PREDICTION OF HEART ATTACKS USING MACHINE LEARNING

Minor project-II report submitted
in partial fulfillment of the requirement for award of the degree of

Bachelor of Technology
in
Computer Science & Engineering

By

Y.AKHILA SIRISHA	(21UECS0693)	**(19564)**
Y.JAYASRI	(21UECS0689)	**(19584)**
M.GAYATHRI GREESHMA	(21UECS0389)	**(19545)**

Under the guidance of
Mr.D.MARICHAMY.,M.E.,
ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF COMPUTING

VEL TECH RANGARAJAN Dr. SAGUNTHALA R&D INSTITUTE OF SCIENCE & TECHNOLOGY

(Deemed to be University Estd u/s 3 of UGC Act, 1956)
Accredited by NAAC with A++ Grade
CHENNAI 600 062, TAMILNADU, INDIA

May, 2024

PREDICTION OF HEART ATTACKS USING MACHINE LEARNING

Minor project-II report submitted
in partial fulfillment of the requirement for award of the degree of

Bachelor of Technology
in
Computer Science & Engineering

By

Y.AKHILA SIRISHA	(21UECS0693)	**(19564)**
Y.JAYASRI	(21UECS0689)	**(19584)**
M.GAYATHRI GREESHMA	(21UECS0389)	**(19545)**

Under the guidance of
Mr.D.MARICHAMY.,M.E.,
ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF COMPUTING

VEL TECH RANGARAJAN Dr. SAGUNTHALA R&D INSTITUTE OF SCIENCE & TECHNOLOGY

(Deemed to be University Estd u/s 3 of UGC Act, 1956)
Accredited by NAAC with A++ Grade
CHENNAI 600 062, TAMILNADU, INDIA

May, 2024

CERTIFICATE

It is certified that the work contained in the project report titled ”PREDICTION OF HEART ATTACKS USING MACHINE LEARNING” by ” Y.AKHILA SIRISHA (21UECS0693) (19564),Y.JAY ASRI (21UECS0689) (19584),M.GAYATHRI GREESHMA (21UECS0389) (19545)” has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

Signature of Supervisor
Computer Science & Engineering
School of Computing
Vel Tech Rangarajan Dr. Sagunthala R&D
Institute of Science & Technology
May, 2024

Signature of Professor In-charge
Computer Science & Engineering
School of Computing
Vel Tech Rangarajan Dr. Sagunthala R&D
Institute of Science & Technology
May, 2024

DECLARATION

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

(Signature)
(Y.AKHILA SIRISHA)
Date: / /

(Signature)
(Y.JAYASRI)
Date: / /

(Signature)
(M.GAYATHRI GREESHMA)
Date: / /

APPROVAL SHEET

This project report entitled " PREDICTION OF HEART ATTACKS USING MACHINE LEARNING" by Y.AKHILA SIRISHA (21UECS0693),Y.JAYASRI (21UECS0689),M.GAYATHRI GREE SHMA (21UECS0389) is approved for the degree of B.Tech in Computer Science & Engineering.

Examiners **Supervisor**

Mr.D.MARICHAMY.,M.E.,
ASSISTANT PROFESSOR.

Date: / /

Place:

ACKNOWLEDGEMENT

We express our deepest gratitude to our respected **Founder Chancellor and President Col. Prof. Dr. R. RANGARAJAN B.E. (EEE), B.E. (MECH), M.S (AUTO),D.Sc., Foundress President Dr. R. SAGUNTHALA RANGARAJAN M.B.B.S.** Chairperson Managing Trustee and Vice President.

We are very much grateful to our beloved **Vice Chancellor Prof. S. SALIVAHANAN,** for providing us with an environment to complete our project successfully.

We record indebtedness to our **Professor & Dean, Department of Computer Science & Engineering, School of Computing, Dr. V. SRINIVASA RAO, M.Tech., Ph.D.,** for immense care and encouragement towards us throughout the course of this project.

We are thankful to our **Head, Department of Computer Science & Engineering, Dr.M.S. MURALI DHAR, M.E., Ph.D.,** for providing immense support in all our endeavors.

We also take this opportunity to express a deep sense of gratitude to our **Internal Supervisor Mr.D.MARICHAMY.,M.E.,ASSISTANT PROFESSOR.,** for his cordial support, valuable information and guidance,she helped us in completing this project through various stages.

A special thanks to our **Project Coordinators Mr. V. ASHOK KUMAR, M.Tech., Ms. U.HEMAVATHI, M.E., Ms. C. SHYAMALA KUMARI, M.E.,** for their valuable guidance and support throughout the course of the project.

We thank our department faculty, supporting staff and friends for their help and guidance to complete this project.

Y.AKHILA SIRISHA	**(21UECS0693)**
JAYASRI	**(21UECS0689)**
M.GAYATHRI GREESHMA	**(21UECS0389)**

ABSTRACT

Heart attack prediction is one of the serious causes of morbidity in the world's population. The clinical data analysis includes a very crucial disease i.e., cardiovascular disease as one of the most important sections for the prediction. Data Science and machine learning (ML) can be very helpful in the prediction of heart attacks in which different risk factors like high blood pressure, high cholesterol, abnormal pulse rate,diabetes, etc... can be considered. The objective of this study is to optimize the prediction of heart disease using ML.In this paper, we are presenting a machine learning-based heart attack prediction (ML-HAP) method in which the analysis of different risk factors and prediction for heart attacks is done using ML approaches of Support Vector Machines, Logistic Regression, Na¨ıve Bayes and XG-Boost. The data of heart disease symptoms has been collected the UCI ML Repository and analysis has been performed on the data using ML methods. The focus has been on optimizing the prediction on the basis of different parameters. XGBoost provided the best prediction among the four. The Area under thecurve achieved with XGBoost is .94 and Logistic Regression is .92. The prediction with ML models in identifying heart attack symptoms is highly efficient, especially with boosting algorithms. The prediction was done to evaluate accuracy, precision, recall, and area under the curve. ML models are being trained to perform optimized predictions.

