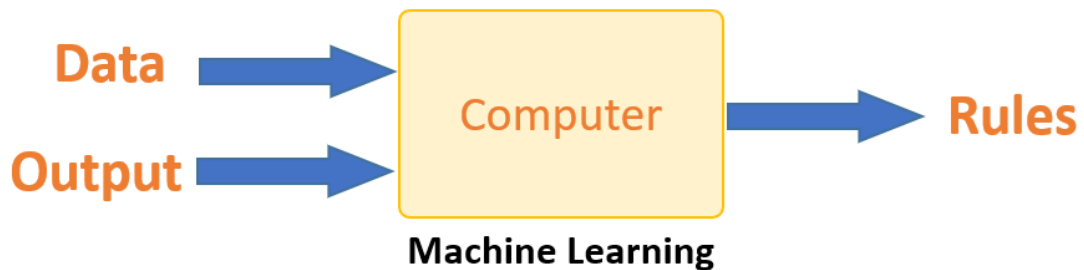


Fig : 1.1 Machine learning vs Traditional programming

The life of Machine Learning programs is straightforward and can be summarized in the following points:

1. Define a question
2. Collect data
3. Visualize data
4. Train algorithm
5. Test the Algorithm
6. Collect feedback
7. Refine the algorithm
8. Loop 4-7 until the results are satisfying
9. Use the model to make a prediction

Once the algorithm gets good at drawing the right conclusions, it applies that

CHAPTER 2

LITERATURE SURVEY

2.1 Inferences from Literature Survey

AUTHORS: Rairikar, A., Kulkarni, V., Sabale, V., Kale, H., & Lamgunde, A.

The healthcare industry collects large amounts of Healthcare data, but unfortunately not all the data are mined which is required for discovering hidden patterns and effective decision making. We propose efficient genetic algorithm with the back propagation technique approach for heart disease prediction. This paper has analyzed prediction systems for Heart disease using more number of input attributes. The System uses medical terms such as Gender, blood pressure, cholesterol like 13 attributes to predict the likelihood of patient getting a Heart disease.

AUTHORS: Gandhi, Monika, and Shailendra Narayan Singh

As huge amount of information is produced in medical associations (healing facilities, therapeutic focuses) yet this information is not properly utilized. The health care system is "data rich" however "knowledge poor ". There is an absence of successful analysis methods to find connections and patterns in health care data. Data mining methods can help as remedy in this circumstance. For this reason, different data mining techniques can be utilized. The paper intends to give details about various techniques of knowledge abstraction by using data mining methods that are being used in today's research for prediction of heart disease. In this paper, data mining methods namely, Naive Bayes, Neural network, Decision tree algorithm are analyzed on medical data sets using algorithms.

AUTHORS: Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, C. H., & Lin, E. J.

The diagnosis of heart disease in most cases depends on a complex combination of clinical and pathological data. Because of this complexity, there exists a significant amount of interest among clinical professionals and researchers regarding the efficient and accurate prediction of heart disease. In this paper, we develop a heart disease predict system that can assist medical professionals in predicting heart disease status based on the clinical data of patients. Our approaches include three steps. Firstly, we select 13 important clinical features, i.e., age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ecg, max heart rate, exercise induced angina, old peak, slope, number of vessels colored, and thal. Secondly, we develop an artificial neural network algorithm for classifying heart disease based on these clinical features. The accuracy of prediction is near 80%. Finally, we develop a user-friendly heart disease predict system (HDPS). The HDPS system will be consisted of multiple features, including input clinical data section, ROC curve display section, and prediction performance display section (execute time, accuracy, sensitivity, specificity, and predict result). Our approaches are effective in predicting the heart disease of a patient. The HDPS system developed in this study is a novel approach that can be used in the classification of heart disease.

AUTHORS: Aldallal, A., & Al-Moosa, A. A. A.

Modernization and commercialization of life lead to an unhealthy Lifestyle that results in increasing non-communicable diseases such as heart diseases and diabetes. Non-communicable diseases have direct impact on inaction, inactivity, and idleness of people. Heart diseases and diabetes are two of the most dangerous killers affecting the society. This research aims to produce application software to be used by doctors and other medical practitioners to predict the occurrence or recurrence of non-communicable diseases (NCDs).

The predictive data-mining model was applied in this project. Patients records obtained from Bahrain Defense Force Hospital were used to examine the proposed software application. This application was executed and tested by the actual practitioner in the mentioned hospital. The results showed that the prediction system is capable of predicting NCDs' diseases effectively, efficiently and most importantly, instantly. This application is capable of helping a physician in making proper decisions towards

AUTHORS: Sultana, Marjia, Afrin Haider, and Mohammad Shorif Uddin

Heart disease is considered as one of the major causes of death throughout the world. It cannot be easily predicted by the medical practitioners as it is a difficult task which demands expertise and higher knowledge for prediction. This paper addresses the issue of prediction of heart disease according to input attributes on the basis of data mining techniques. We have investigated the heart disease prediction using KStar, J48, SMO, Bayes Net and Multilayer Perceptron through Weka software. The performance of these data mining techniques is measured by combining the results of predictive accuracy, ROC curve and AUC value using a standard data set as well as a collected data set. Based on performance factor SMO and Bayes Net techniques show optimum performances than the performances of KStar, Multilayer Perceptron and J48 techniques.

CHAPTER 3

REQUIREMENT ANALYSIS

3.1 Machine Learning Algorithms and Where they are Used?

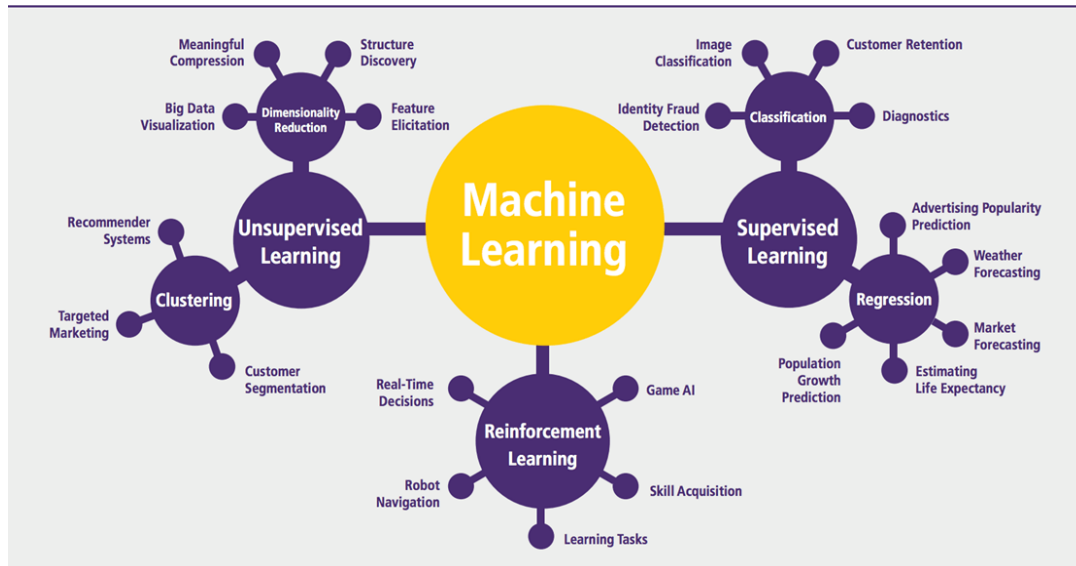


Fig.no: 3.1 Machine learning Algorithms

Machine learning can be grouped into two broad learning tasks: Supervised and Unsupervised. There are many other algorithms

Supervised learning

An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output. For instance, a practitioner can use marketing expense and weather forecast as input data to predict the sales of cans.

You can use supervised learning when the output data is known. The algorithm will predict new data.

There are two categories of supervised learning:

- Classification task
- Regression task

Regression

When the output is a continuous value, the task is a regression. For instance, a financial analyst may need to forecast the value of a stock based on a range of feature like equity, previous stock performances, macroeconomics index. The system will be trained to estimate the price of the stocks with the lowest possible error.

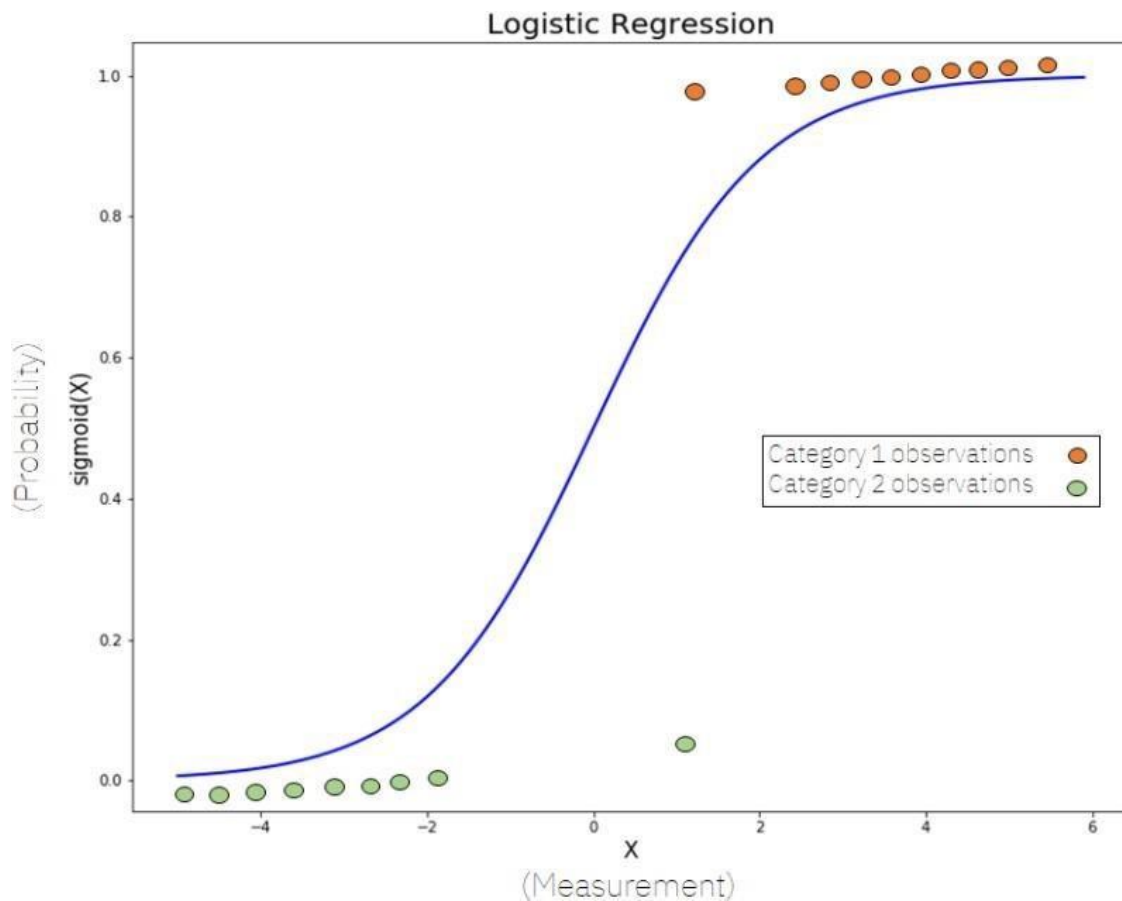


Fig.no 3.2 Logistic Regression

Algorithm Name	Description	Type
Linear regression	Finds a way to correlate each feature to the output to help predict future values.	Regression

Algorithm Name	Description	Type
Logistic regression	Extension of linear regression that's used for classification tasks. The output variable is binary (e.g., only black or white) rather than continuous (e.g., an infinite list of potential colors)	Classification
Decision tree	Highly interpretable classification or regression model that splits data-feature values into branches at decision nodes (e.g., if a feature is a color, each possible color becomes a new branch) until a final decision output is made	Regression Classification
Naive Bayes	The Bayesian method is a classification method that makes use of the Bayesian theorem. The theorem updates the prior knowledge of an event with the independent probability of each feature that can affect the event.	Regression Classification
Support vector machine	Support Vector Machine, or SVM, is typically used for the classification task. SVM algorithm finds a hyperplane that optimally divided the classes. It is best used with a non-linear solver.	Regression (not very common) Classification
Random forest	The algorithm is built upon a decision tree to improve the accuracy drastically. Random forest generates many times simple decision trees and	Regression Classification

Algorithm Name	Description	Type
	uses the 'majority vote' method to decide on which label to return. For the classification task, the final prediction will be the one with the most vote; while for the regression task, the average prediction of all the trees is the final prediction.	
AdaBoost	Classification or regression technique that uses a multitude of models to come up with a decision but weighs them based on their accuracy in predicting the outcome	Regression Classification
Gradient-boosting trees	Gradient-boosting trees is a state-of-the-art classification/regression technique. It is focusing on the error committed by the previous trees and tries to correct it.	Regression Classification

Unsupervised learning

In unsupervised learning, an algorithm explores input data without being given an explicit output variable (e.g., explores customer demographic data to identify patterns)

You can use it when you do not know how to classify the data, and you want the algorithm to find patterns and classify the data for you

Algorithm	Description	Type
K-means clustering	Puts data into some groups (k) that each contains data with similar characteristics (as determined by the model, not in advance by	Clustering

Algorithm	Description	Type
	humans)	
Gaussian mixture model	A generalization of k-means clustering that provides more flexibility in the size and shape of groups (clusters)	Clustering
Hierarchical clustering	Splits clusters along a hierarchical tree to form a classification system. Can be used for Cluster loyalty-card customer	Clustering

3.2 HARDWARE REQUIREMENTS:

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 1.44 Mb.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.
- Ram : 512 Mb.

