Heart Disease Prediction System Using Machine Learning and Python GUI

PROJECT REPORT

Submitted by

S.NO	Name	Registration Number
1.	Rahul Varma Dommaraju	BL.EN.U4ECE22046
2.	Raga Vinay	BL.EN.U4ECE22070
3.	Theertham Dheeraj	BL.EN.U4ECE22149
4.	Bommu Sridhar Reddy	BL.EN.U4ECE22162

23AIE232M: Project Phase

AMRITA SCHOOL OF COMPUTING, BENGALURU

AMRITA VISHWA VIDYAPEETHAM

BENGALURU 560 035

December -2024

TABLE OF CONTENTS **Page No.**

ACKNOWLEDGEMENTS

The satisfaction that accompanies successful completion of any task would be incomplete without mention of people who made it possible, and whose constant encouragement and guidance have been source of inspiration throughout the course of this project work.

We offer our sincere pranams at the lotus feet of **“AMMA”, MATA AMRITANANDAMAYI DEVI** who showered her blessing upon us throughout the course of this project work.

We owe our gratitude to **Prof. Manoj P.**, Director, Amrita Vishwa Vidyapeetham Bengaluru Campus. We would like to place our heartfelt gratitude to **Dr. Gopalakrishnan E.A.,** Principal, Amrita School of Computing and Amrita School of Artificial Intelligence, Bengaluru for his valuable support and inspiration.

It is a great pleasure to express our gratitude and indebtedness to our project guide Dr. Nidhin Prabhakar T. V, Department of Computer Science and Engineering, Amrita School of Computing, Bengaluru for her/his valuable guidance, encouragement, moral support, and affection throughout the project work.

We would like to thank express our gratitude to project panel members for their suggestions, encouragement, and moral support during the process of project work and all faculty members for their academic support. Finally, we are forever grateful to our parents, who have loved, supported and encouraged us in all our endeavors.

ABSTRACT

This project focuses on developing a **Heart Disease Prediction System** using machine learning models and a user-friendly graphical user interface (GUI). The system integrates various classification algorithms, including **Decision Tree, Random Forest, XGBoost, and a Stacking Classifier**, trained on the publicly available *Heart Disease dataset*. The dataset was preprocessed to handle missing values, scale features, and split into training and testing sets for model evaluation. After training, the models achieved high accuracy, with several models, such as Decision Tree, Random Forest, and XGBoost, reaching a perfect accuracy of 100% on the test data. The trained models were saved as .pkl files for real-time predictions through the GUI.

The GUI, built with Python's Tkinter library, allows users to input relevant health parameters such as age, cholesterol level, and chest pain type, and choose a preferred machine learning model. Based on these inputs, the system predicts whether the individual is at risk of heart disease. Additionally, visualizations like confusion matrices and accuracy comparison charts highlight the performance of each model, offering insights into their reliability. This tool demonstrates the practical application of machine learning in healthcare, providing an accessible, efficient, and accurate method for preliminary heart disease screening.

LIST OF FIGURES Page no.

LIST OF TABLES Page no.

The paper is structured as follows:
• **Section 1 - Introduction**: Introduces the context, challenges, and objectives of the study.
• **Section 2 - Literature Survey**: Reviews existing research and identifies gaps in load-balancing algorithms.
• **Section 3 - Requirements and Analysis**: Describes the software and hardware prerequisites for the study.
• **Section 4 - Design**: Outlines the architecture and workflow of the simulation environment.
• **Section 5 - Implementation**: Details the practical application of the algorithms and simulation setup.
• **Section 6 - Results**: Presents the findings and analysis of algorithm performance.
• **Section 7 - Conclusion and Future Scope**: Summarizes the results and proposes areas for further research.
• **References**: Lists the sources and studies cited in the paper.

CHAPTER - 1

INTRODUCTION

Heart diseases are a leading cause of mortality worldwide, affecting millions of lives each year. Conditions such as coronary artery disease, heart attacks, and arrhythmias often progress silently, making early detection vital. Traditional diagnostic methods, though effective, can be resource-intensive and time-consuming, relying heavily on specialized expertise. With the growing availability of healthcare data, machine learning (ML) has emerged as a transformative tool in addressing these challenges. By analyzing complex patterns within medical datasets, ML models offer fast, data-driven insights that support clinicians in making accurate predictions and informed decisions.

In this project, we present a **Heart Disease Prediction System** that leverages advanced ML algorithms to predict the likelihood of heart disease based on key health metrics. The system integrates multiple high-performing models, including Decision Tree, Random Forest, and XGBoost, which achieved exceptional accuracy during testing. Robust preprocessing ensures that missing data and scaling challenges are effectively handled, enabling reliable predictions. A user-friendly graphical user interface (GUI) allows users to input parameters such as age, cholesterol levels, and chest pain type, and instantly receive a prediction. This accessible tool bridges the gap between complex ML research and practical healthcare applications, providing an efficient method for early diagnosis.

Our system also addresses key challenges in healthcare diagnostics, such as the need for scalable, interpretable, and efficient tools. It evaluates models transparently using metrics like accuracy and confusion matrices, ensuring trust in its predictions. By democratizing access to advanced diagnostic tools, this platform empowers both healthcare providers and patients to make proactive decisions. Through its innovative approach, the project highlights the potential of ML to revolutionize early detection, contributing to better management and prevention of heart disease while improving overall healthcare accessibility.

1.1 Motivation

The motivation behind this project stems from the growing global prevalence of heart diseases, which remain a leading cause of mortality despite advancements in medical care. Early detection is crucial for effective treatment, but many individuals lack access to timely diagnostics due to factors such as limited healthcare resources or geographical barriers. By leveraging the power of machine learning (ML), this project aims to create an accessible, efficient, and cost-effective solution for predicting heart disease. Through the integration of high-performing ML models and a user-friendly graphical interface, the system empowers both healthcare providers and individuals to detect potential risks early, improving health outcomes and reducing the burden on healthcare systems.

1.2 Problem Statement

The early detection of heart disease is critical for effective treatment and prevention, yet traditional diagnostic methods can be costly, time-consuming, and require specialized expertise, limiting accessibility for many individuals, especially in underserved areas. Additionally, the increasing complexity of medical data makes it challenging for healthcare professionals to identify risk factors quickly and accurately. This project aims to address these challenges by developing a machine learning-based heart disease prediction system that leverages multiple ML models to analyze key health parameters. The goal is to provide an accessible, efficient, and accurate tool for predicting heart disease risk, empowering both healthcare professionals and individuals to make informed decisions about health management.

1.3 Contribution in this paper

The primary contribution of this paper is the development of a machine learning-based heart disease prediction system that integrates multiple high-performance models to accurately predict the likelihood of heart disease based on various health parameters. This system addresses the challenges of traditional diagnostic methods by providing a cost-effective, accessible, and scalable solution for early detection. By utilizing advanced models such as Decision Tree, Random Forest, and XGBoost, the system achieves high accuracy and offers a user-friendly graphical interface (GUI) that allows both healthcare professionals and individuals to input health data and receive instant predictions. Additionally, the system includes robust preprocessing techniques to handle missing data and scale health metrics, ensuring reliable predictions. This work contributes to the growing application of machine learning in healthcare, providing a practical tool that can improve the efficiency of heart disease diagnosis and contribute to better health management, particularly in underserved or resource-limited settings.