Keywords:

Machine learning,Linear Regression,Decision Tree,Respiratory system,XGBoost, Cardiovascular system, Predictive modeling.

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

ACC American College of Cardiology
AHA : American Heart Association
AI : Artificial Intelligence
API : Application Programming Interface
BP : Blood Pressure
BMI : Body Mass Index
CAD : Coronary Artery Disease
CCS : Canadian Cardiovascular Society
CNN : Convolutional Neural Network
CRP : C-Reactive Protein
CT : Computed Tomography
CV : Cardiovascular
ECG : Electrocardiogram
EHR : Electronic Health Record
HR : Heart Rate
ML : Machine Learning
MRI : Magnetic Resonance Imaging
NLP : Natural Language Processing
PCA : Principal Component Analysis
PET : Positron Emission Tomography
RF : Random Forest
ROC : Receiver Operating Characteristic
SVM : Support Vector Machine

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Introduction

Heart attack which is analogous to acute myocardial infarction (AMI) is one of the most serious diseases in the segment of cardiovascular disease. It occurs due to the interruption of blood circulation to muscle of the heart which damages the heart the muscle. Diagnosing heart disease is also a crucial task. The symptoms, physical examination, and understanding of the different signs of this disease are required to diagnose heart disease. Different factors including cholesterol, genetic heart disease, high blood pressure, low physical activity, obesity, and smoking can be reasons for the occurrence of heart disease. The major reason for heart attacks is the stoppage of blood to the coronary arteries. The red blood cells (RBC) start getting low when blood flow is reduced; due to this the human body stops getting necessary oxygen and loses consciousness. The early diagnosis through symptoms and signs can help prevent patients of heart attacks if the prediction is accurate enough. Figure 1 shows different symptoms of a heart attack. The work presented takes 13 features/attributes as input having number values.

.

It has been stated that little modifications in lifestyle including quitting smoking/alcohol/tobacco, having healthy food habits, and routine exercises can help in the prevention of heart attacks. Any person living a healthy lifestyle with early treatment after diagnosis can greatly increase the positive results. However, it is difficult to identify the high risk of heart disease where different risks like diabetes, high blood pressure, and cholesterol problems are present. In these types of scenarios, ML can help in the early diagnosis of disease In recent years, machine learning (ML) techniques have emerged as promising tools for predicting heart attacks by analysing large datasets of patient information. By ML models can identify patterns and risk factors that may not be evident through traditional methods

1.2 Aim of the Project

The main aim of predicting heart attacks using machine learning is to identify individuals at risk of experiencing a heart attack before it occurs. By analysing various patient data such as medical history, lifestyle factors, and clinical measurements, machine learning models can detect patterns and risk factors that may not be apparent through traditional methods. And targeted medical interventions to prevent or mitigate occurrence it occurs. By analysing various patient data such as medical history, lifestyle factors, and clinical measurements, machine learning models can detect patterns and risk factors that may not be Apparent through traditional methods of heart attacks, ultimately improving of cardiovascular disease.

1.3 Project Domain

The project focuses on certainly this project revolves around using machine learning to predict the likeli- hood of individuals experiencing heart attacks. It involves analyzing diverse medical data encompassing factors like high blood pressure, cholesterol levels, pulse rate ab- normalities, and diabetes. By employing advanced algorithms from data science, the project seeks to construct predictive models that can effectively identify individuals at risk of heart attacks. These models will be trained on extensive datasets, enabling them to discern intricate patterns and correlations within the data. Ultimately, the aim is to furnish healthcare practitioners with a valuable tool for early detection and proactive management of cardiovascular diseases, thus enhancing patient care and mitigating the impact of such ailments on healthcare systems.

1.4 Scope of the Project

The scope of the project involves applying machine learning techniques to predict the occurrence of heart attacks based on various risk factors and symptoms. This encompasses tasks such as data collection from sources like the UCI ML Repository, data preprocessing to clean and prepare the data for analysis, feature selection to identify relevant predictors, model training using algorithms like Support Vector

Machines, Logistic Regression, Naıve Bayes, and XGBoost, model evaluation to assess performance metrics such as accuracy, precision, recall, and area under the curve, and potentially model deployment for real-world use in healthcare settings.

Chapter 2

LITERATURE REVIEW

[1]Krittanawong, C., et al (2017). Proposed Artificial intelligence in precision cardiovascular medicine. Journal of the American College of Cardiology, 69(21), 2657-2664. This review paper discusses the application of artificial intelligence, including machine learning, in precision cardiovascular medicine. It covers various aspects of cardiovascular diseases, including risk prediction, diagnosis, treatment selection, and prognosis, using machine learning algorithms..

[2]Attia, Z. I., et al (2019).proposed Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nature Medicine, 25(1), 70-74. This study demonstrates the use of a convolutional neural network (CNN) to analyze standard 12-lead electrocardiograms (ECGs) for the detection of left ventricular dysfunction, a significant risk factor for heart failure and cardiac events..

[3]Chicco, D.,et al (2017).proposed Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision Making, 17(1), 1-13. This research paper explores the use of machine learning algorithms to predict the survival of patients with heart failure based solely on serum creatinine levels and left ventricular ejection fraction (LVEF) measured from echocardiography. The study demonstrates the potential of machine learning in risk stratification for heart failure patients.