SOFTWARE REQUIREMENTS:

- Operating system : Windows 7.
- Coding Language : Python
- Database : MYSQL

3.3 FUNCTIONAL REQUIREMENTS

- The model is about predicting if the person has heart disease or not by considering some parameters.
- There are 13 parameters that act as attributes,

which are the patients' information.

- These are considered as the functional requirements, which are required by the model.

NON-FUNCTIONAL REQUIREMENTS

- **Reliability:** The structure must be reliable and strong in giving the functionalities. The movements must be made unmistakable by the structure when a customer has revealed a couple of enhancements. The progressions made by the Programmer must be Project pioneer and in addition the Test designer.
- **Maintainability:** The system watching, and upkeep should be fundamental and focus in its approach. There should not be an excess of occupations running on diverse machines such that it gets hard to screen whether the employments are running without lapses.
- **Performance:** The framework will be utilized by numerous representatives all the while. Since the system will be encouraged on a single web server with a lone database server outside of anyone's ability to see, execution transforms into a significant concern. As the structure should not capitulate when various customers would use everything the while.

CHAPTER 4

DESCRIPTION OF PROPOSED SYSTEM

4.1 EXISTING_SYSTEM:

❖ Yun et al. presented different techniques for different type of feature selection, such as feature selection for high-dimensional small sample size data, large-scale data, and secure feature selection. They also discussed some important topics for feature selection have emerged, such as stable feature selection, multi-view feature selection, distributed feature selection, multi-label feature selection, online feature selection, and adversarial feature selection.

❖ Jundong et al. discussed the challenges of feature selection (FS) for big data. It is necessary to decrease the dimensionality of data for various learning tasks due to the curse of dimensionality. Feature selection has great influence in numerous applications such as building simpler, increasing learning performance, creating clean and understandable data. The feature selection from big data is challenging job and create big problems because big data has many dimensions. Further, challenges of feature selection for structured, heterogeneous and streaming data as well as its scalability and stability issues

DISADVANTAGES OF EXISTING SYSTEM:

❖ Existing approaches are low accuracy and high computation time and these might be due the use of irrelevant features in dataset. In order to tackle these problems new methods are needed to detect HD correctly. The improvement in prediction accuracy is a big challenge and research gap.

❖ Accuracy is very low

❖ Computationally complex

❖ More execution time required to generate results

4.2 PROPOSED SYSTEM:

- ❖ We proposed a machine learning based diagnosis method for the identification of HD in this research work. The standard state of the art features selection algorithms have been used to select the features.
- ❖ In the proposed system first try to address the problem of features selection by employing pre-processing techniques for appropriate subset of features and then applied these features for effective training and testing of the classifiers that identify which feature selection algorithm and classifier gives good results in term of accuracy and computation time. The classifiers performances have been compared on features selected by the standard state of the art FS algorithms with the selected features of the proposed FS algorithm. Finally, suggests that heart disease identification system effectively identify the HD.

ADVANTAGES OF PROPOSED SYSTEM:

- ❖ The system has been designed for the identification of heart disease. The performances of various machine learning classifiers for HD identification have been checked on selected features.
- ❖ The performance of the proposed method is high in terms of prediction accuracy.
- ❖ Low computation time.

4.3 USE CASE DIAGRAM

A use case diagram is a visual representation of the interactions between users (actors) and the system, showing the various use cases or functionalities provided by the system. In the context of the Heart Disease Prediction Project, the use case diagram helps identify the different types of users and the actions they can perform within the application.

The use case diagram for the Heart Disease Prediction Project includes two primary actors: User/Player and Quiz. The User/Player is the individual interacting with the system, accessing the quiz and viewing the results. The Quiz represents the functionality of the

system, which includes presenting questions, collecting answers, and calculating scores.

The use case diagram showcases the flow of actions from the User/Player accessing the quiz to viewing the result. It emphasizes the primary functionalities offered by the system, allowing users to interact with the quiz and obtain the prediction result. Each use case is associated with actors and describes the main flow of events, post-conditions, and possible termination scenarios. UML is the modeling toolkit that you can use to build your diagrams. Use cases are represented with a labeled oval shape. Stick figures represent actors in the process, and the actor's participation in the system is modeled with a line between the actor and use case.

To depict the system boundary, draw a box around the use case itself, A use case diagram doesn't go into a lot of detail—for example, don't expect it to model the order in which steps are performed. Instead, a proper use case diagram depicts a high-level overview of the relationship between use case.

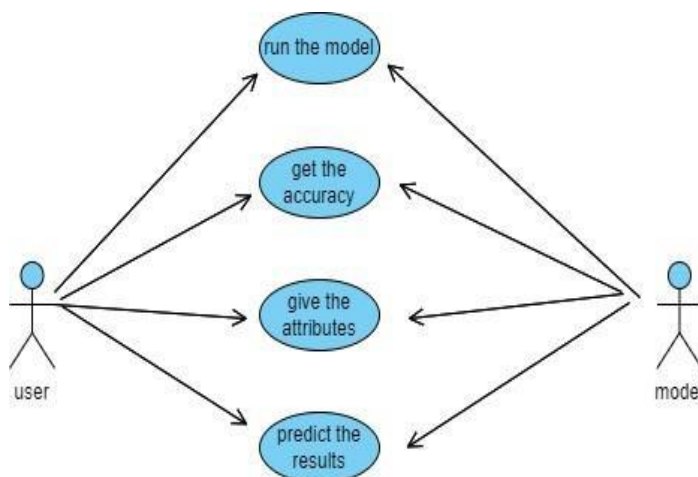


Fig.no : 4.1 Use Case Diagram

4.4 CLASS DIAGRAM

A class diagram is a structural diagram that depicts the classes, attributes, and relationships among objects in a system. In the Heart Disease Prediction Project, the class diagram provides a high-level view of the key entities and their attributes.

The class diagram for the Heart Disease Prediction Project includes two main classes: User/Player and Quiz. The User/Player class represents the users accessing the system, and it contains attributes such as name, age, and gender to capture user-specific information. The Quiz class represents the quiz functionality, with attributes like questions, answers, and score to manage the quiz-related data. The class diagram showcases the relationships between the classes, indicating that a User/Player can access the Quiz functionality. It helps in understanding the overall structure of the system and the dependencies between different entities.

Class diagram is a static diagram and it is used to model the static view of a system. The static view describes the vocabulary of the system. Class diagram

is also considered as the foundation for component and deployment diagrams. Class diagrams are not only used to visualize the static view of the system, but they are also used to construct the executable code for forward and reverse engineering of any system. In a class diagram, you would illustrate the relationships between classes using various types of associations such as aggregation, composition, generalization (inheritance), and dependency. You can also include additional details such as method signatures, access modifiers, and multiplicity (to represent cardinality) for associations.

When designing a class diagram for a disease prediction system, you would typically identify the main classes and their relationships within the system. It's important to note that the actual design of a class

diagram can vary depending on the specific requirements and context of your disease prediction system. The above examples provide a starting point, but you may need to adapt and extend the diagram based on your unique needs.

In the context of a disease prediction system, a class diagram would represent the different classes or entities involved in the system and their associations.

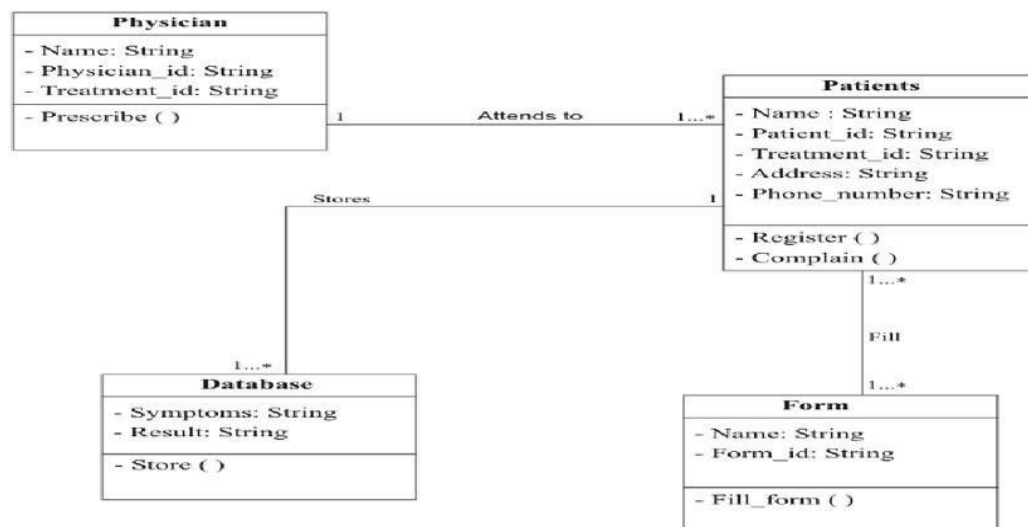


Fig.no : 4.2 Class Diagram

CHAPTER 5

IMPLEMENTATION DETAILS

AGILE DEVELOPMENT MODEL:

The Agile development model is an iterative and incremental approach to software development. It emphasizes flexibility, collaboration, and customer satisfaction. The Agile model breaks the development process into small iterations called sprints, where each sprint focuses on delivering a working software increment. The Agile model promotes adaptive planning, continuous improvement, and early delivery of valuable software. The Heart Disease Prediction Project adopts the Agile development model as its chosen methodology. Agile is a flexible and iterative approach that emphasizes collaboration, adaptability, and customer satisfaction. It allows for continuous improvement, early delivery of working software, and the ability to respond to changing requirements throughout the development process.

The Agile model is well-suited for this project as it involves developing a software application for heart disease prediction. By following the Agile principles, the development team can work closely with stakeholders, gather feedback, and make necessary adjustments to ensure the final product meets the desired objectives. The iterative nature of Agile allows for incremental development, where the project is divided into smaller iterations or sprints, each delivering a functional increment of the software. The Agile methodology promotes regular communication and collaboration among team members.

including developers, testers, and stakeholders. It encourages adaptive planning, where requirements are continuously refined and reprioritized based on feedback and changing needs. The development process is driven by user stories and frequent customer involvement, ensuring that the software aligns with the users' expectations. By adopting Agile, the Heart Disease Prediction Project benefits from increased flexibility, faster feedback loops, and the ability to address risks and issues early in the development cycle. It also allows for continuous integration and testing, ensuring a high-quality software product. However, it is important to effectively manage scope and prioritize features to avoid scope creep and deliver the most valuable functionality within the project's time and resource constraints.