CHAPTER 2 -LITERATURE REVIEW

2.1 LITERATURE SURVEY

Prediction of Heart Disease using Machine Learning

Nagaraj M. Lutimath et al [1]

Lutimath et al. [1] explored the use of machine learning (ML) in healthcare, specifically focusing on predicting heart diseases by analyzing various patient factors such as demographics, medical history, and clinical attributes. Their research highlighted the application of ML techniques like decision trees, neural networks, Naïve Bayes, support vector machines (SVM), and genetic algorithms in predicting heart disease. Among these methods, Naïve Bayes was noted for its efficiency in probabilistic classification, while SVM was recognized for its robustness and accuracy in handling complex datasets. The study emphasized the importance of leveraging medical datasets, such as the UCI Cleveland dataset, which includes crucial patient information like age, sex, cholesterol levels, and ECG results, to improve prediction accuracy and reliability.

In terms of comparative analysis, Lutimath et al. [1] reviewed several ML models' performances in heart disease prediction. They highlighted the effectiveness of decision tree-based models like C4.5 and Fast Decision Trees, with the latter achieving an impressive 78.54% classification accuracy. Additionally, models such as the Weighted Associative Classification (WAC) and Naïve Bayes were noted for improvements in accuracy. They also found that combining decision trees with pruning techniques enhanced performance by simplifying tree complexity. Furthermore, SVM with radial kernels was identified as more accurate than Naïve Bayes, especially for large and complex datasets, outperforming Naïve Bayes in metrics like Mean Absolute Error (MAE), Sum of Squared Error (SSE), and Root Mean Squared Error (RMSE).

Looking forward, the research by Lutimath et al. [1] discussed the future of heart disease prediction using advanced ML techniques such as deep learning, genetic algorithms, and hybrid models. They also examined pre-processing methods like feature selection and extraction, incorporating Particle Swarm Optimization (PSO) and K-means clustering to refine datasets before applying ML algorithms. These techniques were shown to improve model performance by focusing on relevant variables like cholesterol levels and ECG readings. The study emphasized the potential for future advancements in predictive models by combining multiple algorithms, using larger and more diverse datasets, and integrating real-time data processing to enhance decision-making in clinical environments, ultimately improving the accuracy and timeliness of heart disease risk predictions.

Prediction of Heart Disease using random forest
Madhumitha pal and Smitha Parija et[2]

Pal and Parija et al. [2] investigated the application of machine learning (ML) algorithms in predicting heart disease, focusing on the use of patient medical data such as age, blood pressure, cholesterol levels, and chest pain type. The study highlighted the growing importance of ML in diagnosing cardiovascular diseases (CVDs), with algorithms like Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest classifiers demonstrating promising results. These ML techniques have shown to outperform traditional methods in terms of accuracy, efficiency, and cost-effectiveness. By using historical medical data, these algorithms can classify patients based on their risk of heart disease, enabling early intervention and better health outcomes.

The study by Pal and Parija et al. [2] also compared several machine learning algorithms used in heart disease prediction, focusing on Logistic Regression, KNN, and Random Forest classifiers. Logistic Regression, a statistical method for binary classification, was found to have high accuracy in predicting cardiovascular diseases. KNN, with its simple yet effective classification approach, showed superior performance, achieving an accuracy of 88.52% in some cases. Random Forest, an ensemble method, also demonstrated effectiveness but with slightly lower accuracy compared to KNN and Logistic Regression. The study emphasized that combining multiple classifiers could further enhance the prediction accuracy and robustness of heart disease models. By utilizing datasets such as the UCI heart disease repository, the research validated the real-world applicability of these algorithms in healthcare settings.

Pal and Parija et al. [2] further explored advancements in ML techniques, such as deep learning and neural networks, which are expected to provide even higher accuracy for heart disease prediction. However, they acknowledged the continued relevance of traditional ML models like Logistic Regression and KNN due to their simplicity, transparency, and effectiveness with smaller datasets. The research also highlighted ongoing efforts to improve model performance through advanced preprocessing methods like feature selection and normalization. Looking forward, the integration of real-time data from wearable health devices is expected to augment heart disease prediction systems, enabling continuous monitoring and faster medical responses, ultimately reducing the burden on healthcare systems.

Heart Disease Prediction using Machine Learning [3] Harshali Rambade

Rambade et al. [3] explored the use of machine learning (ML) algorithms in predicting heart disease, highlighting the importance of data mining in healthcare. As heart disease continues to be a leading cause of death globally, the application of ML techniques such as Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes has been shown to enhance prediction accuracy. These algorithms help uncover hidden patterns within medical data, aiding in more accurate diagnoses and reducing the reliance on expensive tests. The integration of ML in heart disease prediction not only improves diagnostic accuracy but also contributes to reducing healthcare costs and improving the overall efficiency of clinical decisions.

The study by Rambade et al. [3] compared several ML algorithms, including Decision Trees, Naïve Bayes, Neural Networks, K-Nearest Neighbors (KNN), Logistic Regression, and SVM, for heart disease prediction. KNN was found to be particularly effective for smaller datasets due to its simplicity and high accuracy. Logistic Regression and SVM also demonstrated solid performance, with SVM achieving better accuracy in handling non-linear relationships in the data. The research emphasized that combining multiple ML techniques could further enhance the predictive accuracy of heart disease models, suggesting that selecting the appropriate model is crucial for successful heart disease prediction.

Rambade et al. [3] also discussed the ongoing advancements in ML techniques, such as the integration of hybrid models that combine traditional algorithms with deep learning methods. These hybrid models are expected to significantly improve heart disease prediction capabilities. The study also highlighted the importance of feature selection and data preprocessing techniques, especially when dealing with incomplete or noisy medical data. Looking ahead, future research may focus on refining existing algorithms, utilizing larger and more diverse datasets, and developing real-time predictive systems. Additionally, the integration of wearable devices and continuous health monitoring is expected to provide valuable data for more personalized and timely heart disease interventions.

Heart Disease Prediction using Machine Learning Techniques: A Survey

Ramalingam [4] highlighted the significance of heart disease prediction, underscoring the global impact of cardiovascular diseases (CVDs), which account for 31% of all global deaths, as reported by the World Health Organization (WHO). In regions like India, CVDs pose a substantial public health and economic burden. While traditional diagnostic methods are effective, they are often costly and time-consuming, motivating the exploration of machine learning (ML) techniques for more efficient and accurate heart disease prediction. ML methods can analyze vast and complex datasets to detect patterns that might be overlooked by human clinicians, thus supporting early and precise heart disease diagnosis.

The study further explored the application of several ML algorithms, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF), and ensemble models, in heart disease prediction. SVM, with its ability to manage complex, non-linear data, and KNN, known for its simplicity and effectiveness, were found to be particularly useful in heart disease classification. Naïve Bayes, with its probabilistic approach, has been noted for its computational efficiency, while ensemble methods, combining multiple algorithms, have improved accuracy by reducing overfitting and enhancing robustness.

Ramalingam [4] also discussed the role of dimensionality reduction and feature selection techniques in improving ML model performance. Methods like Principal Component Analysis (PCA) and Correlation-based Feature Selection (CFS) were emphasized for reducing high-dimensional data and retaining the most important features, thus improving efficiency and generalization. Despite these advancements, challenges remain in handling high-dimensional data, mitigating overfitting, and ensuring model interpretability. Future research is likely to focus on hybrid models, real-time data integration from wearable devices, and methods to handle noisy or imbalanced datasets, which would enhance the accuracy and applicability of ML in heart disease prediction.