[4]Shah, S. J.,et al (2015).proposed Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation, 131(3), 269-279. This study introduces a data-driven approach called ”phenomapping” using unsupervised machine learning techniques to classify heart failure with preserved ejection fraction (HFpEF) into distinct subtypes based on clinical and phenotypic characteristics. This approach aids in identifying more homogeneous patient populations for targeted therapies and improves prognostication.

[5]Poplin, R.,et al (2018).proposed Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering, 2(3),

158-164. This study demonstrates the use of deep learning algorithms to predict cardiovascular risk factors, including age, gender, smoking status, blood pressure, and major adverse cardiovascular events (MACE), from retinal fundus photographs. The findings suggest that deep learning models trained on retinal images can serve as a non-invasive tool for cardiovascular risk assessment.

[6]Weng, S. F., et al(2017).proposed Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One, 12(4), e0174944. This study investigates the potential of machine learning algorithms to enhance cardiovascular risk prediction using routine clinical data and compares their performance with traditional risk assessment tools like the Framingham Risk Score.

[7]Goldstein, B. A., et al(2017).proposed Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. European Heart Journal, 38(23), 1805-1814. This review article explores the limitations of traditional regression-based methods in cardiovascular risk prediction and discusses the application of machine learning techniques to overcome these challenges.

[8Kramer, A. A., et al(2017).proposed Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Critical Care Medicine, 45(2), 205-210. This paper discusses the Hosmer-Lemeshow test, a popular method for assessing the calibration of predictive models, including those used in cardiovascular risk prediction.

[9]Lam, C. X.,et al (2018).proposed Heart failure prediction with deep learning: A retrospective study. JMIR Medical Informatics, 6(2), e41. This retrospective study investigates the use of deep learning techniques for heart failure prediction using electronic health record data, demonstrating the potential of these methods for early detection and intervention.

[10]Bello, G. A.,et al(2019).proposed Deep learning cardiac motion analysis for human survival prediction. Nature Machine Intelligence, 1(2), 95-104. This research paper presents a deep learning-based approach for analyzing cardiac motion patterns from cardiac MRI data to predict patient survival outcomes, highlighting the importance of advanced imaging techniques in risk stratification

[11]Gupta, D.,et al(2020).proposed Predicting the risk of heart disease using ma-

chine learning algorithms: A systematic review. SN Comprehensive Clinical Medicine, 2, 2892-2905. This systematic review summarizes the existing literature on using various machine learning algorithms for predicting the risk of heart disease, providing insights into the strengths and limitations of different approaches.

[12]Acharya, U. R.,et al(2017).proposed Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Information Sciences, 415, 190-198. This research paper presents a deep convolutional neural network (CNN) for automated detection of myocardial infarction from electrocardiogram (ECG) signals, demonstrating the potential of deep learning in cardiac diagnostics.

Chapter 3

PROJECT DESCRIPTION

3.1 Existing System

The existing system for predicting heart attacks using machine learning typically involves the utilization of various algorithms and methodologies to analyze medical data and identify patterns associated with cardiovascular risk factors. Researchers and practitioners often gather datasets containing information such as patient demo- graphics, medical history, lifestyle factors, and diagnostic test results. They then preprocess the data to handle missing values, normalize features, and remove noise. Next, machine learning models are trained on these preprocessed datasets using algorithms like logistic regression, support vector machines, decision trees, random forests, and deep learning approaches. These models learn from the data to recognize complex relationships between different variables and make predictions about the likelihood of an individual experiencing a heart attack within a certain timeframe. Once trained, the models are evaluated using metrics such as accuracy, precision, recall, and area under the curve to assess their performance. Researchers often com- pare the performance of different models and select the one that provides the most accurate and reliable predictions.

The existing system also involves ongoing research and development efforts to improve the accuracy and efficiency of predictive models.This includes exploring new algorithms, feature engineering techniques, and data sources, as well as validating the models in real-world clinical settings to ensure their effectiveness in practical applications.Additionally, efforts are made to address challenges such as data privacy concerns, model interpretability, and scalability to enable widespread adoption of machine learning-based heart attack prediction systems in healthcare practice.

Disadvantages of existing system are

1. Prediction of cardiovascular disease results is not accurate
2. Data mining techniques does not help to provide effective decision making.

3.Over-reliance on ML predictions may lead to neglect of traditional risk factors.
4.ML models might not always generalize well to diverse populations, leading to biased predictions.
5.Privacy concerns regarding the sensitive health data required for accurate predictions. 6.False positives could cause unnecessary anxiety and medical interventions.
7.ML models may lack transparency, making it difficult to understand how predictions are made.
8.Continuous updates and recalibration of models are necessary to maintain accuracy over time.
9.ML models may struggle with interpreting complex and subtle patterns in heart health data.
10.Implementation costs for integrating ML into healthcare systems can be significant.
11.Lack of interpretability could lead to distrust among healthcare professionals and patient.

3.2 Proposed System

The proposed While the integration of machine learning (ML) into the prediction of heart attacks presents numerous advantages, it also comes with inherent disadvantages that warrant consideration. Firstly, reliance on ML algorithms may lead to a reduction in the importance placed on traditional risk factors. Overemphasis on algorithmic predictions could overshadow the significance of well-established factors such as smoking, high blood pressure, and family history, potentially resulting in missed opportunities for preventive interventions. Moreover, ML models may suffer from biases and inaccuracies, particularly when applied to diverse populations. If the training data used to develop these models are not representative of the entire population, predictions may be skewed, leading to misdiagnoses and inappropriate treatment plans for certain demographic groups.

The privacy and security concerns associated with the use of sensitive health data pose significant challenges. ML algorithms rely heavily on large datasets, including personal health information, to make accurate predictions. However, the collection and storage of such data raise ethical dilemmas regarding patient privacy and confidentiality. Unauthorized access, data breaches, and misuse of health information could result in serious consequences for individuals and healthcare organizations alike. Striking a balance between leveraging ML for predictive analytics and safeguarding patient privacy remains a critical challenge in the development and implementation of predictive models for heart attack prevention.