ADVANTAGES OF AGILE DEVELOPMENT MODEL

- **Flexibility:** Agile allows for changes and modifications throughout the development process, accommodating evolving requirements.
- **Customer Satisfaction:** Agile involves the customer in the development process, ensuring their continuous feedback and satisfaction.
- **Iterative Development:** The project is divided into small iterations, allowing for frequent releases of working software and early feedback.
- **Superior quality product:** Testing is an integrated part of the project execution phase, which means that the overall quality of the final product is greater.
- **Adaptability:** Agile allows teams to adapt to changes and make adjustments based on feedback and

DISADVANTAGES OF AGILE DEVELOPMENT MODEL

- **Lack of Predictability:** Due to the flexible nature of Agile, it can be challenging to estimate and predict project timelines and costs accurately.
- **Resource Requirements:** Agile development requires a highly skilled and dedicated team, as well as frequent communication and collaboration.
- **Documentation:** Agile focuses more on working software rather than extensive documentation, which can be a challenge for maintaining comprehensive documentation.
- **Scope Creep:** Without proper management, the project scope can expand continuously during each iteration, leading to potential delays and increased costs.
- **Fragmented Output:** Incremental delivery may help bring products to market faster, but it's also a big disadvantage of Agile methodology.
- **Limited Documentation:** In Agile, documentation happens throughout a project and often "just in-time" for building the output, not at the beginning.
- **Limited Scope:** Agile Methodology focuses on delivering small increments of the software solution at a time.
- **More time and commitment:** Communication and collaboration is great but the constant interaction takes more time and energy for everyone involved.
- **Difficult Measurement:** Since Agile delivers in increments, tracking progress requires you to look across cycles. You can't set many requirements at the start of the

project.

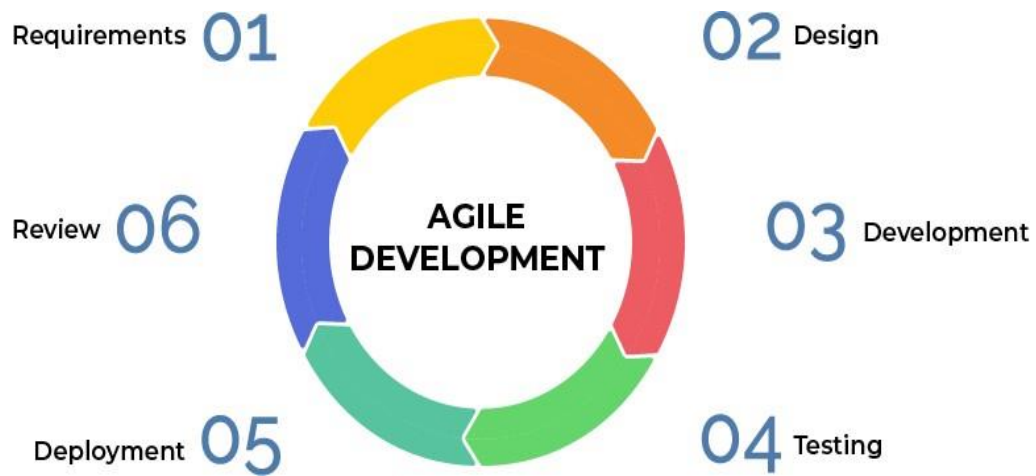


Fig.no : 5.1 Agile Cycle Model

REQUIREMENT

- Gather and analyze all the requirements for the Heart Disease Prediction project.
- Identify the stakeholders, their needs, and the expected functionality of the system.
- Document the requirements in a requirement specification document.

DESIGN

- Based on the requirements, design the system architecture for the Heart Disease Prediction project.
- Specify the hardware and system requirements necessary for the project.

DEVELOPMENT

- Develop the Heart Disease Prediction system in small units or modules.
- Each module should be developed and tested
- Unit testing should be performed to ensure that each module works as expected.

TESTING

- Integrate the developed modules into a complete Heart Disease Prediction system.
- Test each integrated module for any faults or failures.
- Conduct regular testing throughout the development process to ensure the system functions correctly.

DEPLOYMENT

- Once the Heart Disease Prediction system has undergone functional and non-functional testing, deploy it in the customer environment.
- Ensure the system is ready for use and meets the requirements of the stakeholders.
- Release the system into the market or customer's environment.
- Agile deployment may be seen as simply another testing step since multiple development deployments

AGILE DEVELOPMENT APPROACH

- Agile development emphasizes flexibility, adaptability, and iterative development.
- The Heart Disease Prediction project can be implemented using an Agile framework such as Scrum or Kanban.
- Instead of following a sequential waterfall model, Agile focuses on delivering working software in short iterations called sprints.
- The project team collaborates closely with stakeholders, continuously gathering feedback and adapting the system accordingly.
- The Agile software development cycle relies on the iteration or single development cycles.
- The development teams must keep the software running smoothly while also teaching users exactly how to use it.
- Agile development allows for regular inspection and adaptation, ensuring that the Heart Disease Prediction system meets evolving requirements and customer

5.1 COLLECTION OF DATASETS

Initially, we collected a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model. For this project, 70% of training data is used and 30% of data is used for testing. The dataset used for this project is Heart Disease UCI. The dataset consists of 76 attributes; out of which, 14 attributes are used for the system.

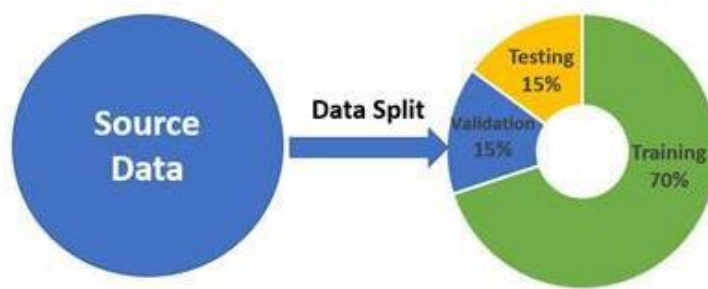


Fig.no :5.2 Collection of Datasets

5.2 SELECTION OF ATTRIBUTES

After the collection of the dataset, we split the dataset into training data and testing Attribute or Feature selection includes the selection of appropriate attributes for the prediction system. This is used to increase the efficiency of the system. Various attributes of the patient like gender, chest pain type, fasting blood pressure, serum cholesterol, exang, etc are selected for the prediction. The Correlation matrix is used for attribute selection for this model.

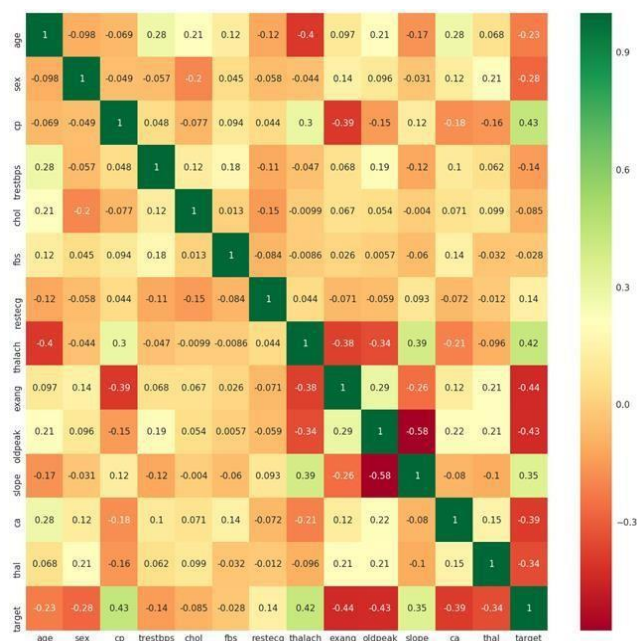


Fig.no: 5.3 Correlation matrix

5.3 PRE-PROCESSING OF DATA

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data pre-processing has activities like importing datasets, splitting datasets, attribute scaling, etc. Preprocessing is required to increase the efficiency and production of the given model.

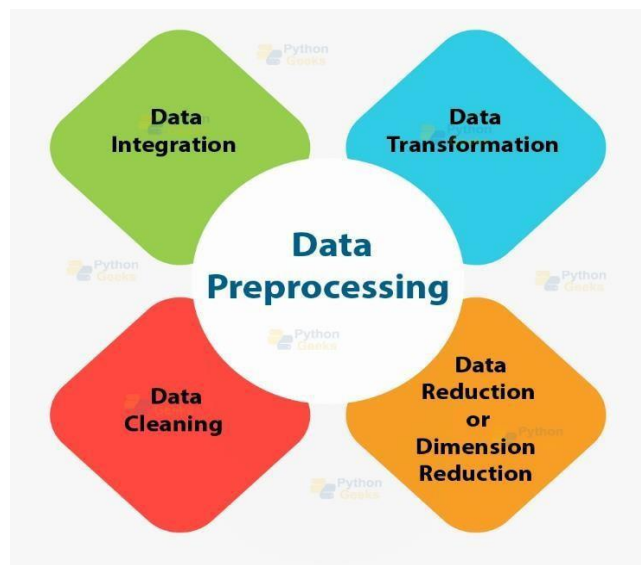


Fig.no : 5.4 Data Pre-Processing

5.4 BALANCING OF DATA

Imbalanced datasets can be balanced in two ways. They are Under Sampling and Over Sampling. In Under Sampling, dataset balance is done by the reduction of the size of the ample class. This process is considered when the amount of data is adequate. In Over Sampling, dataset balance is done by increasing the size of the scarce samples. This process is considered when the amount of data is inadequate.

Imbalanced data is not always a bad thing, and in real data sets, there is always some degree of imbalance. That said, there should not be any big impact on your model performance if the level of imbalance is relatively low. Over- sampling increases the number of minority class members in the training set. The advantage of over-sampling is that no information from the original training set is lost, as all observations from the minority and majority classes are kept. On the other hand, it is prone to overfitting.

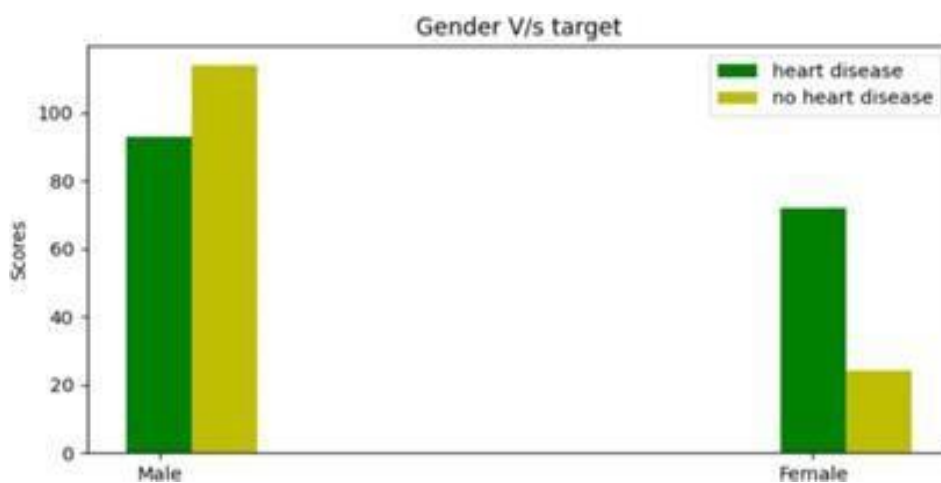


Fig.no : 5.5 Balancing of data

5.5 PREDICTION OF THE DISEASE

Various machine learning algorithms like SVM, Naive Bayes, Decision Tree, Random Tree, Logistic Regression, Ada-boost, Xg-boost are used for classification. Comparative analysis is performed among algorithms and the algorithm that gives the highest accuracy is used for heart disease prediction.

Machine learning model predictions allow businesses to make highly accurate guesses as to the likely outcomes of a question based on historical data, which can be about all kinds of things – customer churn likelihood, possible fraudulent activity, and more. These provide the business with insights that result in tangible business value. For example, if a model predicts a customer is likely to churn.