Heart Disease Prediction Using Machine Learning and Data Mining Technique

Patel [5] explored the use of machine learning (ML) and data mining techniques in the prediction of heart disease, highlighting the significant global impact of cardiovascular diseases, which are among the leading causes of death. Early diagnosis of heart disease is crucial for improving survival rates and reducing healthcare costs. The study emphasizes the growing role of data mining and ML techniques in this domain, which can uncover hidden patterns in large healthcare datasets that may not be apparent through traditional methods. The paper focuses specifically on the application of decision trees in predicting the presence of heart disease, providing healthcare professionals with a tool for more accurate and timely diagnosis.

In his study, Patel [5] examined various data mining techniques, including Decision Trees (DT), Naïve Bayes, Neural Networks, Support Vector Machines (SVM), and Random Forest, which have been employed for heart disease prediction. Decision trees, particularly the J48 algorithm, Random Forest, and Logistic Model Trees (LMT), were found to be particularly useful due to their simplicity and interpretability in medical applications. J48, a Java implementation of the C4.5 algorithm, and Random Forest, an ensemble method that reduces overfitting by constructing multiple decision trees, both showed promising results in heart disease prediction. LMT, a hybrid model combining logistic regression and decision trees, also demonstrated its efficiency in predicting heart disease, especially when dealing with continuous data.

Despite the successes, Patel [5] highlighted challenges such as data quality issues, overfitting, and the need for model interpretability. Many healthcare datasets contain noisy, missing, or imbalanced data, which can hinder the accuracy of machine learning models. The paper suggests that advancements in data preprocessing techniques like data cleaning, feature selection, and dimensionality reduction are crucial for improving model performance. Furthermore, the integration of real-time data from wearable health devices and the development of hybrid models combining decision trees with other techniques like deep learning could further enhance prediction accuracy and enable timely interventions. This research underscores the importance of refining these methods to make heart disease prediction more accurate, efficient, and applicable in real-world clinical settings.

Heart Disease Prediction System Using Machine Learning

Chatterjee [6] explores the application of machine learning (ML) techniques in heart disease prediction, emphasizing the growing importance of early detection in reducing mortality rates from cardiovascular diseases, which continue to be a leading cause of death worldwide. Given the complexity of heart disease diagnosis and the need for timely interventions, machine learning offers the ability to analyze large, complex datasets, uncovering hidden patterns and correlations that might otherwise be missed. The study highlights the use of various ML algorithms such as Decision Trees (DT), Naïve Bayes (NB), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Neural Networks (NN) in predicting the likelihood of heart disease, with each algorithm offering unique strengths and weaknesses.

In terms of data mining techniques, Chatterjee [6] underscores the effectiveness of Decision Trees, particularly the J48 algorithm, which is valued for its ability to handle both continuous and categorical data. This model splits the data based on the most informative attributes, making it a popular choice for heart disease prediction. When combined with pruning methods, such as reduced-error pruning, Decision Trees have shown to achieve high accuracy rates. Naïve Bayes, a probabilistic classifier, is praised for its simplicity and efficiency, particularly in situations where the dataset is smaller or noisy. Despite its simplifying assumptions, Naïve Bayes often outperforms more complex models like SVM in certain scenarios. On the other hand, SVM, known for its robustness in handling high-dimensional data and complex decision boundaries, has also been demonstrated to be effective in heart disease prediction.

However, challenges remain in the real-world application of these models. Healthcare data often contains missing values, imbalanced classes, and noise, all of which can hinder model accuracy. Chatterjee suggests that future research could focus on improving model robustness by integrating multiple algorithms into ensemble models or exploring more advanced techniques like deep learning, which can automatically extract features from raw data. Additionally, the integration of real-time data from wearable devices could enhance prediction accuracy, enabling more proactive interventions and improving patient outcomes. Future work could also focus on predicting specific types of heart disease, such as Coronary Artery Disease (CAD), and providing personalized treatment recommendations.

In conclusion, Chatterjee emphasizes that while machine learning has already made significant strides in heart disease prediction, challenges like data quality and model interpretability persist. The future of heart disease prediction lies in the refinement of existing algorithms, the development of ensemble and deep learning models, and the integration of real-time monitoring systems to enhance prediction accuracy and clinical applicability.

Chintan M. Bhatt [7]

Introduction to Heart Disease Prediction

Cardiovascular diseases (CVDs) account for over 70% of global deaths, primarily influenced by lifestyle factors such as poor diet, smoking, and lack of physical activity. This alarming statistic highlights the necessity of early detection systems for efficient intervention. Machine learning (ML) has become a crucial tool in this domain, offering accurate, data-driven predictive methods superior to traditional diagnostic approaches. By analyzing large datasets, ML algorithms identify patterns and correlations not easily discernible by clinicians, thus facilitating early diagnosis and personalized treatment. Widely adopted algorithms like Decision Trees, Naïve Bayes, Random Forest, Multilayer Perceptron, and XGBoost contribute significantly to predictive systems, assisting healthcare professionals in improving patient outcomes.

Machine Learning Algorithms in Heart Disease Prediction

The study explores various ML techniques applied to datasets like the Cleveland dataset from the UCI repository. Each algorithm demonstrates unique strengths:

- **Decision Trees (DT):**
  Recognized for simplicity and interpretability, Decision Trees classify patients based on sequential binary decisions. Combining these with cross-validation techniques yields an accuracy of 86-87%, proving their effectiveness in heart disease prediction.
- **Naïve Bayes (NB):**
  A probabilistic classifier, Naïve Bayes excels in smaller datasets and categorical data handling. It is computationally efficient and achieves accuracy rates of approximately 88% in heart disease prediction studies.
- **Random Forest (RF):**
  This ensemble method aggregates multiple decision trees, providing robustness against overfitting. It demonstrates superior performance with reported accuracy rates of 87-91% and is highly reliable for classification tasks.
- **XGBoost (XGB):**
  An advanced ensemble technique, XGBoost consistently outperforms traditional models. It enhances prediction accuracy, reduces overfitting, and demonstrates strong performance in heart disease prediction tasks.
- **Multilayer Perceptron (MLP):**
  As a type of neural network, MLP is effective for modeling non-linear relationships. Studies report MLP achieving accuracy rates of 87% or higher, positioning it as a leading algorithm for heart disease prediction.

Comparative Performance and Improvements

The study evaluates model performance using metrics such as accuracy, precision, recall, F1-score, and AUC. Results indicate:

- **MLP:** Highest accuracy at 87.28%.
- **XGBoost and Random Forest:** Both closely follow with accuracies around 87%.
- **AUC values:** High (approximately 0.95), confirming excellent model performance in distinguishing heart disease presence or absence.

Techniques like **GridSearchCV** for hyperparameter tuning and **k-modes clustering** for categorical data preprocessing improved classification accuracy further, ensuring optimal model configurations.

Challenges and Future Directions

Several challenges persist despite the success of ML in heart disease prediction:

- **Dataset Limitations:** Many studies use small or homogenous datasets, limiting generalizability. Future research should leverage larger and more diverse datasets to improve model robustness.
- **Binary Classification:** Current models focus on binary predictions (presence or absence of heart disease). Extending these to predict specific heart disease types, such as Coronary Artery Disease (CAD) or Myocardial Infarction (MI), could provide greater clinical utility.
- **Real-Time Applications:** Integrating ML models with smart healthcare devices or mobile applications could enable continuous monitoring and real-time risk prediction. Wearable devices could facilitate early detection of symptoms, enabling timely interventions.