Advantages of proposed system are

1.Early detection of potential heart attacks, allowing for timely interventions and preventive measures.

2.Personalized risk assessment based on individual health data, leading to tailored treatment plans.

3.Enhanced accuracy compared to traditional risk assessment methods, improving patient outcomes.

4.Identification of subtle patterns and correlations in large datasets that may not be detectable by human experts.

5.Integration of ML algorithms into existing healthcare systems for real-time monitoring and prediction.

6.Potential for cost savings by reducing hospitalizations and emergency interventions through proactive management of heart health.

7.Facilitation of population-level risk stratification, enabling targeted public health interventions and resource allocation.

8.Continuous learning and refinement of prediction models based on new data and insights, improving their efficacy over time.

9.Empowerment of healthcare providers with decision support tools to assist in clinical decision-making and patient management.

3.3 Feasibility Study

A feasibility study for predicting heart attacks using machine learning (ML) would assess the practicality and viability of integrating ML algorithms into existing healthcare systems. This study would involve evaluating various factors such as data availability, quality, and privacy concerns, as well as the technical expertise and resources required for model development and deployment. Additionally, it would analyze the potential benefits of ML-based prediction, including early detection, personalized risk assessment, and cost savings, against the associated implementation challenges and limitations. By conducting a comprehensive feasibility study, healthcare stakeholders can make informed decisions regarding the adoption of ML for heart attack prediction, ensuring alignment with organizational goals, regulatory requirements, and patient care objectives.

3.3.1 Economic Feasibility

The economic feasibility of predicting heart attacks using machine learning (ML) involves assessing the cost-effectiveness of implementing ML-based prediction models within healthcare systems. This evaluation includes considerations such as the initial investment required for model development, infrastructure upgrades, and staff training, balanced against the potential long-term savings from reduced hospitalizations, emergency interventions, and improved patient outcomes. Additionally, economic feasibility analysis would examine the scalability of ML solutions, their impact on healthcare resource utilization, and the potential for generating revenue through improved patient care and population health management. By conducting a thorough economic feasibility study, healthcare organizations can determine the financial viability of integrating ML into heart attack prediction strategies, ensuring efficient allocation of resources and sustainable implementation in the long run.

Furthermore, economic feasibility entails evaluating the return on investment (ROI) associated with ML-based prediction of heart attacks, considering both direct and indirect benefits. Direct benefits include reduced healthcare costs, such as fewer hospital admissions and emergency room visits, as well as savings from preventive interventions and timely treatments. Indirect benefits may encompass improvements in patient satisfaction, quality of life, and overall population health, leading to potential long-term economic gains for healthcare systems and society as a whole. By quantifying the economic impact of ML-driven prediction models, stakeholders can make informed decisions about resource allocation, funding priorities, and strategic investments to maximize the value of predictive analytics in cardiovascular care.

3.3.2 Technical Feasibility

The technical feasibility of predicting heart attacks using machine learning (ML) involves assessing the capability of ML algorithms to effectively analyze and interpret diverse healthcare data sources. This includes evaluating the availability and quality of electronic health records, medical imaging data, genetic information, and other relevant clinical variables required for accurate prediction. Additionally, technical feasibility entails considering the scalability and computational requirements of ML models, ensuring that they can handle large volumes of data and operate efficiently within existing healthcare IT infrastructures. Furthermore, integration with electronic health record systems, interoperability with other healthcare applications, and compliance with data privacy regulations are essential aspects to be addressed to ensure seamless implementation and adoption of ML-based prediction methods in clinical practice.

Moreover, technical feasibility also involves assessing the readiness of healthcare professionals and IT staff to implement and support ML-driven prediction systems. This includes evaluating the availability of skilled personnel with expertise in machine learning, data science, and healthcare informatics who can develop, deploy, and maintain predictive models. Furthermore, considerations such as system compatibility, data standardization, and algorithmic transparency need to be addressed to ensure seamless integration with existing clinical workflows and decision-making processes. By conducting a comprehensive technical feasibility analysis, healthcare organizations can identify potential challenges, develop mitigation strategies, and optimize the implementation of ML-based prediction systems for enhancing cardiovascular risk assessment and improving patient outcomes.

3.3.3 Social Feasibility

The social feasibility of predicting heart attacks using machine learning (ML) involves assessing the acceptance, trust, and ethical implications of implementing ML-based prediction models within healthcare systems and communities. This includes considering factors such as patient and healthcare provider attitudes towards AI-driven healthcare interventions, potential concerns about algorithmic bias and discrimination, and the impact on patient-doctor relationships and communication. Additionally, social feasibility analysis encompasses addressing cultural and socio-economic disparities in access to healthcare and technology, ensuring equitable distribution of ML-based predictive tools and interventions. By fostering transparency, inclusivity, and collaboration among stakeholders, healthcare organizations can navigate social complexities and leverage ML technology responsibly to improve cardiovascular risk assessment and promote public health outcomes.

Furthermore, the project promotes environmental awareness and education by fos- tering a deeper understanding of the sources and patterns of air pollution. Through the dissemination of air quality data, stakeholders, including policymakers, environ- mental agencies, and the general public, gain valuable insights into the impact of human activities on the environment. Additionally, the project's reliance on machine learning underscores the potential for technological innovation to address societal challenges, contributing to a positive perception of technology as a force for en- vironmental stewardship. In conclusion, the social feasibility of the "Air Quality Analysis Using Machine Learning" project is significant, as it aligns with societal values of health, safety, environmental awareness, and technological innovation for the greater good.