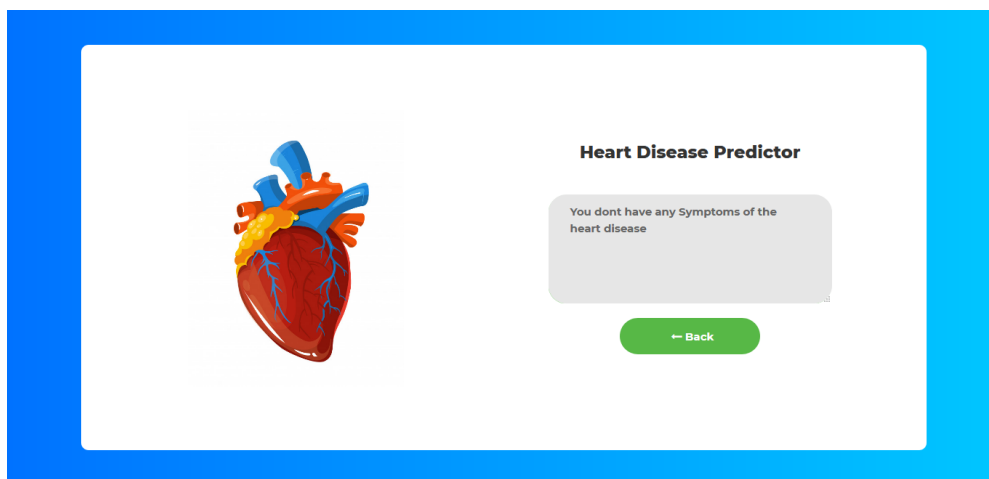


Fig.no : 5.6 Prediction of Disease

CHAPTER 6

RESULT AND DISCUSSION

In this study, an efficient machine learning based diagnosis system has been developed for the diagnosis of heart disease. Machine learning classifiers include LR, K-NN, ANN, SVM, NB, and DT are used in the designing of the system. Four standard feature selection algorithms including Relief, MRMR, LASSO, LLBFS, and proposed a novel feature selection algorithm FCMIM used to solve feature selection problem. LOSO cross-validation method is used in the system for the best hyperparameters selection. The system is tested on Cleveland heart disease dataset. Furthermore, performance evaluation metrics are used to check the performance of the identification system. According to the specificity of ANN classifier is best on Relief FS algorithm as compared to the specificity of MRMR, LASSO, LLBFS, and FCMIM feature selection algorithms. Therefore for ANN with relief is the best predictive system for detection of healthy people. The sensitivity of classifier NB on selected features set by LASSO FS algorithm also gives the best result as compared to the sensitivity values of Relief FS algorithm with classifier SVM (linear). The classifier Logistic Regression MCC is 91% on selected features selected by FCMIM FS algorithm. The processing time of Logistic Regression with Relief, LASSO, FCMIM and LLBFS FS algorithm best as compared to MRMR FS algorithms, and others classifiers. Thus the experimental results show that the proposed features selection algorithm select features that are more effective and obtains high classification accuracy than the standard feature selection algorithms. According to feature selection algorithms, the most important and suitable features are Thallium Scan type chest pain and Exercise-induced Angina. All FS algorithms results show that the feature Fasting blood sugar (FBS) is not a suitable heart disease diagnosis. The accuracy of SVM with the proposed feature selection algorithm (FCMIM) is 92.37% which is very good as compared previously proposed method the performance of machine learning based method FCMIMSVM is high then Deep neural network for detection of HD. A little improvement in prediction accuracy have great influence in diagnosis of critical diseases. The novelty of the study is developing a diagnosis system for

identification of heart disease. In this study, four standard feature selection algorithms along with one proposed feature selection algorithm is used for features selection. LOSO CV method and performance measuring metrics are used. The Cleveland heart disease dataset is used for testing purpose. As we think that developing a decision support system through machine learning algorithms it will be more suitable for the diagnosis of heart disease. Furthermore, we know that irrelevant features also degrade the performance of the diagnosis system and increased computation time. Thus another innovative touch of our study to used features selection algorithms to selects the appropriate features that improve the classification accuracy as well as reduce the processing time of the diagnosis system.

In the future, we will use other features selection algorithms, optimization methods to further increase the performance of a predictive system for HD diagnosis. The controlling and treatment of disease is significance after diagnosis, therefore, i will work on treatment and recovery of diseases in future also for critical disease such as heart, breast, Parkinson, diabetes.

CHAPTER 7

CONCLUSION AND FUTURE WORK

CONCLUSION

In conclusion, heart disease prediction projects hold great potential for improving risk assessment and early intervention. However, these projects come with their limitations and challenges. The accuracy of heart disease prediction models is influenced by the quality and representativeness of the data used, as well as the complexity and dynamic nature of heart disease itself. False positives and false negatives can occur, and predictive models may not capture all relevant risk factors or adequately account for temporal changes.

Ethical considerations surrounding privacy, consent, and potential discrimination also require careful attention. Additionally, integrating prediction models into clinical practice and ensuring their effective implementation pose practical challenges.

Despite these limitations, heart disease prediction projects offer valuable insights and can guide preventive strategies. Collaborative efforts between researchers, healthcare providers, and technology experts are essential for refining these models, incorporating additional risk factors, and addressing ethical concerns.

By continuously improving data quality, refining algorithms, and considering personalized risk assessment, heart disease prediction projects can contribute to more effective preventive measures, personalized interventions, and ultimately reduce the burden of heart disease on individuals and healthcare systems.

Despite these limitations, heart disease prediction projects offer valuable insights and can guide preventive strategies. Collaborative efforts between researchers, healthcare providers, and technology experts are essential for refining these models, incorporating additional risk factors, and addressing ethical concerns.

By continuously improving data quality, refining algorithms, and considering personalized risk assessment, heart disease prediction projects can contribute to more effective preventive measures, personalized interventions, and ultimately reduce the burden of heart disease on individuals and healthcare systems.

General physicians can utilize this tool for initial diagnosis of cardio-patients. There are many possible improvements that could be explored to improve the scalability and accuracy of this prediction system. As we have developed a generalized system, in future we can use this system for the analysis of different data sets.

The performance of the health's diagnosis can be improved significantly by handling numerous class labels in the prediction process, and it can be another positive direction of research. In DM warehouse, generally, the dimensionality of the heart database is high, so identification and selection of significant attributes for better diagnosis of heart disease are very challenging

FUTURE WORK

The correct prediction of heart disease can prevent life threats, and incorrect prediction can prove to be fatal at the same time. In this paper different machine learning algorithms and deep learning are applied to compare the results and analysis of the UCI Machine Learning Heart Disease dataset. The dataset consists of 14 main attributes used for performing the analysis. Various promising results are achieved and are validated using accuracy and confusion matrix. The dataset consists of some irrelevant features which are handled using Isolation Forest, and data are also normalized for getting better results. And how this study can be combined with some multimedia technology like mobile devices is also discussed. Using deep learning approach, 94.2% accuracy was obtained.

Heart disease prediction models have several limitations that need to be considered. First, the quality and availability of data can impact the accuracy of predictions. Inconsistent data collection practices, missing or incomplete data, and limited access to relevant variables can hinder the development of robust prediction models. Additionally, the predictive power of these models is limited by the complexity of heart disease, with numerous risk factors, variable interactions, and individual variability making it challenging to achieve perfect accuracy.

Furthermore, heart disease prediction models may not adequately capture the dynamic nature of risk factors, as they can change over time due to lifestyle modifications, medical interventions, and aging. This can lead to reduced accuracy and relevance of the models without regular updates and recalibration. Moreover, while some models incorporate genetic information, many predominantly focus on traditional risk factors, potentially overlooking the complex genetic component of heart disease risk.

There is also the issue of uncertainty associated with predictions. Despite advancements in predictive analytics, there is always a degree of uncertainty in predicting individual health outcomes. False positives and false negatives are possible, leading to both overdiagnosis and missed diagnoses, potentially causing unnecessary medical interventions or overlooking individuals at risk.

Moreover, ethical considerations arise regarding privacy, consent, and potential discrimination. Proper handling of sensitive health data, obtaining informed consent, and addressing bias and fairness concerns are essential to ensure the responsible and equitable use of heart disease prediction models.

Finally, the implementation and adoption of heart disease prediction models in real-world clinical settings pose challenges. Integrating these models into existing healthcare systems, training healthcare professionals, and ensuring effective communication of predictive results to patients require careful planning and considerations.

Predictive models rely on data inputs to make accurate predictions. If the dataset used to develop the model lacks comprehensive or up-to-date information, it may not capture all

REFERENCES:-

1. Abegunde DO, Mathers CD, Adam T, et al. The burden and costs of chronic diseases in low-income and middle-income countries. *Lancet* 2007;370:1929–38. [[PubMed](#)] [[Google Scholar](#)]
2. Barzi F, Patel A, Gu D, et al. Cardiovascular risk prediction tools for populations in Asia. *J Epidemiol Community Health* 2007;61:115–21. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
3. Conti AA, Minelli M, Gensini GF. Global management of high risk patients: integrated primary cardiovascular prevention in diabetics. *Int Congr Ser* 2003;207:10–20. [[Google Scholar](#)]
4. Chow CK, Joshi R, Celermajer DS, et al. Recalibration of a Framingham risk equation for a rural population in India. *J Epidemiol Community Health* 2009;63:379–85. [[PubMed](#)] [[Google Scholar](#)]
5. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci* 1963;107:539–56. [[PubMed](#)] [[Google Scholar](#)]
6. D'Agostino RB, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117:743–53. [[PubMed](#)] [[Google Scholar](#)]
7. D'Agostino RB, Grundy S, Sullivan LM, et al. Validation of the Framingham Coronary Heart Disease prediction scores. *JAMA* 2001;286:180–7. [[PubMed](#)] [[Google Scholar](#)]
8. Dalton M, Cameron AJ, Zimmet PZ, et al. Waist circumference, waist-hip ratio and body mass index and their correlation with cardiovascular disease risk factors in Australian adults. *J Intern Med* 2003;254:555–63. [[PubMed](#)] [[Google Scholar](#)]
9. Giampaoli S, Palmieri L, Mattiello A, et al. Definition of high risk individuals to optimise strategies for primary prevention of cardiovascular diseases. *Nutr Metab Cardiovas Dis* 2005;15:79–85. [[PubMed](#)] [[Google Scholar](#)]
10. Greenland P, Grundy S, Pasternak RC, et al. Problems on the pathway from risk assessment to risk reduction. *Circulation* 1998;97:1761–2. [[PubMed](#)] [[Google Scholar](#)]
11. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort. *BMJ* 2007;335:136–47. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
12. Jackson R. Cardiovascular risk prediction: are we there yet? *Heart* 2008;94:1–3. [[PubMed](#)] [[Google Scholar](#)]
13. Kinosian B, Glick H, Garland G. Cholesterol and coronary heart disease: predicting risks by levels and ratios. *Ann Intern Med* 1994;121:641–7. [[PubMed](#)] [[Google Scholar](#)]

14. Kannel WB, D'Agostino RB, Sullivan L, et al. Concept and usefulness of cardiovascular risk profiles. *Am Heart J* 2004;148:16–26. [[PubMed](#)] [[Google Scholar](#)]
15. Kannel WB. Clinical misconceptions dispelled by epidemiological research. *Circulation* 1995;92:3350–60. [[PubMed](#)] [[Google Scholar](#)]
16. Khaw K-T, Wareham N, Luben R, et al. Glycated haemoglobin, diabetes and mortality in men in Norfolk Cohort of European Perspective Investigation of Cancer and Nutrition (EPIC-Norfolk). *BMJ* 2001;322:15–20. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
17. L'Italien G, Ford I, Norrie J, et al. The Cardiovascular Event Reduction Tool (CERT)—a simplified cardiac risk prediction model developed from the West Scotland Coronary Prevention Study (WOSCOPS). *Am J Cardiol* 2000;85:720–4. [[PubMed](#)] [[Google Scholar](#)]
18. Lloyd-Jones DM, Leip EP, Larson MG, et al. Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. *Circulation* 2006;113:791–8. [[PubMed](#)] [[Google Scholar](#)]
19. Liu J, Hong Y, Ralph B, et al. Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-provincial Cohort Study. *JAMA* 2004;291:2591–9. [[PubMed](#)] [[Google Scholar](#)]
20. Matheny ME, Ohno-Machado L, Resnic FS. Discrimination and calibration of mortality risk prediction models in interventional cardiology. *J Biomed Inform* 2005;38:367–75. [[PubMed](#)] [[Google Scholar](#)]
21. Niccoli G, Iacoviello L, Cianflone D, et al. Coronary risk factors: new perspective. *Int J Epidemiol* 2001;30:S41–7. [[PubMed](#)] [[Google Scholar](#)]
22. O'Donnell CJ, Elosua R. Cardiovascular risk factors. Insights from Framingham Heart Study. *Rev Esp Cardiol* 2008;61:299–310. [[PubMed](#)] [[Google Scholar](#)]
23. Panagiotakos DB, Stavrinou V. Methodological issues in cardiovascular epidemiology: the risk of determining absolute risk through statistical models. *Vasc Health Risk Manag* 2006;2:309–15. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
24. Pastor-Barriuso R, Banegas JR, Damian J, et al. Systolic blood pressure, diastolic blood pressure and pulse pressure: an evaluation of their joint effect on mortality. *Ann Intern Med* 2003;139:731–9. [[PubMed](#)] [[Google Scholar](#)]
25. Ridker PM, Buring JE, Rifai N, et al. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA* 2007;297:611–9. [[PubMed](#)] [[Google Scholar](#)]
26. Redberg RF, Benjamin EJ, Bittner V, et al. ACCF/AHA 2009 performance measures for primary prevention of cardiovascular disease in adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on performance measures (writing committee to