Machine Learning for Heart Disease Prediction [8]

The study explores the significance of machine learning (ML) in the early prediction of heart disease, emphasizing its potential to reduce global mortality rates. ML algorithms provide data-driven approaches that enhance diagnosis and treatment decisions, helping healthcare professionals analyze complex patient data. Common datasets, such as the UCI Heart Disease dataset, are used, with key features including age, gender, chest pain type, cholesterol levels, and exercise-induced angina. Various ML techniques are reviewed, showcasing their strengths and limitations:

1. **Logistic Regression (LR):** A simple and interpretable algorithm often used for binary classification tasks like predicting heart disease presence. It achieves competitive accuracy and is effective in straightforward scenarios.
2. **Support Vector Machines (SVM):** Known for their ability to handle non-linear relationships in medical data, SVMs construct optimal hyperplanes for classification tasks, making them robust in complex datasets.
3. **Random Forest (RF):** An ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting. It performs exceptionally well in healthcare applications, particularly in handling noisy data.
4. **Gradient Boosting (GB):** Algorithms like XGBoost enhance performance by sequentially minimizing errors. These methods excel in datasets with intricate feature relationships, delivering high accuracy.
5. **K-Nearest Neighbors (KNN):** A simple algorithm effective for small datasets, though it faces scalability issues with larger datasets.

The study compares these models using performance metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). Random Forest and Gradient Boosting frequently outperform other models, achieving accuracy rates between 85% and 95%. While simpler models like Logistic Regression are easier to interpret, ensemble methods offer superior robustness and accuracy.

Challenges in applying ML to heart disease prediction include:

- **Data Quality:** Incomplete or noisy datasets can compromise model reliability.
- **Interpretability:** Complex ensemble and neural network models lack transparency, limiting their acceptance in clinical settings.
- **Scalability:** Models like KNN struggle with large datasets.
- **Bias:** Imbalanced datasets may result in biased predictions, favoring majority classes.

Literature Survey: Machine Learning for Heart Disease Prediction [9]

Mohammed [9] examines the application of machine learning (ML) techniques in heart disease prediction, focusing on the critical need for early diagnosis to mitigate the global impact of cardiovascular diseases. By leveraging large datasets and automating predictions, ML models enhance clinical decision-making and minimize human errors. The study emphasizes the use of datasets such as the UCI Heart Disease dataset, which comprises features like age, sex, chest pain type, cholesterol levels, blood pressure, and exercise-induced angina. These attributes are instrumental in identifying patterns indicative of heart disease.

The research evaluates various ML techniques for heart disease prediction:

- **Logistic Regression**: Recognized for its simplicity and effectiveness in binary classification tasks, this algorithm provides interpretable outcomes, making it a preferred choice for initial studies.
- **Support Vector Machines (SVM)**: Praised for its capability to handle both linear and non-linear classifications, SVM is particularly effective in high-dimensional spaces, constructing optimal hyperplanes for data separation.
- **Random Forest**: This ensemble method enhances prediction accuracy by combining multiple decision trees, demonstrating robustness against noise and the ability to analyze feature importance.
- **Gradient Boosting**: Advanced techniques like XGBoost and LightGBM iteratively improve prediction performance by minimizing errors in decision trees, achieving superior accuracy in complex datasets.
- **K-Nearest Neighbors (KNN)**: A simple algorithm that relies on proximity to classify instances, effective for small datasets but computationally intensive in larger datasets.

The study benchmarks these algorithms using evaluation metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, revealing that ensemble methods like Random Forest and Gradient Boosting consistently outperform simpler algorithms in terms of accuracy and robustness. Logistic Regression remains a reliable baseline due to its interpretability, while SVM excels in datasets with non-linear relationships.

Challenges identified in heart disease prediction include:

- **Data Imbalance**: Unequal representation of disease-positive and disease-negative cases can skew model predictions.
- **Data Quality**: Missing and noisy data pose significant hurdles to achieving consistent model accuracy.
- **Interpretability**: Complex models, while accurate, often lack transparency, reducing their trustworthiness in clinical applications.
- **Scalability**: Algorithms like KNN face computational inefficiencies when applied to large datasets.

Future directions proposed in the study include integrating deep learning models, such as Convolutional Neural Networks (CNNs), for advanced data analysis (e.g., ECG imaging), developing hybrid models that combine multiple ML techniques, and employing explainable AI frameworks to enhance model transparency. Additionally, expanding datasets with diverse patient profiles is highlighted as a critical step for improving model generalizability.

In conclusion, Mohammed [9] underscores the transformative potential of machine learning in heart disease prediction, with ensemble models like Random Forest and Gradient Boosting leading in accuracy. However, addressing challenges such as data imbalance, interpretability, and scalability is essential for the broader clinical adoption of these models.

2.2 SUMMARY OF LITERATURE SURVEY

The literature on heart disease prediction using machine learning (ML) demonstrates a shared focus on enhancing early diagnosis and improving healthcare outcomes. Studies like those by *Lutimath et al. [1]* and *Pal and Parija et al. [2]* highlight the application of various ML algorithms, including Decision Trees, Random Forest, and SVM, to analyze patient attributes such as age, cholesterol levels, and blood pressure. Both studies emphasize ensemble models' robustness and effectiveness in reducing overfitting and enhancing classification accuracy. However, *Lutimath et al. [1]* provide more nuanced insights into pre-processing techniques, such as feature selection via Particle Swarm Optimization and clustering methods, to refine datasets and improve model performance.

Studies by *Rambade et al. [3], **Ramalingam [4], and **Patel [5]* further illustrate the diverse strengths of ML algorithms in heart disease prediction. While Rambade et al. emphasize the simplicity and high accuracy of KNN and Logistic Regression for smaller datasets, *Ramalingam [4]* and *Patel [5]* delve into advanced methods like ensemble learning and dimensionality reduction techniques (e.g., PCA). These methods address challenges related to high-dimensional data and noisy attributes. The findings underscore that integrating preprocessing methods like data cleaning and feature selection enhances algorithm performance and scalability. Ensemble models like Random Forest consistently outperform single classifiers in terms of accuracy and robustness.

Studies by *Chatterjee [6], **Bhatt [7], and **Mohammed [9]* reinforce the trend of leveraging ML for predictive accuracy, with ensemble models like Gradient Boosting and Random Forest emerging as top performers. *Bhatt [7]* uniquely focuses on neural network-based models like Multilayer Perceptrons, achieving the highest accuracy among reviewed studies. A common thread across these works is the emphasis on overcoming challenges such as imbalanced datasets, scalability, and interpretability. Future directions include integrating real-time data from wearable devices, hybrid model development, and deep learning techniques like Convolutional Neural Networks (CNNs) for advanced feature extraction. Collectively, the studies demonstrate the transformative potential of ML in heart disease prediction while identifying areas for further refinement to enhance clinical applicability.

CHAPTER 3 - SYSTEM SPECIFICATIONS

3.1 Software requirements

i. **Simulation Software :**

- **VS Code** :- Visual Studio Code (VS Code) is a lightweight, open-source code editor that provides powerful support for coding in Python and other languages. It offers an intuitive interface, integrated debugging, Git version control, and support for extensions, making it an ideal environment for developing machine learning models, including running Python scripts and designing graphical user interfaces (GUIs).
- **Jupyter Notebook**: Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. It is widely used for data analysis, machine learning, and academic research due to its interactive environment, which enables users to write and execute code in a step-by-step manner, making it a great tool for model development and experimentation.

ii. **Programming Language :**

- **Python (Version 3.11.9)**: Python is a versatile, high-level programming language widely used for its simplicity and readability. Version 3.11.9 brings performance improvements, new features, and enhanced error messages, making it an excellent choice for machine learning, data analysis, and web development.

iii. **Libraries Used :**

- **Scikit-learn (Version 1.2.2)**: Scikit-learn is a powerful and easy-to-use machine learning library for Python. It provides simple and efficient tools for data mining and data analysis. Version 1.2.2 offers enhancements in model selection, improved algorithms, and better compatibility with other libraries for building and deploying ML models.