3.4 System Specification

- System: Intel Dual Core
- Monitor: 15" LED or above
- Ram: 1GB or above

3.4.1 Hardware Specification

- System: Intel Dual Core or above
- Hard Disk: 120GB or above
- Monitor:15" LED or above
- Input devices: Keyboard and Mouse
- Ram: 1GB or above

3.4.2 Software Specification

1. Programming Language:

• Python: Python is widely used in the machine learning community and offers a rich ecosystem of libraries and frameworks such as scikit-learn, TensorFlow, and Keras.

2.Machine Learning Frameworks:

• scikit-learn: For implementing machine learning models like Linear Regression and Decision Trees.
• TensorFlow or PyTorch: For more complex machine learning models, including neural networks.

3. Data Processing:

• Pandas: For data manipulation and preprocessing.
• NumPy: For numerical operations on data arrays.

4. Visualization:

• Matplotlib and Seaborn: For creating visualizations and plots.
• Plotly: Interactive and web-based visualization library.

5. Development Environment:

• Google colab or Integrated Development Environment (IDE) like VSCode or PyCharm.

6. Testing Framework:

• Pytest: For unit testing of code component.

3.4.3 Standards and Policies

Google Colab:

Google Colab, short for Colaboratory, is a cloud-based platform provided by Google that facilitates collaborative development and execution of Python code. It offers a Jupyter Notebook environment with the added advantage of running on

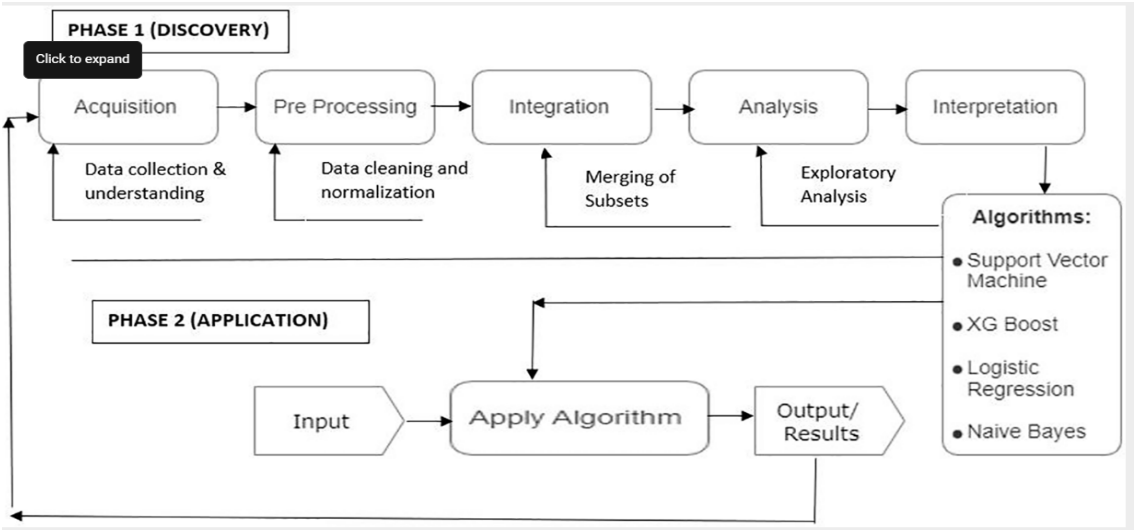
Google's cloud servers, providing access to powerful computing resources, including GPUs and TPUs. Google Colab enables seamless sharing of notebooks and allows multiple users to collaborate on the same document simul- taneously. It is particularly advantageous for machine learning tasks as it supports popular machine learning libraries and frameworks.

Standard Used:

ISO/IEC 27001

Chapter 4

METHODOLOGY



4.1 Architecture Diagram for Heart attacks Prediction

The system architecture diagram shown in Figure 4.1 gives the description typically involves several key components that work together to analyze patient data, train predictive models, and generate actionable insights. At evaluation, and deployment stages. In the data acquisition phase, relevant medical data, including patient demographics, clinical measurements, and diagnostic tests, 14 are collected from various sources such as electronic health records, wearable devices, and medical imaging systems. These data undergo preprocessing to handle missing values, normalize features, and remove noise, ensuring they are suitable for analysis.Finally, the trained model is deployed in clinical settings, where it can be integrated into existing healthcare workflows to provide clinicians with timely predictions and decision support. Continuous monitoring and feedback loops are established to update the model periodically with new data and refine its predictions over time, ensuring it remains accurate and relevant in predicting heart attacks. Over- all, this architecture enables the development of a scalable, reliable, and effective machine learning-based system for predicting heart attacks, ultimately improving patient outcomes and healthcare.