- develop performance measures for primary prevention of cardiovascular disease): developed in collaboration with the American Academy of Family Physicians, American Association of Cardiovascular and Pulmonary Rehabilitation and Preventive Cardiovascular Nurses Association. *J Am Coll Cardiol* 2009;54:1364–405. [[PubMed](#)] [[Google Scholar](#)]
27. Swan HJC. The Framingham offspring Study: a commentary. *J Am Coll Cardiol* 1999;33:1136–40. [[PubMed](#)] [[Google Scholar](#)]
28. Suka M, Sugimori H, Yoshida K. Application of the updated Framingham Risk Score to Japanese men. *Hypertens Res* 2001;24:685–9. [[PubMed](#)] [[Google Scholar](#)]
29. Smith SC. Current and future directions of cardiovascular risk prediction. *Am J Cardiol* 2006;97:28–32. [[PubMed](#)] [[Google Scholar](#)]
30. Tonkin AM, Lim SS, Schirmer H. Cardiovascular risk factors: when should we treat? *Med J Aust* 2003;178:101–2. [[PubMed](#)] [[Google Scholar](#)]
31. Tunstall-Pedoe H. The Dundee Coronary Risk-Disk for management of change in risk factors. *BMJ* 1991;303:744–7. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
32. Ueshima H, Kasagi F, Kodama K, et al. Risk assessment chart for death from cardiovascular disease based on a 19-year follow-up study of a Japanese representative population—NIPPON DATA 80. *Circ J* 2006;70:1249–55. [[PubMed](#)] [[Google Scholar](#)]
33. Volpe M, Erhardt LR, Williams B. Managing cardiovascular risk: the need for change. *J Hum Hypertens* 2008;22:154–7. [[PubMed](#)] [[Google Scholar](#)]
34. Wallis EJ, Ramsey LE, Haq IU, et al. Is coronary risk an accurate surrogate for cardiovascular risk for treatment decisions in mild hypertension? A population validation. *Clin Pharmacol Ther* 2001;19:691–6. [[PubMed](#)] [[Google Scholar](#)]
35. Wang Z, Hoy WE. Waist circumference, body mass index, hip circumference and waist-to-hip ratio as predictors of cardiovascular disease in Aboriginal people. *Eur J Clin Nutr* 2004;58:888–93. [[PubMed](#)] [[Google Scholar](#)]
36. World Health Organization. *The top 10 leading causes of death by broad income group (2004)*. Geneva: World Health Organization, 2008. [[Google Scholar](#)]
37. Wu Y, Liu X, Li X, et al. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. *Circulation* 2006;114:2217–25. [[PubMed](#)] [[Google Scholar](#)]
38. Yusuf S, Reddy S, Ounpuu S, et al. Global Burden of Cardiovascular Diseases: Part II: Variations in cardiovascular disease by specific ethnic groups and geographic regions and prevention strategies. *Circulation* 2001;104:2855–64. [[PubMed](#)] [[Google Scholar](#)]
39. Yusuf S, Hawken S, Ounpuu S, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the

INTERHEART Study): case-control study. *Lancet* 2004;364:937–52.
[\[PubMed\]](#) [\[Google Scholar\]](#)

APPENDIX

A.Source code

```
1  from flask import Flask, render_template, flash, redirect
2  import joblib
3  from flask import request
4  import numpy as np
5  import pickle
6  import pandas as pd
7
8  app = Flask(__name__)
9
10 file = open('final_model.pkl', 'rb')
11 ml_model = pickle.load(file)
12
13 @app.route('/')
14
15 @app.route('/index')
16 def index():
17     return render_template('index.html')
18 @app.route('/login')
19 def login():
20     return render_template('login.html')
21 @app.route('/abstract')
22 def abstract():
23     return render_template('abstract.html')
24 @app.route('/elements')
25 def elements():
26     return render_template('elements.html')
27
28 @app.route('/chart')
29 def chart():
30     return render_template('chart.html')
```

```

31 @app.route('/upload')
32 def upload():
33     return render_template('upload.html')
34 @app.route('/preview',methods=["POST"])
35 def preview():
36     if request.method == 'POST':
37         dataset = request.files['datasetfile']
38         df = pd.read_csv(dataset,encoding = 'unicode_escape')
39         df.set_index('Id', inplace=True)
40         return render_template("preview.html",df_view = df)
41
42
43 @app.route('/home')
44 def home():
45     return render_template('heart.html')
46
47 @app.route("/predict",methods = ["POST"])
48 def predict():
49     if request.method == "POST":
50         print(request.form)
51         cp = request.form['cp']
52         trestbps = request.form['trestbps']
53         chol = request.form['chol']
54         fbs = request.form['fbs']
55         restecg = request.form['restecg']
56         thalach = request.form['thalach']
57         exang = request.form['exang']
58         pred_args = [cp,trestbps,chol,fbs,restecg,thalach,exang]
59         pred_args_arr = np.array(pred_args)
60         pred_args_arr = pred_args_arr.reshape(1,-1)

```

```

61         output = ml_model.predict(pred_args_arr)
62         print(output)
63
64         if (int(output)==1):
65             prediction = "You have Symptoms of getting heart disease. Please consult the Doctor"
66         else:
67             prediction = "You dont have any Symptoms of the heart disease"
68
69         return render_template("result.html",prediction_text=prediction)
70
71
72
73 if __name__ == "__main__":
74     app.run(debug=True)

```

```

1  <!DOCTYPE HTML>
2  <html>
3      <head>
4          <title>Cardio Sentinel Prognosticator</title>
5          <!-- <Link rel="icon" type="image/x-icon" href="D:\Project\New_folder\App\SOURCE_CODE\heart\templates\images\image2.jpg" />
6          <link rel="icon" href=" ../templates/images/fresh_hehy_food_melon_watermelon_fruit_icon_261718.ico" type="image/x-icon">
7          <meta charset="utf-8" />
8          <meta name="viewport" content="width=device-width, initial-scale=1" />
9          <link rel="stylesheet" href=" ../static/css/mains.css" />
10     </head>
11     <body>
12
13         <!-- Header -->
14         <header id="header">
15             <div class="logo"><a href="index.html">Cardio Sentinel <span>Prognosticator</span></a></div>
16             <a href="#menu"><span>Menu</span></a>
17         </header>
18
19         <!-- Nav -->
20         <nav id="menu">
21             <ul class="links">
22                 <li><a href="{{url_for('index')}}">Home</a></li>
23                 <li><a href="{{url_for('login')}}">Login</a></li>
24                 <li><a href="{{url_for('abstract')}}">Abstract</a></li>
25                 <li><a href="{{url_for('elements')}}">Future</a></li>
26
27                 <li><a href="{{url_for('chart')}}">chart</a></li>
28             </ul>
29         </nav>
30
31
32         <!-- Banner -->
33
34         <section id="banner" class="bg-img" data-bg=" ../static/images/image.jpg">
35             <div class="inner">
36                 <header>
37                     <h1> Cardio Sentinel Prognosticator </h1>
38                 </header>
39             </div>
40
41         </section>
42         <!-- Scripts -->
43         <script src=" ../static/js/jquery.min.js"></script>
44         <script src=" ../static/js/jquery.scrolly.min.js"></script>
45         <script src=" ../static/js/jquery.scrollex.min.js"></script>
46         <script src=" ../static/js/skel.min.js"></script>
47         <script src=" ../static/js/util.js"></script>
48         <script src=" ../static/js/mains.js"></script>
49
50     </body>
51 </html>

```

```

1  <!DOCTYPE html>
2  <html lang="en">
3
4  <head>
5
6      <meta charset="utf-8">
7      <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
8      <meta name="description" content="">
9      <meta name="author" content="">
10
11     <title>PREVIEW</title>
12
13     <!-- Custom fonts for this theme -->
14     <link href="../../static/vendor/fontawesome-free/css/all.min.css" rel="stylesheet" type="text/css">
15     <link href="https://fonts.googleapis.com/css?family=Montserrat:400,700" rel="stylesheet" type="text/css">
16     <link href="https://fonts.googleapis.com/css?family=Lato:400,700,400italic,700italic" rel="stylesheet" type="text/css">
17
18     <!-- Theme CSS -->
19     <link href="../../static/css/freelancer.min.css" rel="stylesheet">
20     <style>
21     #loading {
22     background: url('../../static/ajax-loader.gif') no-repeat center center;
23     position: absolute;
24     top: 0;
25     left: 99;
26     height: 100%;
27     width: 90%;
28     z-index: 9999999;
29     }
30     </style>
31
32 </head>
33
34 <body id="page-top">
35
36     <!-- Navigation -->
37     <nav class="navbar navbar-expand-lg bg-secondary text-uppercase fixed-top" id="mainNav">
38         <div class="container">
39             <a class="navbar-brand js-scroll-trigger" href="#page-top">Heart Disease Identification</a>
40             <button class="navbar-toggler navbar-toggler-right text-uppercase font-weight-bold bg-primary text-white rounded" type="button" data-toggle="collapse" data-target="#navbar" data-ks="true">
41                 Menu
42             </button>
43         </div>
44     </nav>
45
46     <!-- Contact Section -->
47     <section class="page-section" id="contact">
48         <div class="container">
49             <br>
50             <br>
51             <!-- Contact Section Heading -->
52             <h2 class="text-center text-uppercase text-secondary mb-0">Preview</h2>
53
54             <!-- Icon Divider -->
55             <div class="divider-custom">
56                 <div class="divider-custom-line"></div>
57                 <div class="divider-custom-icon">
58                     <i class="fas fa-star"></i>
59                 </div>
60             </div>

```

```

61     </div>
62     <div class="divider-custom-line"></div>
63     </div>
64
65     <!-- Contact Section Form -->
66     <div class="row" style="margin-left:-350px">
67         <div class="col-lg-8 mx-auto">
68             <!-- To configure the contact form email address, go to mail/contact_me.php and update the email address in the PHP file on line 19. -->
69             {{ df_view.to_html(classes="table striped",na_rep="-") | safe}}
70         </div>
71     </div>
72
73 </section>
74
75 <div class="form-group" style="padding:0px 250px 10px 40px;height:200px">
76     <input style="margin-left:500px" type="button" onclick="hideLoader()" class="btn btn-primary" value="Click to Train | Test" />
77     <div id="loading" style="display:none;margin-top:552950px"></div>
78 </div>
79
80 <!-- Copyright Section -->
81 <section class="copyright py-4 text-center text-white">

```

```

<!-- Bootstrap core JavaScript -->
<script src="../static/vendor/jquery/jquery.min.js"></script>
<script src="../static/vendor/bootstrap/js/bootstrap.bundle.min.js"></script>

<!-- Plugin JavaScript -->
<script src="../static/vendor/jquery-easing/jquery.easing.min.js"></script>

<!-- Contact Form JavaScript -->
<script src="../static/js/jqBootstrapValidation.js"></script>
<script src="../static/js/contact_me.js"></script>

<!-- Custom scripts for this template -->
<script src="../static/js/freelancer.min.js"></script>

<script type='text/javascript' src='https://ajax.googleapis.com/ajax/libs/jquery/2.2.4/jquery.min.js'></script>
<script type='text/javascript'>

function hideLoader() {
$( '#loading' ).show(0).delay(1000).hide(0,function(){
    alert("Training finished!");
    window.location = "{url_for('home')}}"";
});
}