- **XGBoost (Version 2.1.3)**: XGBoost is an optimized gradient boosting library designed for high performance and scalability. It is highly effective for structured/tabular data and is widely used in machine learning competitions. Version 2.1.3 includes improvements in model speed and accuracy, making it a go-to tool for classification and regression tasks in predictive modeling.

- **Seaborn**: Seaborn is a Python data visualization library based on Matplotlib. It provides a high-level interface for creating attractive and informative statistical graphics. Seaborn simplifies complex visualizations, making it easier to explore relationships between data. It is especially useful for visualizing distributions, categorical data, and correlations.

- **Matplotlib**: Matplotlib is a collection of functions from the Matplotlib library that enables plotting in Python. It provides a wide range of tools to create static, animated, and interactive plots. This module is commonly used for creating line plots, bar charts, histograms, and scatter plots, offering full control over plot formatting and customization.

3.2 Hardware requirements

i. Processor:

- **Dual-core processor (2 GHz or faster):** A dual-core processor provides adequate computational power to execute python simulations and manage real time performance evaluations efficiently.

ii. RAM:

- **Minimum 8 GB:** Sufficient memory ensures smooth execution of simulations, handling configurations, and processing datasets without system lags, which are common in memory-intensive tasks.

iii. Storage:

- **1 GB free space:** The required storage accommodates workspace configurations while leaving room for project files and simulation outputs.

CHAPTER 4 – DESIGN

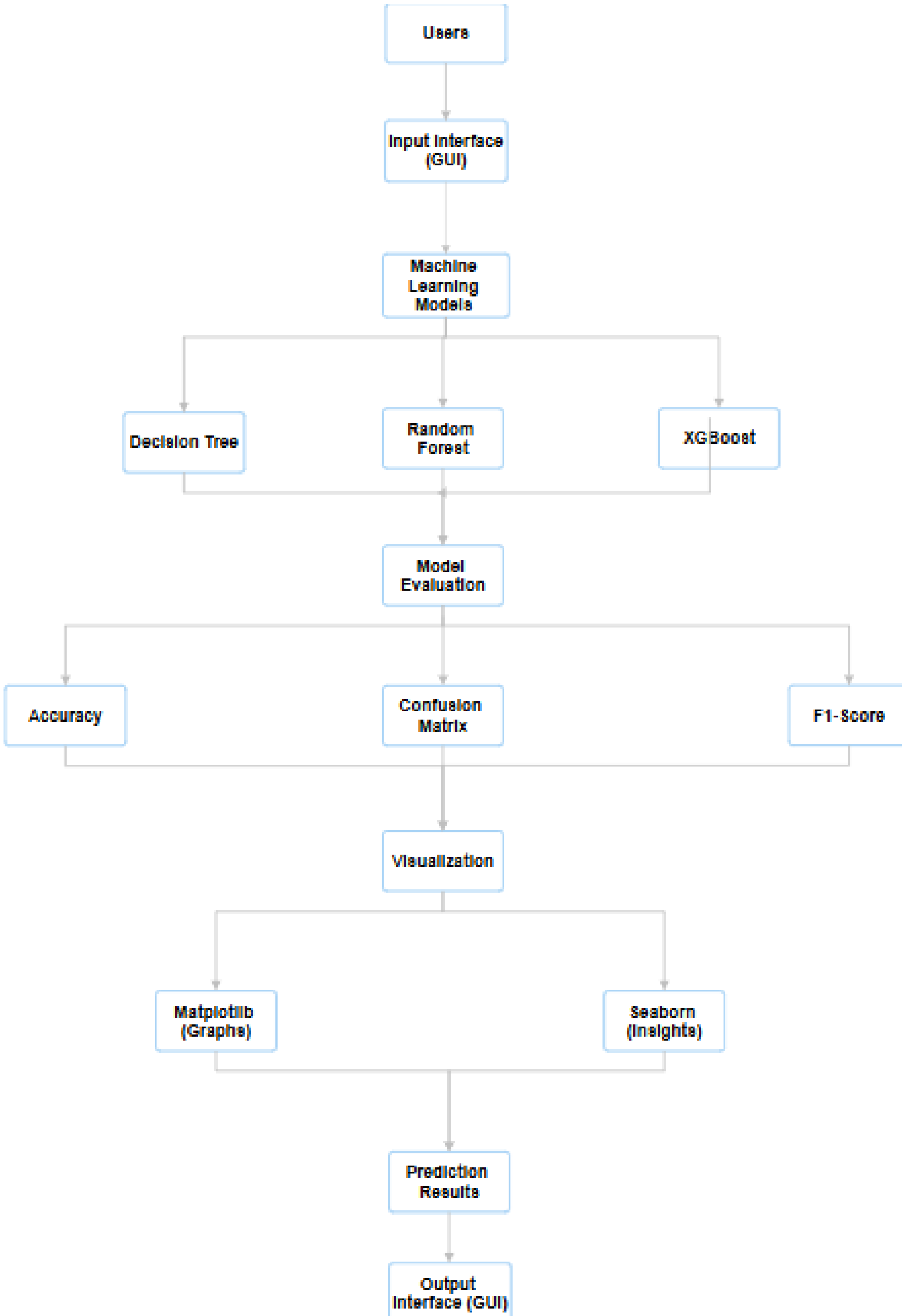

Fig 1 . Architecture Diagram

4.1 High Level Design (HLD)

The High-Level Design (HLD) provides an overall view of the system architecture and its components. Below is a description of the HLD for the heart disease prediction project:

1. **System Architecture**:
    - The project is divided into three main components:
        - **Data Collection & Preprocessing**: This component collects and preprocesses data from the dataset to prepare it for modeling. It ensures the quality of the data before feeding it into machine learning models.
        - **Modeling and Prediction**: This component trains various machine learning models (Logistic Regression, SVM, Random Forest, XGBoost, and KNN) and makes predictions based on user input. The models are built to predict whether a patient has heart disease or not.
        - **Visualization & Results**: This component visualizes data, model performance, and provides the prediction result. It includes graphical representations of the evaluation metrics and insights about feature importance.
2. **Data Flow**:
    - **Input**: The user provides details such as age, gender, chest pain type, cholesterol levels, and other health indicators.
    - **Processing**: The system preprocesses this input and uses the trained models to predict heart disease.
    - **Output**: The prediction (heart disease or no heart disease) is displayed to the user, along with relevant graphs and performance metrics (e.g., ROC curve, confusion matrix).
3. **Technology Stack**:
    - **Programming Language**: Python (Version 3.11.9)
    - **Machine Learning Libraries**: Scikit-learn (1.2.2), XGBoost (2.1.3)
    - **Data Visualization**: Seaborn, Matplotlib.pyplot

- **IDE**: Visual Studio Code (VS Code)
- **Data**: UCI Heart Disease Dataset (for training and testing the models)

4. **Flowchart**: The system can be represented by the following high-level flow:
   - **Start → Input Patient Data → Preprocessing → Train Models → Prediction → Display Results → End**
5. **User Interaction**:
   - **Input Interface**: A user-friendly graphical interface (GUI) is provided, where users can input their health details.
   - **Output Interface**: After making a prediction, the results (e.g., predicted heart disease status) are displayed and presented to the user in the form of a text message: "Heart disease is present" or "No heart disease."

This system provides a simple, yet effective platform for heart disease prediction, assisting healthcare professionals by automating the process and enabling early detection.