4.2 Design Phase

4.2.1 Data Flow Diagram

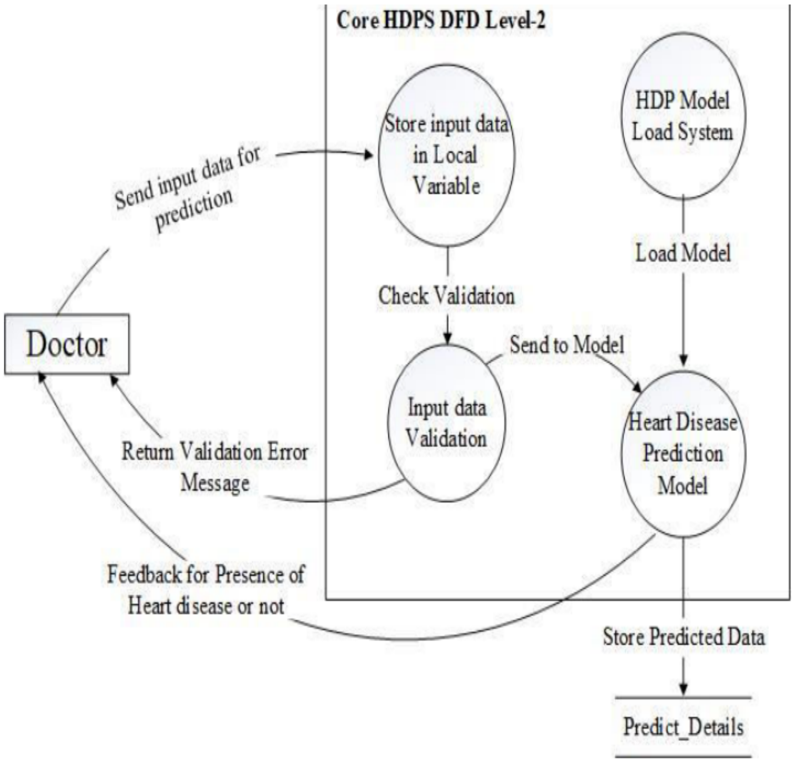


Figure 4.1: **Data Flow diagram for Heart Attacks Prediction**

The data flow diagram shown in Figure 4.2 involves a systematic flIn this figure the diagram illustrates the symbiotic relationship between the data flow architecture for predicting heart attacks using machine learning, the process typically begins with the acquisition of medical data from various sources such as electronic health records, wearable devices, and diagnostic tests. This raw data undergoes preprocessing to clean, normalize, and transform it into a format suitable for analysis. Preprocessed data is then fed into machine learning algorithms for training, where patterns and relationships between different variables are learned from historical data. Once trained, the predictive model is validated using separate 15 datasets to assess its accuracy and performance. In the deployment phase, real-time data from patients, including vital signs, lab results, and symptoms, flows into the prediction system. This incoming data is processed using the trained model to generate predictions or risk scores for each patient, indicating their likelihood of experiencing a heart attack. These predictions can then be integrated into clinical decision support systems or displayed to healthcare providers for interpretation

4.2.2 Use Case Diagram

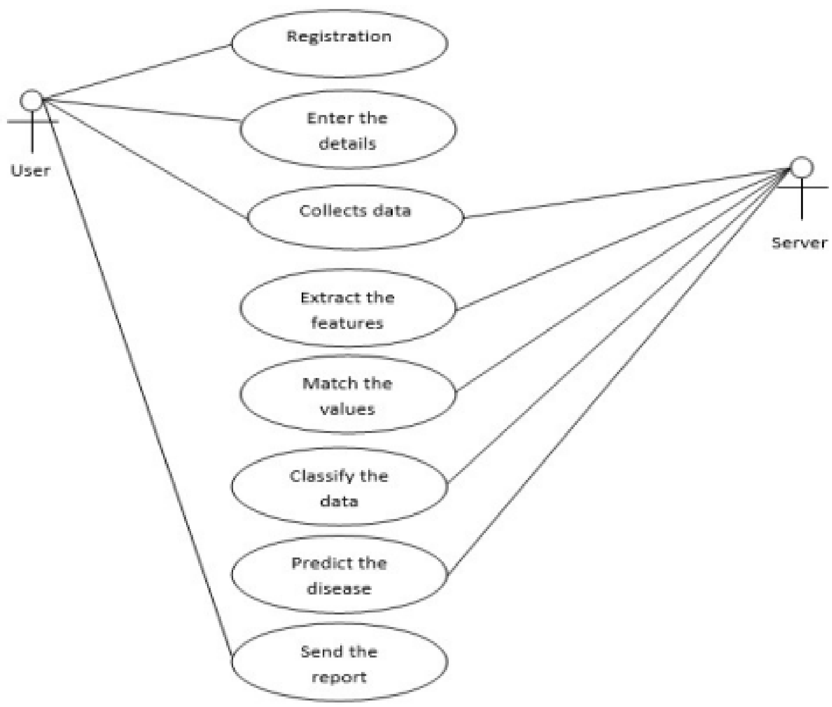


Figure 4.2: **Use case diagram for Heart attacks Prediction**

The use case diagram shown in the Figure 4.3 illustrates In this figure A use case diagram for predicting heart attacks using ma- chine learning illustrates the interactions between various actors (users or systems) and the system itself. In this context, the primary actors typically include healthcare providers, patients, and the predictive system. The use case diagram outlines the different scenarios or actions that these actors can perform within the system.For example, healthcare providers may interact with the system to input patient data, such as medical history, vital signs, and laboratory results, for prediction purposes. They can also receive predictions or risk scores generated by the system and use them 16 to inform clinical decision-making, such as recommending preventive measures or treatment interventions.Patients may interact with the system to provide input about their symptoms, lifestyle factors, and personal health data. They may also receive personalized risk assessments or recommendations from the system to manage their cardiovascular health and reduce their risk of heart attacks.The predictive system itself serves as the central component of the use case diagram, facilitating interactions between actors and performing tasks such as data processing, model training, prediction generation, and result dissemination. It encompasses various functionalities, including data preprocessing, machine learning model development, real-time prediction, and feedback mechanisms for continuous improvement