```

```

1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4     <title>Heart Disease Predictor</title>
5     <meta charset="UTF-8">
6     <meta name="viewport" content="width=device-width, initial-scale=1">
7     <!-- =====>
8     <link rel="icon" type="image/png" href="images/icons/favicon.ico"/>
9     <!-- =====>
10    <link rel="stylesheet" type="text/css" href="{url_for('static',filename='vendor/bootstrap/css/bootstrap.min.css')}}">
11    <!-- =====>
12    <link rel="stylesheet" type="text/css" href="{url_for('static',filename='fonts/font-awesome-4.7.0/css/font-awesome.min.css')}}">
13    <!-- =====>
14    <link rel="stylesheet" type="text/css" href="{url_for('static',filename='vendor/animate/animate.css')}}">
15    <!-- =====>
16    <link rel="stylesheet" type="text/css" href="{url_for('static',filename='vendor/css-hamburgers/hamburgers.min.css')}}">
17    <!-- =====>
18    <link rel="stylesheet" type="text/css" href="{url_for('static',filename='vendor/select2/select2.min.css')}}">
19    <!-- =====>
20    <link rel="stylesheet" type="text/css" href="{url_for('static',filename='css/util.css')}}">
21    <link rel="stylesheet" type="text/css" href="{url_for('static',filename='css/main.css')}}">
22 </head>
23 </head>
24 <body>
25
26     <div class="contact1">
27         <div class="container-contact1">
28             <div class="contact1-pic js-tilt" data-tilt>
29                 
30             </div>

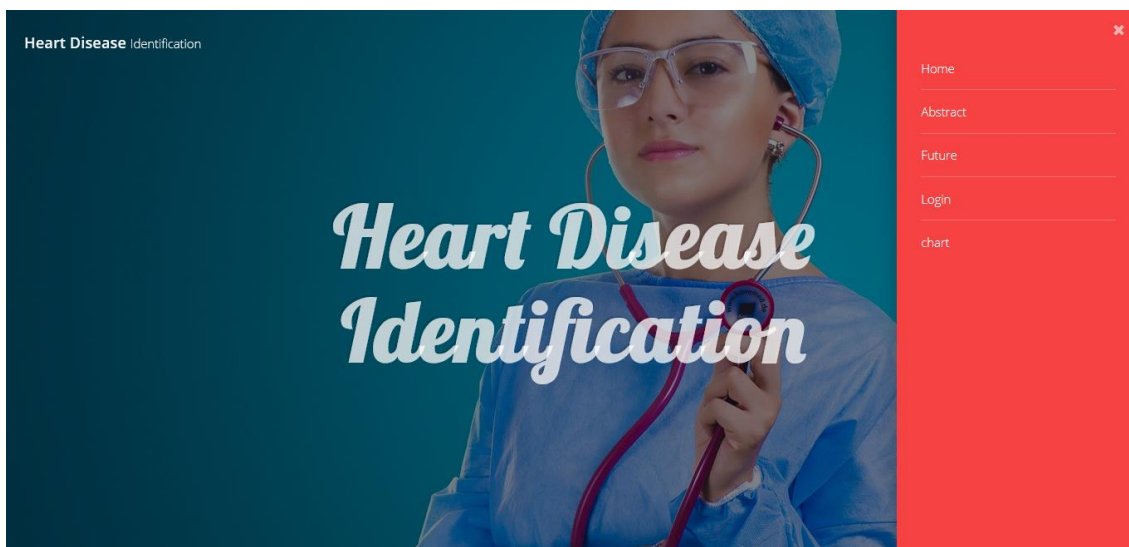
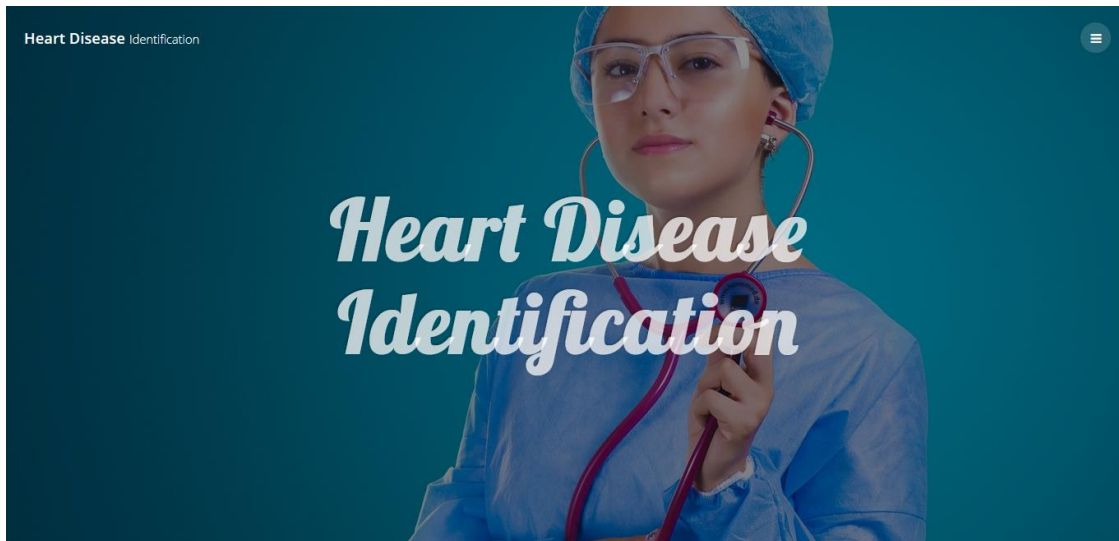
```

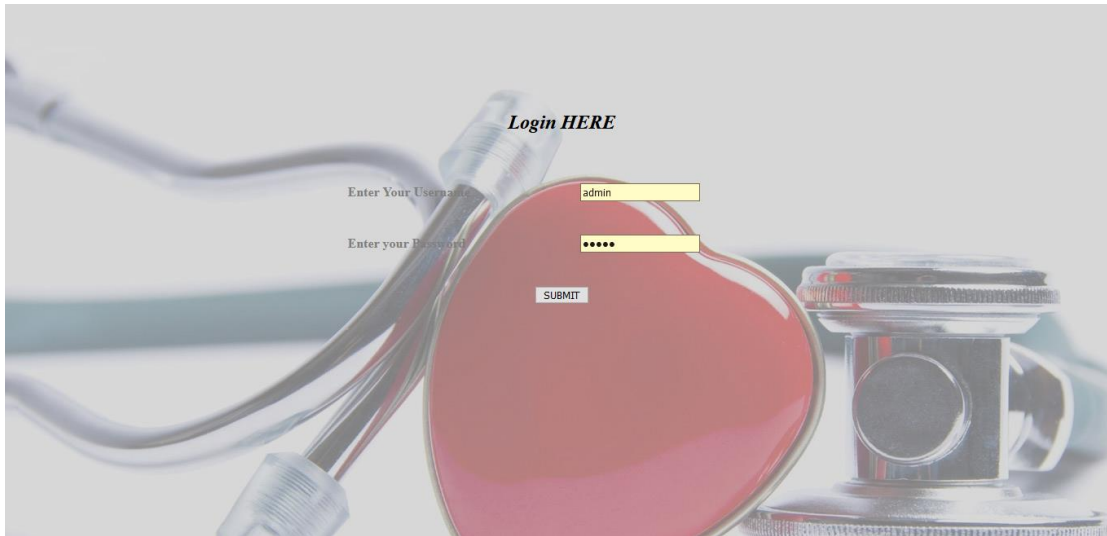
```

32 <form class="contact1-form validate-form" action="/predict" method="POST">
33 <span class="contact1-form-title" style="text-shadow: 2px;">
34   Heart Disease Predictor
35 </span>
36
37 <div class="wrap-input1 validate-input" data-validate = "Name is required">
38   <label for="cp">Chest Pain Type</label>
39   <select class="input1" id="cp" name="cp" required="required" placeholder="Chest Pain Type">
40     <option value="0">Typical Angina</option>
41     <option value="1">Atypical Angina</option>
42     <option value="2">Non-Anginal Pain</option>
43     <option value="3">Asymptomatic</option>
44   </select>
45   <span class="shadow-input1"></span>
46 </div>
47
48 <div class="wrap-input1 validate-input" data-validate = "Valid email is required: ex@abc.xyz">
49   <label for="trestbps"> Resting BP(mm Hg)</label>
50   <input class="input1" type="number" name="trestbps" id="trestbps" required="required" placeholder="Resting Blood Pressure">
51   <span class="shadow-input1"></span>
52 </div>
53
54 <div class="wrap-input1 validate-input" data-validate = "Subject is required">
55   <label for="chol">Serum Cholestoral (mg/dl)</label>
56   <input class="input1" type="number" step="any" id="chol" name="chol" required="required" placeholder="Serum Cholestoral">
57
58
59 <div class="wrap-input1 validate-input" data-validate = "Message is required">
60   <label for="restecg">Resting Electro-cardiographic Result</label>
61   <select class="input1" type="number" id="restecg" name="restecg" required="required" placeholder="Resting Electro-cardiographic">
62     <option value="0">Normal</option>
63     <option value="1">having ST-T wave abnormality</option>
64     <option value="2">showing probable or definite left ventricular hypertrophy</option>
65   </select>
66   <span class="shadow-input1"></span>
67 </div>
68
69 <div class="wrap-input1 validate-input" data-validate = "Subject is required">
70   <label for="thalach">Maximum Heart Rate Achieved</label>
71   <input class="input1" id="thalach" name = "thalach" required="required" placeholder="Max Heart Rate Achieved" >
72   <span class="shadow-input1"></span>
73 </div>
74
75 <div class="wrap-input1 validate-input" data-validate = "Message is required">
76   <label for="exang">Exercise Induced Angina</label>
77   <select class="input1" id="exang" name = "exang" required="required" >
78     <option value="1">Yes</option>
79     <option value="0">No</option>
80   </select>
81   <span class="shadow-input1"></span>
82 </div>
83
84
85
86
87
88
89
90
91

```

B. SCREENSHOTS





HEART DISEASE IDENTIFICATION

UPLOAD DATASET

———— ★ ————

Browse... hearts.csv

Upload

HEART DISEASE IDENTIFICATION

PREVIEW

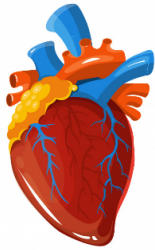


	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
Id														
1	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
2	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
3	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0

HEART DISEASE IDENTIFICATION

1023	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1024	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1025	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

[Click to Train | Test](#)



Heart Disease Predictor

Chest Pain Type
Typical Angina

Resting BP(mm Hg)
Resting Blood Pressure

Serum Cholestorol (mg/dl)
Serum Cholestorol

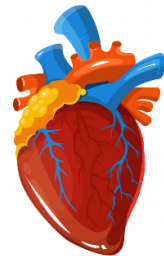
Fasting Blood Sugar
Fasting Blood Sugar < 120 mg/dl

Resting Electro-cardiographic Result
Normal

Maximum Heart Rate Achieved
Max Heart Rate Achieved

Exercise Induced Angina
Yes

Predict →



Heart Disease Predictor

Chest Pain Type
Typical Angina

Resting BP(mm Hg)
125

Serum Cholestorol (mg/dl)
212

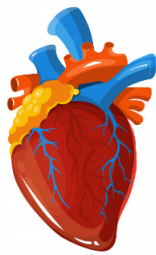
Fasting Blood Sugar
Fasting Blood Sugar < 120 mg/dl

Resting Electro-cardiographic Result
Normal

Maximum Heart Rate Achieved
165

Exercise Induced Angina
No

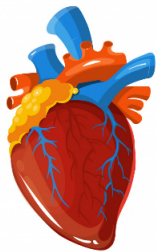
Predict →



Heart Disease Predictor

You dont have any Symptoms of the heart disease

← Back



Heart Disease Predictor

You have Symptoms of getting heart disease. Please consult the Doctor

← Back

C.RESEARCH PAPER

Cardio Sentinel Prognosticator Using ML

<i>Raveen^{1*}</i>	<i>Rahul²</i>	<i>Dr.A.Arul Prakash³</i>
<i>“Dept.of Computer Science and Engineering</i>	<i>“Dept.of Computer Science and Engineering</i>	<i>“Dept.of Computer Science and Engineering</i>
<i>Sathyabama Institute of Technology</i>	<i>Sathyabama Institute Science and Technology</i>	<i>ofSathyabama Institute of Science and Technology</i>
<i>Chennai,Tamilnadu, India.”</i>	<i>Chennai, Tamilnadu, India”</i>	<i>ChennaiTamilNadu,I ndia”.</i>
<i>raveen119raveen@gmail.c om</i>	<i>riorahul848352@gmail.com</i>	<i>prakash875@gmail. com.</i>

Abstract: Myocardial infarction and strokes are major contributors to the fatalities associated with cardiovascular diseases (CVDs), often occurring before the age of 70. Annually, an estimated 17.9 million people succumb to CVDs. Due to the limitations of accurate individual patient monitoring and the constraints on clinicians consulting patients every 24 hours, a more efficient approach is needed. In this research, we utilized a machine learning technique, specifically logistic regression, on a publicly available dataset from Kaggle, incorporating various cardiac characteristics to develop predictive models for heart disease. Assessing the risk of developing coronary heart disease (CHD) over a ten-year period was the main goal of our investigation. The dataset comprised over 4,000 records, featuring 15 attributes and patient data. For binary and multi-class classification, logistic regression—a technique for predicting a dependent variable based on sets of independent variables—was used. The main goals were to determine the most important heart disease risk factors and use logistic regression to estimate the total risk. Globally, machine learning is extensively used in many different fields, and the healthcare sector is no exception. It is useful in determining if cardiac conditions, movement abnormalities, and other conditions will manifest or not. Healthcare practitioners can customize their diagnostic and treatment plans for each patient with the help of accurate predictions, which can provide insightful information. Our work revolves around leveraging machine learning algorithms to predict potential heart diseases in individuals, contributing to the ongoing advancements in healthcare predictive analytics.