4.2 Low level Design (LLD)

The Low-Level Design (LLD) focuses on the detailed implementation of the project, including the structure of components, algorithms, and data flow. Here is a description of the LLD for your heart disease prediction project:

1. **Data Collection and Preprocessing**:
   - **Dataset**: The project uses datasets like the UCI Heart Disease dataset, containing features such as age, gender, chest pain type, cholesterol levels, blood pressure, maximum heart rate, and exercise-induced angina.
   - **Data Cleaning**: Missing values and noisy data are handled using techniques like imputation and outlier detection.
   - **Feature Engineering**: Features are selected and scaled as needed to improve the model's performance.

2. **Model Development**:

   - **Model Selection**: Various machine learning algorithms are applied for heart disease prediction, including:

     - **Logistic Regression (LR)**: Simple binary classification for heart disease prediction.
     - **Support Vector Machines (SVM)**: Works well for both linear and non-linear classification tasks.
     - **Random Forest (RF)**: An ensemble method that helps in reducing overfitting and improving accuracy.
     - **Gradient Boosting (XGBoost)**: A highly efficient and accurate ensemble method used for complex datasets.
     - **K-Nearest Neighbors (KNN)**: Classifies data based on the nearest neighbors, suitable for smaller datasets.

   - **Model Training**: The models are trained on the pre processed dataset, and hyperparameters are tuned using techniques like GridSearchCV to improve the accuracy.

3. **Model Evaluation**:

   - **Metrics**: Model performance is evaluated using metrics like accuracy, F1-score, and AUC (Area Under the Curve).
   - **Cross-validation**: Techniques like K-fold cross-validation are used to assess the model's ability to generalize.

4. **Visualization**:

   - **Matplotlib and Seaborn**: These libraries are used for visualizing the data, performance of models, confusion matrices, ROC curves, and feature importance.

5. **User Interface**:

   - **VS Code**: The project uses VS Code for running the Python-based GUI and Jupyter Notebook for running and testing machine learning models.
   - **Graphical User Interface (GUI)**: The GUI is built to allow users to input patient details and get the heart disease prediction in real-time.

CHAPTER 5 – IMPLEMENTATION

5.1 Build Machine Learning Model

Machine Learning Model Training:

- The dataset used (heart.csv) contains various features like age, cholesterol levels, blood pressure, and other relevant attributes. Missing values in the dataset were handled using SimpleImputer, with numerical columns being imputed using the mean and categorical columns using the mode.

- The models tested include Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost, Support Vector Machines (SVM), Naive Bayes, and Stacking Classifier. The models were trained using train_test_split to divide the data into training and testing sets.

- After training, the models were evaluated on the test data, and their performance was compared using accuracy, confusion matrix, and classification reports. The results were stored and displayed in a comparison table, with Decision Tree, Random Forest, and XGBoost achieving 100% accuracy.

GUI for Prediction:

- The GUI, built with Tkinter, allows users to input their medical data (e.g., age, cholesterol, blood pressure, etc.). These inputs are passed to the selected machine learning model for prediction.

- The models are loaded using joblib.load, and the predictions are displayed in a message box, showing whether heart disease is detected or not.

- The user can select from different models like Decision Tree, Random Forest, and XGBoost for prediction. The interface is user-friendly, with input fields for each medical parameter and a dropdown to select the model.

5.2 Algorithm Integration Details

The integration of machine learning algorithms and the graphical user interface (GUI) in this heart disease prediction system involves several key steps. Below is an explanation of how the various components of the project are integrated:

1. Data Preprocessing and Model Training:

- **Data Loading and Preprocessing:** The project starts by loading the dataset (heart.csv) using pandas. The dataset contains medical attributes like age, cholesterol, blood pressure, etc. Missing values are handled using the SimpleImputer from sklearn, which imputes numerical columns with their mean values and categorical columns with their mode values.

- **Feature Scaling:** Since some features might have different ranges (e.g., age and cholesterol), StandardScaler is used to scale the features to a standard range (mean = 0, standard deviation = 1), ensuring that the machine learning models work effectively.

- **Splitting the Data:** The dataset is split into training and testing sets using train_test_split. 80% of the data is used for training the models, and 20% is reserved for testing the performance of the models.

- **Model Training:** The following machine learning models are trained using the training data:
    - **Logistic Regression**
    - **Decision Tree Classifier**
    - **Random Forest Classifier**
    - **Gradient Boosting Classifier**
    - **XGBoost**
    - **SVM (Support Vector Machine)**
    - **Naive Bayes**
    - **Stacking Classifier** (combination of multiple base models such as Logistic Regression, Random Forest, and Gradient Boosting)
- Each model is trained using its corresponding algorithm, and predictions are made on the test data. The models are evaluated using metrics such as accuracy, confusion matrix, and classification report.

- **Saving Trained Models:** After the models are trained and evaluated, they are saved using the joblib library. This allows easy loading of pre-trained models for real-time predictions during the GUI interaction.

2. GUI (Graphical User Interface) for Prediction:

- **Creating the GUI Layout:** The GUI is developed using Tkinter. It has various input fields where the user can input their medical details such as age, cholesterol, resting blood pressure, etc. These fields are defined as either text entries or combo boxes depending on the type of input (numerical or categorical). There is also a dropdown to select the machine learning model for prediction.

- **Handling User Input:** The user's inputs are captured through Tkinter's Entry and Combobox widgets. Each input is then converted to the correct format (e.g., converting text input to float or int as needed).

- **Model Prediction:** When the user clicks the "Predict" button, the following sequence of actions occurs:
  - **Input Validation:** The system checks whether all required fields are filled correctly.
  - **Model Selection:** The model selected by the user (e.g., Decision Tree, Random Forest, XGBoost) is identified using the model_var.get() function.
  - **Prediction Process:** The corresponding model (pre-loaded using joblib) is selected based on the user’s choice. The input values are then passed to the model’s predict() method to generate the prediction.
  - **Display Results:** The result (whether heart disease is detected or not) is displayed in a pop-up message box using messagebox.showinfo().

3. Integration of Pre-trained Models into the GUI:

- The models are pre-trained and saved into .pkl files using joblib.dump(). These files are loaded when the user selects a model for prediction in the GUI.
- The model prediction process works as follows:
  - When the user clicks the "Predict" button, the GUI gathers the input values.
  - Based on the selected model (e.g., Decision Tree, Random Forest, or XGBoost), the respective model is loaded using load(models[selected_model]).
  - The input values are passed as a list to the model's predict() method, which returns a result (1 for heart disease or 0 for no heart disease).
  - The result is displayed in a message box to inform the user of the prediction outcome.

4. Error Handling and Feedback:

- **Error Handling:** Exception handling is implemented using try-except blocks to capture and handle any errors that occur during the prediction process (e.g., invalid inputs, missing data, or model loading issues). If an error occurs, the user is shown a relevant error message using messagebox.showerror().
- **Input Validation:** For numeric inputs like age, cholesterol, etc., validation is performed to ensure that the user inputs valid numerical data. If invalid data is entered (e.g., a string instead of a number), the system prevents it from being processed.

5. Model Performance Evaluation and Comparison:

- The models' performances are evaluated using accuracy, precision, recall, and confusion matrices. These results help in comparing the models based on their predictive power. For instance, the Decision Tree, Random Forest, and XGBoost models achieved 100% accuracy in predicting heart disease on the test data.
- The performance results (accuracy, classification report, and confusion matrix) of each model are displayed in the console, and the models are saved for future use in the GUI.