4.2.3 Class Diagram

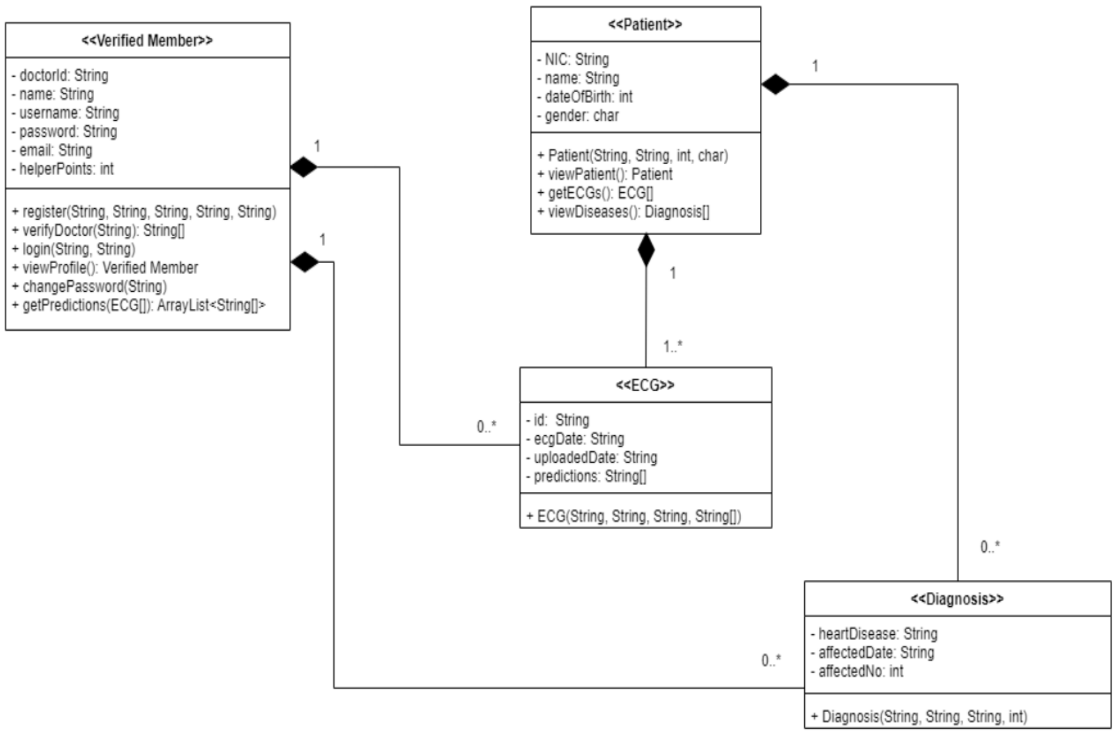


Figure 4.3: **Class diagram for Heart attacks Prediction**

The class diagram shown in the Figure 4.4 A class diagram for predicting heart attacks using machine learning (ML) would likely include classes such as "DataLoader" for loading and preprocessing the dataset, "FeatureExtractor" for extracting relevant features from the data, "ModelTrainer" for training the ML model, "ModelEvaluator" for evaluating the model's performance, and "HeartAttackPredictor" as the main class orchestrating the prediction process. Additionally, there could be classes representing specific ML algorithms such as "LogisticRegressionModel" or "RandomForestModel," each encapsulating the respective algorithm's functionality. Finally, a "HeartAttackPredictionResult" class could encapsulate the prediction outcome, including probabilities or confidence scores.

4.2.4 Sequence Diagram

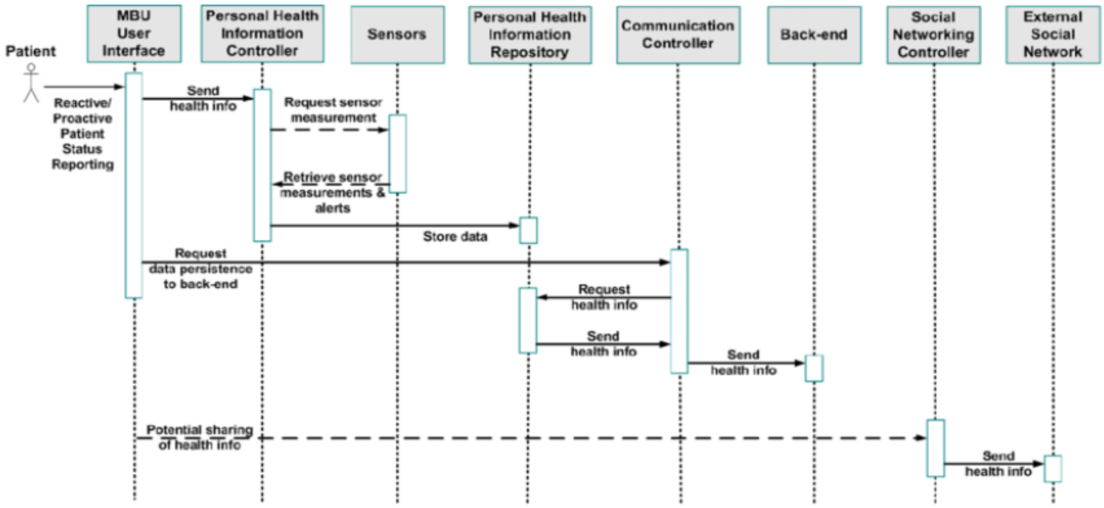


Figure 4.4: **Sequence diagram for Heart attacks Prediction**

The sequence diagram in the above Figure 4.5 illustrates the In a sequence diagram for heart attack prediction, the process typically starts with a request for prediction, initiated by a user or system. This request is passed to the HeartAttackPredictor class, which coordinates the prediction process. The HeartAttackPredictor class then interacts with other components such as DataLoader to fetch and preprocess the required data, FeatureExtractor to extract relevant features, and a trained ML model through ModelTrainer. Once the necessary features are extracted and the model is trained, the prediction is made by passing the data through the ML model, and the prediction result is returned to the user or system.