Keywords: cardiovascular diseases (CVDs), ML, Logistic regression, Kaggle.

I.INTRODUCTION

Heart disease claims 12 million lives globally, emerging as a predominant cause of mortality worldwide.

Rising Burden and Silent Threats

The escalating burden of cardiovascular disease over recent years underscores the urgency of predictive analytics in the realm of data analysis. Heart disease's often silent manifestation, leading to fatalities without evident symptoms, accentuates the need for proactive measures.

As cardiovascular diseases surge, the exploration of predictive analytics becomes imperative. Machine learning, a powerful tool in data analysis, stands out as a key player in forecasting heart-related ailments.

Project Objectives

Using patient data and machine learning algorithms, our initiative attempts to forecast future heart disorders by categorizing people according to whether or not they have heart disease. The goal is to create a predictive model that is both user-friendly and successful.

Logistic Regression: An Analytical Engine

Logistic regression, a popular supervised learning algorithm, takes center stage in our project. It predicts categorical dependent variables by fitting an "S"-shaped logistic function, offering probabilistic values between 0 and 1.

Advantages and Utility

Logistic regression's simplicity, ease of implementation, and efficiency in training make it a suitable choice. Its ability to provide insights into feature importance and associations enhances model interpretability.

Limitations and Challenges

While logistic regression offers advantages, it faces challenges such as overfitting on high-dimensional datasets and the inability to address non-linear problems. These limitations prompt the need for careful consideration and potential feature transformations.

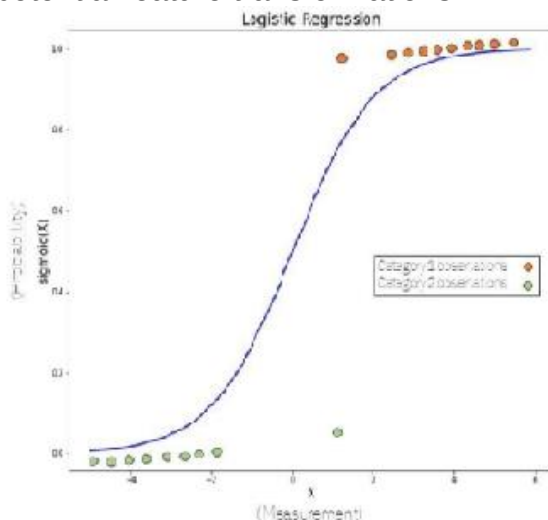


Fig.1 Logistic Regression

II.LITERATURE SURVEY

Accurate cardiovascular disease (CVD) risk assessment is paramount for effective preventive strategies and personalized healthcare. The need for a sufficiently large and representative sample cannot be overstated in ensuring the reliability of risk prediction models. Age, sex, and ethnicity-based sample stratification recognizes the differences in CVD risk factors among groups.

But disagreements emerge about how well-suited the current instruments are for a range of age groups, which forces one to carefully evaluate the inclusion requirements.

Age has a significant effect on the risk of CVD; Lloyd-Jones and colleagues found that if a person had cardiovascular risk factors, a critical age of 50 years old increases their lifetime risk of developing CVD. However, the underrepresentation of elderly people (70–80 years old) and younger adults (30–40 years old) in earlier studies raises concerns regarding the applicability of these techniques for these particular populations. A balance must be struck to prevent an unnecessary increase in the age range, potentially blunting the incidence of CVD events.

Sex-specific differences in risk factors necessitate the inclusion of representative samples from both men and women. Longitudinal prospective cohorts offer a valuable approach to observe actual outcomes based on existing lifestyles and risk factors, minimizing selection bias. This methodology aligns with the shift towards prevention medicine research, providing a robust foundation for assessing CVD risk.

Existing risk-assessment instruments, which were developed from research done in the 1970s and 1980s, continue to face difficulties. The high incidence of diabetes and CVD events in the historical setting restricts the applicability of these methods to the present day. A critical evaluation of a 5- or 10-year follow-up period becomes essential for clinically relevant risk management strategies.

Distinguishing between coronary heart disease (CHD) and overall CVD risks poses critical questions, as they are correlated but not equivalent. The utilization of the Cox proportional hazards model, with its incorporation of event timing, enhances the precision of risk assessments compared to logistic regression.

The selection of risk factors for inclusion in assessment tools remains a point of contention, influenced by measurement methods, resource availability, and their impact on predictive accuracy. Lifestyle factors, aside from smoking, are underrepresented, partly due to the challenges in their assessment and their limited independent significance in multivariate models.

Uncertainties about the relative relevance of cholesterol and blood pressure components in predicting CVD risk continue despite improvements. Innovations in risk assessment, such as exploring novel risk factors and adapting tools to local epidemiological data, are crucial for enhancing predictive accuracy and relevance.

In conclusion, while existing risk-assessment tools provide valuable insights into CVD risk, continuous validation, and adaptation to population trends are essential.

Addressing the challenges and uncertainties surrounding age groups, sex-specific differences, and evolving risk factors will contribute to the refinement and precision of CVD risk prediction models in diverse populations.

III.RELATED WORKS

Existing system:

Heart disease, often dubbed as a silent killer due to its ability to lead to a person's demise without apparent symptoms, has become a major source of concern. The elusive nature of the disease has sparked growing anxiety, prompting continuous efforts to predict its occurrence in the future. Many

instruments and methods are being continuously tested to meet the demands of modern medicine in response to this urgent health concern. In this endeavor, machine learning approaches show great promise. Although there are many different ways that heart disease might present itself, Many people are susceptible to the disease because they have a similar set of risk factors. The approach involves collecting data from diverse sources, categorizing them under relevant headings, and conducting a comprehensive analysis to extract pertinent information. This methodology is deemed effective in predicting heart disease, aligning with the adage, "Prevention is better than cure." Early prediction and subsequent intervention hold the potential to reduce mortality rates associated with heart disease.

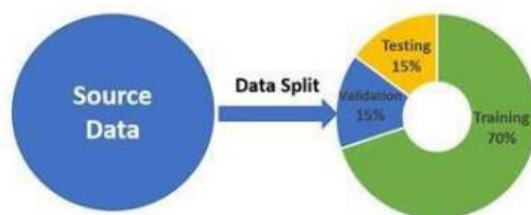
Proposed system:

The proposed system operates by initiating the process with data collection and the selection of crucial attributes. Preprocessing is done on the collected data to ensure that it is in the format needed. Separate sets of data are then created for testing and training purposes. The model is trained with the designated training data by applying machine learning methods. The system is then evaluated using separate testing data to find out how accurate it is. The predictive model is constructed with several modules, and each one helps it work smoothly. The systematic workflow ensures that relevant data is identified, processed, and utilized for training and testing purposes. The suggested method is an attempt to proactively address the significant health problem of heart disease by using machine learning algorithms to improve the accuracy of illness prediction.

IV.METHODOLOGY

Collection of data sets:

We gathered information for our heart disease prediction model throughout the first phase. After that, the dataset was divided into subgroups for testing and training. The training dataset is essential to the prediction model's learning process, whereas the testing data is vital for assessing the prediction model's efficacy. Thirty percent of the data in this project is set aside for testing, while the remaining seventy percent is used for training. Heart Disease UCI is the dataset used in this study, and the algorithm focuses on using twenty-four of the 76 attributes that are present in the



dataset.

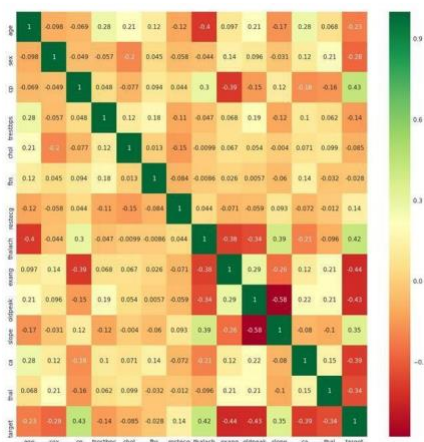
Fig.2 Data set

Attributes selection:

After acquiring the dataset, we proceeded to partition it into testing and training sets. The selection of appropriate attributes for the prediction system constitutes a crucial aspect of the attribute or feature selection process, aimed at enhancing the system's efficiency.

Numerous patient characteristics contribute to the prediction, encompassing

factors such as gender, type of chest pain, fasting blood pressure, serum cholesterol, and exang, among others. The attribute selection process of this model leverages the correlation matrix to discern the interrelationships



among various attributes.

Fig.3 Coorelation matrix

Data pre-processing

An essential part of developing a machine learning model is getting the data ready. Incorrect findings might have resulted from the model first receiving data that was not clean or properly prepared. Data must be transformed into the required format during pre-processing. Dataset issues including noise, duplication, and missing values may be remedied with it. Data pre-processing comprises activities such as loading and dividing datasets, scaling attributes, and comparable tasks. To improve the model's productivity and efficiency, preprocessing is necessary.

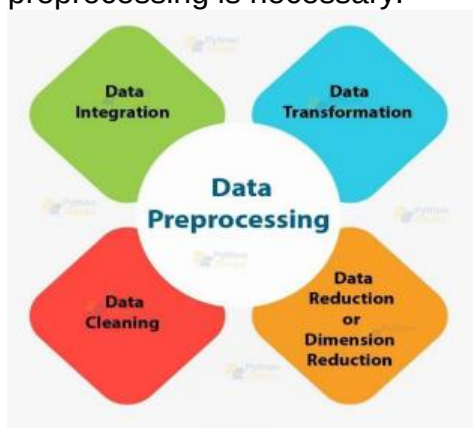


Fig.4 Data pre- processing

Balancing of data:

Using strategies like under- and oversampling is part of managing imbalanced datasets. When sufficient data is available, undersampling, which lowers the size of the majority class, produces balance. On the other hand, Over Sampling, which is used when the dataset is insufficiently informational, expands the size of the minority class. The type and volume of the available data determine which of these approaches is best.

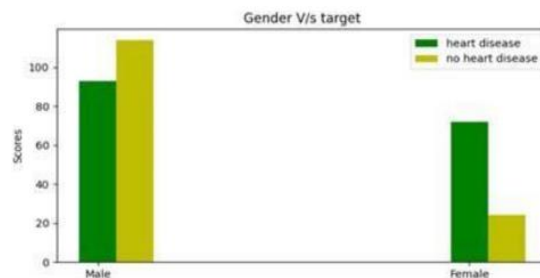
While imbalanced datasets are prevalent in real-world scenarios, their presence doesn't necessarily lead to a negative impact on model performance. A certain degree of imbalance is expected, and it typically doesn't significantly affect the model's performance, especially when the level

of imbalance is relatively low.