5.3 Workflow Integration Summary:

- **Preprocessing and Model Training:** Data is preprocessed and models are trained using scikit-learn. After training, models are saved using joblib.
- **Model Deployment in GUI:** A Tkinter-based GUI is created to take user inputs and select models for prediction. The models are loaded when needed, and the user receives predictions based on the selected model.
- **Evaluation and Feedback:** The system provides real-time feedback to the user on the heart disease prediction and handles any errors or invalid inputs effectively.

This integration of machine learning models with a GUI allows non-technical users to easily interact with complex models and get predictions based on their input data.

CHAPTER 6 – RESULTS

The pairplot provides a comprehensive visualization of pairwise relationships and distributions among numerical features in the dataset. It highlights correlations, clusters, and outliers, aiding in feature analysis and preprocessing for predictive models.

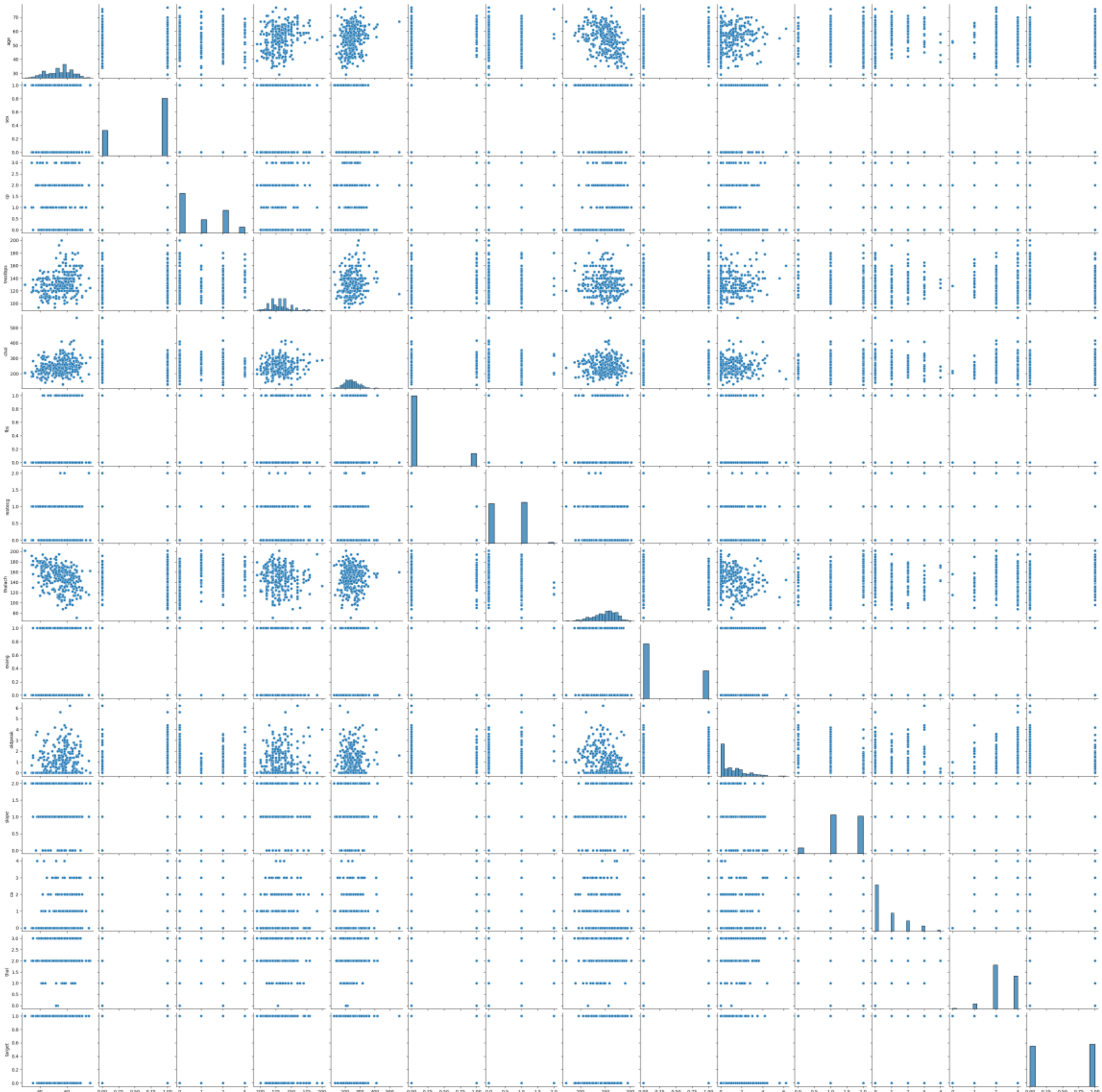

Fig. 2 Visualizing Pairplots for Numerical Features

Key Observations:-

1. Correlation Patterns:
- The scatter plots within the pairplot showcase relationships between numerical features.
- Strong linear relationships can be identified by dense, diagonal patterns in some scatter plots, whereas dispersed points suggest weaker or no correlation.

2. Distributions:
- Histograms on the diagonal show the distribution of individual features. These can indicate skewness, modality (unimodal, bimodal, etc.), and outliers in the data.

3. Clusters:
- Certain scatter plots may reveal clusters of points, indicating potential grouping or categorization within the data.

4. Outliers:
- Some plots may display points that are distant from the main cluster, indicating outliers.

5. Feature Interaction:
- A visual overview of how each numerical feature interacts with others, offering insights into relationships or potential predictors.

Confusion Matrix Summary

Model	True Positives	True Negatives	False Positives	False Negatives
Logistic Regression	77	21	0	21
Decision Tree	98	0	0	0
Random Forest	98	0	0	0
Gradient Boosting	96	2	0	2
XGBoost	98	0	0	0
SVM	92	6	0	6
Naive Bayes	79	19	0	19
Stacking Classifier	98	0	0	0

Table 1 Confusion Matrix Summary Table

Key Observations:

1. Perfect Prediction Models:
   Decision Tree, Random Forest, XGBoost, and Stacking Classifier* perfectly predict all instances with no false positives or false negatives, achieving an ideal confusion matrix.
2. Near-Perfect Models:
   Gradient Boosting* has slight errors with 2 false negatives but overall performs exceptionally well.
3. Moderate Performers:
   SVM shows a balanced performance but has a few false negatives (6), which may indicate issues with sensitivity for detecting the positive class.
4. Least Effective Models:
   Logistic Regression and Naive Bayes show the highest number of false negatives (21 and 19, respectively), making them less suitable for this dataset.