Over Sampling specifically focuses on augmenting instances from the minority class within the training set. Its advantage lies in retaining all observations from the original training set, encompassing both minority and majority classes. However, Over Sampling introduces a potential drawback: the risk of overfitting.

A model is said to have overfit when it incorporates noise or random fluctuations into its learning of the training data to an excessive degree. While Over Sampling ensures adequate representation of the minority class, it elevates the likelihood of the model tailoring itself too closely to the training data, potentially hindering generalization to new, unseen data.

Therefore, when choosing a balancing technique, it's essential to strike a balance between addressing class imbalance and averting overfitting. The objective is to enhance the model's capacity to recognize patterns in both minority and majority classes without compromising its ability to generalize effectively to new, unseen data. Each balancing method presents its advantages and trade-offs, and the selection should be guided by the specific characteristics and requirements of the dataset under



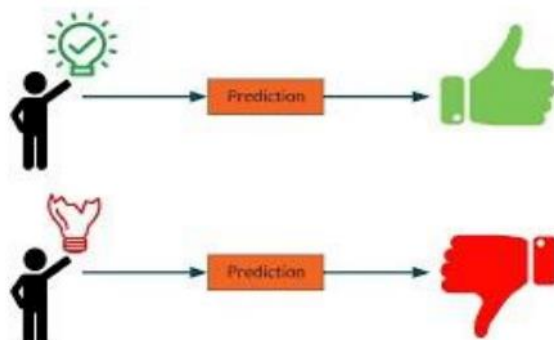
consideration.

Fig.5 Data balancing

Prediction of disease:

Heart disease prediction uses a variety of machine learning methods for classification, including SVM, Naive Bayes, Decision Tree, Random Tree, Logistic Regression, Ada-boost, and Xg-boost. To determine which of these algorithms has the best accuracy in predicting heart disease, a thorough comparison study is carried out.

In the realm of machine learning, predictive models generated by these algorithms empower businesses to make highly accurate predictions based on historical data. The applications of these predictions span a broad spectrum, encompassing areas like identifying potential fraudulent activity, predicting customer churn likelihood, and more. The tangible insights derived from these models deliver substantial business value, offering precise estimations, such as the probability of a customer



churning.

Fig.6 Prediction of Disease

Agile Architecture Methodology:

The Heart Disease Prediction Project was built using the Agile development style, which is known for being an incremental and iterative approach to software development. Emphasizing flexibility, collaboration, and customer satisfaction, Agile breaks down the development process into small, manageable iterations known as sprints, each dedicated to delivering a functional software increment. This framework promotes adaptive planning, continuous improvement, and the early delivery of valuable software.

The Heart Disease Prediction Project's alignment with the Agile development model is driven by its flexibility, iterative characteristics, and the emphasis on collaboration and adaptability. Adhering to Agile principles allows the development team to engage closely with stakeholders, collect timely feedback, and make necessary adjustments throughout the project's lifecycle. The iterative nature of Agile facilitates incremental development, with the project segmented into smaller sprints, each delivering a functional increment of the software.

The Agile methodology fosters contact between developers, testers, and stakeholders by encouraging regular communication and collaboration among team members. It facilitates adaptive planning, which allows requirements to be continuously improved and rearranged in accordance with input and changing needs. User stories serve as a roadmap for the development process, and customer involvement is encouraged to make sure that the software meets user expectations as intended.

Adopting the Agile model offers several advantages for the Heart Disease Prediction Project:

Flexibility: Agile accommodates changes and modifications throughout the development process to address evolving requirements.

Customer Satisfaction: Involving customers in the development process ensures continuous feedback and satisfaction.

Iterative Development: The project is divided into small iterations, allowing for frequent releases of working software and early feedback.

Superior Quality Product: Testing is integrated into the project execution phase, ensuring overall product quality.

Adaptability: Agile allows teams to adapt to changes and make adjustments based on feedback and lessons learned during development.

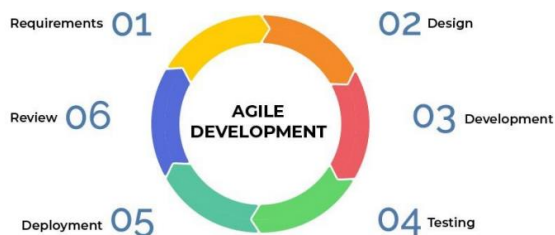


Fig.7 Agile Architecture methodology

Requirement:

Gathering and analyzing all requirements for the Heart Disease Prediction project involves identifying stakeholders, understanding their needs, and specifying the expected system functionality. These requirements are documented in a requirement specification document.

Design: Based on the requirements, the system architecture for the Heart

Disease Prediction project is designed. Hardware and system requirements are specified to guide the development process.

Development:

The system is constructed in manageable chunks, or modules, and the functionality of each module is tested separately. Every module is put through unit testing to make sure it functions as intended.

Testing:

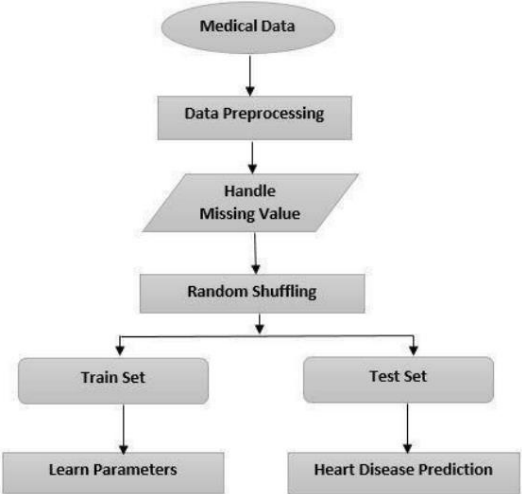
Integrated modules are tested for faults or failures. Regular testing throughout the development process ensures the system functions correctly.

Deployment:

Once the Heart Disease Prediction system undergoes testing, it is deployed in the customer environment. The system is released into the market or the customer's environment.

Review:

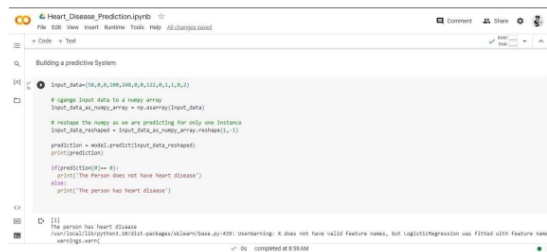
Ongoing maintenance, updates, and enhancements are performed based on customer feedback, ensuring the continuous improvement of the Heart Disease Prediction system. Addressing issues and releasing patches contribute to the system's robustness and effectiveness. Agile deployment is considered a continuous process, allowing for adjustments and improvements even after the initial



release.

Fig.8 Flow diagram

2) When the person has heart disease



```
Building a predictive system
In [ ]:
# Input data = (0, 0, 0, 100, 100, 0, 0, 1, 1, 0, 1)
# Change input data to a numpy array
input_data_numpy_array = np.array(input_data)
# Reshape the numpy array to be predicting for only one instance
input_data_reshaped = input_data_numpy_array.reshape(1,-1)
prediction = model.predict(input_data_reshaped)
print(prediction)

if(prediction[0]==0):
    print("The person does not have heart disease")
else:
    print("The person has heart disease")

Out [ ]:
The person has heart disease
Out [ ]: completed 0.000s
```

Fig.10 When the person have heart disease

V.CONCLUSION

In conclusion, heart disease prediction projects present significant potential for enhancing risk assessment and facilitating early intervention. However, these endeavors are not without their limitations and challenges. The accuracy of heart disease prediction models is intricately tied to the quality and representativeness of the data utilized, compounded by the complex and dynamic nature of heart disease itself. False positives and false negatives may arise, and predictive models may not comprehensively capture all relevant risk factors or adequately adapt to temporal changes.

Ethical considerations surrounding privacy, consent, and the potential for discrimination demand careful attention in the implementation of these projects. Additionally, integrating prediction models into clinical practice and ensuring their effective implementation pose practical challenges that must be navigated.

Despite these limitations, heart disease prediction projects offer valuable insights that can guide preventive strategies. Collaborative efforts between researchers, healthcare providers, and technology experts are indispensable for refining these models, incorporating additional risk factors, and addressing ethical concerns. A continuous commitment to improving data quality, refining algorithms, and embracing personalized risk assessment is crucial. This evolution in heart disease prediction projects has the potential to contribute to more effective preventive measures, personalized interventions, and ultimately alleviate the burden of heart disease on individuals and healthcare systems.

Future works:

As we go forward with this project, primary care doctors may be able to use this tool to make the first diagnosis of heart patients. Important factors to examine include investigating potential enhancements to the prediction system's scalability and accuracy. The developed system, being generalized, holds the potential for application in the analysis of diverse datasets.

A promising direction for future research involves improving the performance of health diagnosis by addressing numerous class labels in the prediction process. This approach represents a positive trajectory for further exploration and development in the field of heart disease prediction. Collaborative endeavors and a commitment to ongoing research will contribute to advancing the capabilities and impact of heart disease prediction projects, ultimately benefiting healthcare professionals and individuals alike.

REFERENCES

- [1] World Health Organization. The top 10 leading causes of death by broad income group (2004). Geneva: World Health Organization, 2008.

- [2] Abegunde DO, Mathers CD, Adam T, et al. The burden and costs of chronic diseases in low-income and middle-income countries. *Lancet* 2007;370:1929–38.
- [3] Giampaoli S, Palmieri L, Mattiello A, et al. Definition of high-risk individuals to optimise strategies for primary prevention of cardiovascular diseases. *Nutr Metab Cardiovas Dis* 2005;15:79–85.
- [4] Yusuf S, Hawken S, ōunpuu S, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART Study): case-control study. *Lancet* 2004;364:937–52.
- [5] Volpe M, Erhardt LR, Williams B. Managing cardiovascular risk: the need for change. *J Hum Hypertens* 2008;22:154–7.
- [6] Barzi F, Patel A, Gu D, et al. Cardiovascular risk prediction tools for populations in Asia. *J Epidemiol Community Health* 2007;61:115–21.
- [7] Ueshima H, Kasagi F, Kodama K, et al. Risk assessment chart for death from cardiovascular disease based on a 19-year follow-up study of a Japanese representative population—NIPPON DATA 80. *Circ J* 2006;70:1249–55.
- [8] Wu Y, Liu X, Li X, et al. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. *Circulation* 2006;114:2217–25.
- [9] Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci* 1963;107:539–56.
- [10] Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort. *BMJ* 2007;335:136–47.
- [11] Swan HJC. The Framingham offspring Study: a commentary. *J Am Coll Cardiol* 1999;33:1136–40.
- [12] D'Agostino RB, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117:743–53.
- [13] L'Italien G, Ford I, Norrie J, et al. The Cardiovascular Event Reduction Tool (CERT)—a simplified cardiac risk prediction model developed from the West Scotland Coronary Prevention Study (WOSCOPS). *Am J Cardiol* 2000;85:720–4.
- [14] Ridker PM, Buring JE, Rifai N, et al. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA* 2007;297:611–9.
- [15] Tunstall-Pedoe H. The Dundee Coronary Risk-Disk for management of change in risk factors. *BMJ* 1991;303:744–7.

CERTIFICATE



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
CATEGORY-1 UNIVERSITY BY UGC
Accredited 'A++' Grade by NAAC (I2B Status by UGC) Approved by AICTE



www.sathyabama.ac.in

SCHOOL OF COMPUTING

Department of Computer Science and Engineering

Certificate of Presentation

This is to certify that

RAVEEN.

of

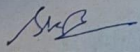
SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY.

has presented a paper entitled

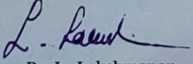
CARDIO SENTINEL PROGNOSTICATOR USING ML.

in the

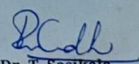
International Conference on Cognitive Computing and Artificial Intelligence (ICCCAI-2024) in association with Taylor's University, Malaysia held on 7th & 8th March 2024.



Dr. Sayan Kumar Ray
Head, School of Computer Science
Taylor's University



Dr. L. Lakshmanan
HOD, CSE Dept
Sathyabama Institute of Science and Technology



Dr. T. Sasikala
Dean, SOC
Sathyabama Institute of Science and Technology