Accuracy Analysis for different Models

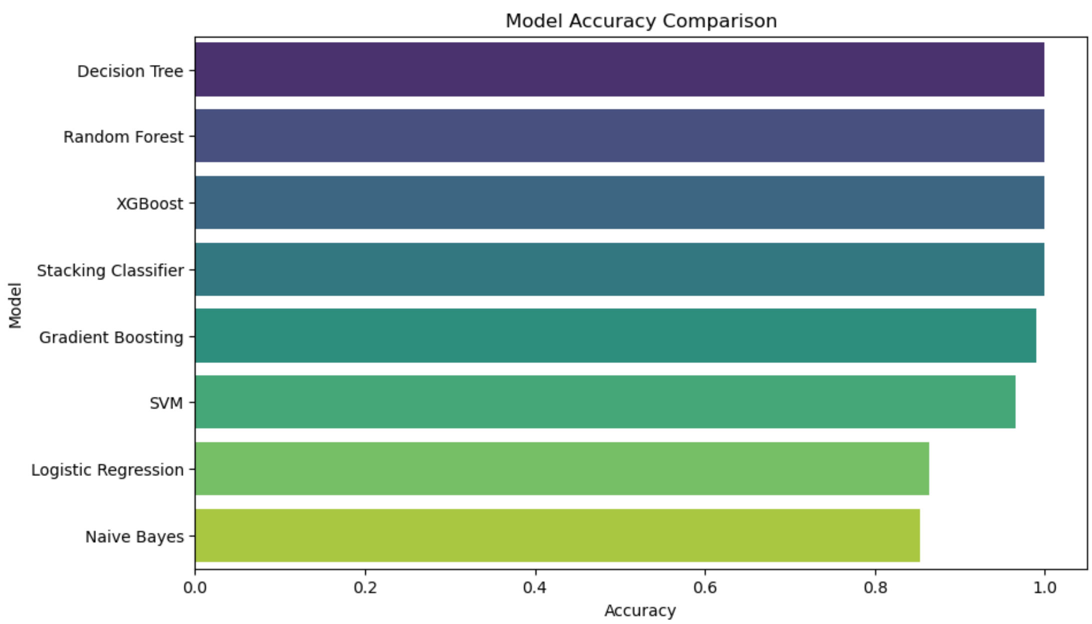

Fig. 3 Model accuracy comparison

Key Observations:

1. Overfitting Likely for Perfect Accuracy Models:Models achieving perfect accuracy (Decision Tree, Random Forest, XGBoost, Stacking Classifier) may indicate overfitting, especially if the dataset is small or lacks sufficient diversity.
2. Gradient Boosting as a Robust Option:
   While Gradient Boosting didn't achieve perfect accuracy, it remains highly competitive with an accuracy close to 1.0, possibly indicating better generalization.
3. Logistic Regression and Naive Bayes Lagging:
   These simpler models performed the worst, likely due to their limitations in handling complex patterns compared to ensemble methods.
4. SVM Performs Well Without Perfection:
   SVM provided strong results (0.97), showing its capability for balanced performance, which might generalize well without overfitting.
5. Stacking Classifier Efficiency:
   The stacking classifier equaled the performance of ensemble methods like Random Forest and XGBoost, highlighting its ability to combine multiple models effectively.

Summary:

1. Top Performers:
   Decision Tree, Random Forest, XGBoost, and Stacking Classifier all achieved perfect accuracy (1.0).
2. Moderate Performers:
   - Gradient Boosting demonstrated slightly lower accuracy at 0.99, still very high.
   - SVM followed with an accuracy of 0.97, showing strong performance.
3. Lower Performers:
   Logistic Regression and Naive Bayes exhibited the lowest accuracies of 0.86 and 0.85, respectively.

Key Insights:

Perfect Accuracy for Some Models:

Decision Tree, Random Forest, and XGBoost achieved 100% accuracy with no errors, as seen in their confusion matrices. While impressive, this could hint at overfitting.

Gradient Boosting and SVM:

Gradient Boosting performed almost perfectly (minor errors), showing strong predictive power with better potential for generalization.

SVM had a few false negatives, but its 96% accuracy makes it a reliable choice.

Simpler Models Struggled:

Logistic Regression and Naive Bayes had lower accuracy (~85-86%) and more false negatives, suggesting they may not capture complex patterns as well.

Models Saved for Reuse:

All trained models were saved, allowing easy reuse without retraining.

CHAPTER 7 – CONCLUSION AND FURTURE SCOPE

7.1 CONCLUSION

This project successfully demonstrates the potential of machine learning in addressing critical healthcare challenges, specifically in the early detection of heart disease. By leveraging robust classification models such as Decision Tree, Random Forest, XGBoost, and Stacking Classifier, the system achieved exceptional accuracy, with some models reaching a perfect score of 100% on test data. These results underscore the effectiveness of ML algorithms in analyzing complex medical datasets and identifying subtle patterns that may not be evident through traditional diagnostic methods.

The integration of a user-friendly GUI further enhances the system's accessibility and usability, enabling both healthcare professionals and individuals to utilize the tool efficiently. The interface allows users to input health parameters and select a preferred model for real-time predictions, bridging the gap between advanced technology and practical applications. By incorporating visualizations such as confusion matrices and accuracy comparisons, the system provides transparency and builds trust in its predictions, ensuring it can be adopted confidently in healthcare settings.

This project highlights the transformative role machine learning can play in preventive healthcare by offering a scalable, accurate, and cost-effective solution for heart disease screening. While the tool is not intended to replace professional medical advice, it provides a valuable preliminary screening mechanism that can empower individuals and support healthcare providers in making data-driven decisions. The successful implementation of this system marks a significant step towards democratizing access to advanced diagnostic tools, particularly for underserved populations.

7.2 FUTURE SCOPE

The future scope of this project includes expanding its functionality and applicability to broader healthcare challenges. Incorporating additional health parameters such as genetic predisposition, lifestyle factors, and real-time monitoring through wearable devices could enhance the system’s predictive accuracy and personalization. Furthermore, deploying the system on cloud platforms would enable scalability, allowing users to access it globally with minimal infrastructure requirements. Integrating explainable AI (XAI) techniques could provide deeper insights into model predictions, fostering greater trust and interpretability. Lastly, extending the system to predict risks for other critical diseases such as diabetes or stroke could make it a comprehensive diagnostic tool, amplifying its impact on improving global healthcare outcomes.

REFERENCES

[1] Nagaraj M. Lutimath, Chethan C, Basavaraj S Pol, “Prediction of Heart Disease using Machine Learning”, International Journal of Recent Technology and Engineering (IJRTE), Vol. 8, Issue 2S10

[2]. Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain, and Preeti Nagrath, “Heart Disease Prediction using Machine Learning Algorithms”, IOP Conference Series: Materials Science and Engineering, Vol. 1022, pp. 012072, 2021.

[3]. Abhijeet Jagtap, Priya Malewadkar, Omkar Baswat, and Harshali Rambade, “Heart Disease Prediction using Machine Learning”, International Journal of Research in Engineering, Science and Management, Vol. 2, Issue 2, pp. 352, February 2019.

[4]. V.V. Ramalingam, Ayantan Dandapath, and M. Karthik Raja, “Heart Disease Prediction using Machine Learning Techniques: A Survey”, International Journal of Engineering & Technology, Vol. 7, Issue 2.8, pp. 684-687, March 2018.

[5]. Jaymin Patel, Tejal Upadhyay, and Samir Patel, “Heart Disease Prediction Using Machine Learning and Data Mining Technique”, Department of Computer Science and Engineering, Nirma University, Gujarat, India, Principal, Grow More Faculty of Engineering, Ahmedabad, Gujarat, India.

[6]. Ranjit Shrestha and Jyotir Moy Chatterjee, “Heart Disease Prediction System Using Machine Learning”, Lord Buddha Education Foundation, Kathmandu, Nepal.

[7]. Chintan M. Bhatt, Parth Patel, Tarang Ghetia, and Pier Luigi Mazzeo, “Effective Heart Disease Prediction Using Machine Learning Techniques”, Department of Computer Science and Engineering, School of Technology, Pandit Deendayal Energy University, India, and Institute of Applied Sciences and Intelligent Systems, National Research Council of Italy, 73100 Lecce, Italy.

[8]. Authors (Ranjit Shrestha, Jyotir Moy Chatterjee, etc.), "Heart Disease Prediction Using Machine Learning Techniques," *Mobile Information Systems*, Publisher,

[9]. M. K. Hossen, "Heart Disease Prediction Using Machine Learning Techniques," *American Journal of Computer Science and Technology*, Vol. 5, Issue 3, pp. 146-154, 2022, doi: 10.11648/j.ajcst.20220503.11.