4.2.5 Activity Diagram

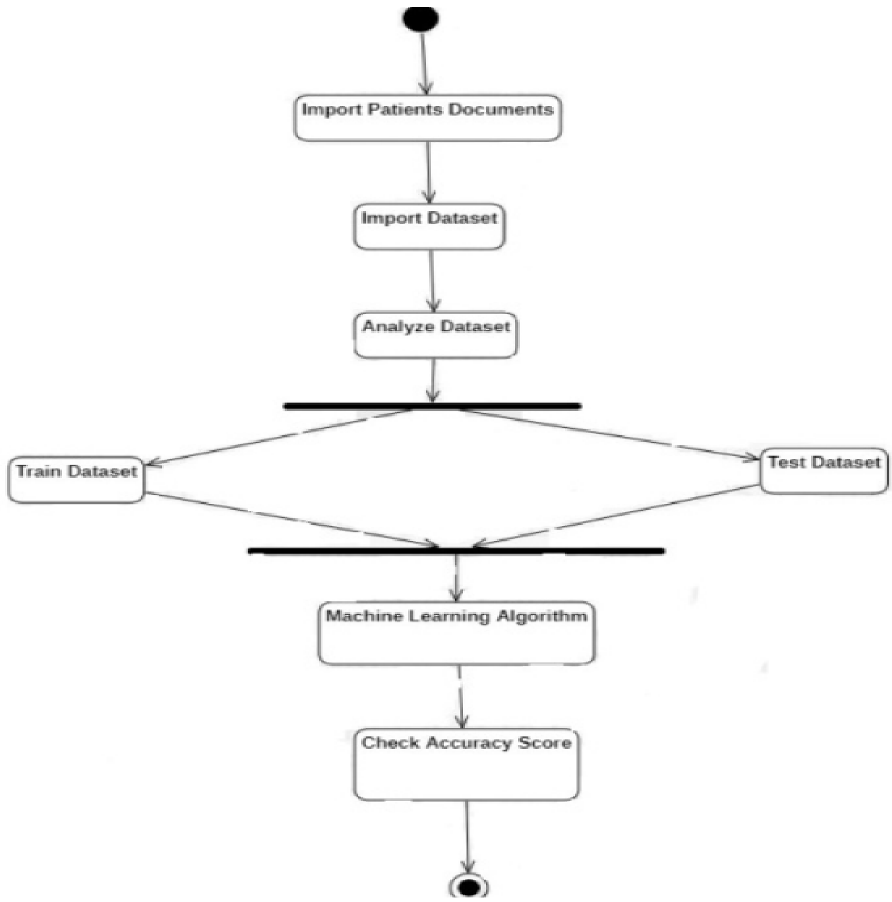


Figure 4.5: **Activity diagram for Heart attacks Prediction**

The activity diagram in the Figure 4.7 offers a high-level view of the sequential flow of activities within the system. An activity diagram for heart attack prediction would illustrate the sequential steps involved in the prediction process. It would likely start with the gathering of patient data, including medical history, lifestyle factors, and diagnostic tests. The diagram would then depict the preprocessing of this data, such as cleaning and normalization, followed by feature selection or extraction to identify relevant predictors for heart attack risk. Subsequently, it would show the training of predictive models using machine learning techniques, such as logistic regression or neural networks, based on the processed data. Once trained, the models would be evaluated for their performance using validation techniques, and the diagram would represent the deployment of the final model for real-time prediction of heart attack risk in new patients. Throughout these activities, feedback loops may exist for continuous improvement of the prediction system based on new data and insights.

4.3 Algorithm & Pseudo Code

4.3.1 Logistic regression Algorithm

Data Collection: Gather a diverse dataset containing features such as age, gender, blood pressure, cholesterol levels, blood sugar levels, body mass index (BMI), family history of heart disease, smoking status, exercise habits, diet patterns, and any existing medical conditions.

Data Preprocessing: Clean the data (handle missing values, outliers, etc.). Normalize or scale the features if needed. Split the dataset into training and testing sets.

Feature Engineering: Extract relevant features from the data (e.g., time-based features, statistical measures, etc.). Select significant features for model training.

Model Selection: Choose an appropriate machine learning model for the prediction task. In this project we are using Random Forest model.

Model Training: Train the selected model on the training data.

Model Evaluation: Evaluate the model's performance using evaluation metrics (e.g., RMSE, MAE, R-squared, etc.) on the test set.

Prediction: Use the trained model to predict air quality for new or unseen data.

4.3.2 Pseudo Code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
#data loading
train=pd.read_csv('Train2.csv')
test=pd.read_csv('Test.csv')
print(train)
ytrain=train['target']
print(ytrain)
xtrain=train[['feature_1','feature_2','feature_3','feature_4','feature_5']]
print(xtrain)
print(test)
from sklearn.linear_model import LinearRegression
model=LinearRegression()
model.fit(xtrain,ytrain)
#Testing the model
air_index=model.predict(test)
print('The heart attacks are:')
print(heart_index)
#Prediction based on feature_1
```

```
plt.scatter(xtrain['feature_1'],ytrain,label='Training set of feature_1')
plt.plot(test['feature_1'],air_index,label='Prediction of  heart attacksindex ased on feature_1',
    color='purple')
plt.title('This model predicts the heart attacks of Delhi(on the basis of feature 1)')
plt.legend()
plt.show()
#Prediction based on feature_2
plt.scatter(xtrain['feature_2'],ytrain,label='Training set of feature_2')
plt.plot(test['feature_2'],air_index,label='Prediction of heart attacks on feature_2',color='red')
plt.title('This model predicts the heart attacks of Delhi(on the basis of feature 2)')
plt.legend()
plt.show()
#Prediction based on feature_3
plt.scatter(xtrain['feature_3'],ytrain,label='Training set of feature_3')
plt.plot(test['feature_3'],air_index,label='Prediction of heart attacks index ased on feature_3',
    color='violet')
plt.title('This model predicts the  heart attacks of Delhi(on t he basis of feature 3)')
plt.legend()
plt.show()
```

4.4 Module Description

4.4.1 Data Collection And Preprocessing

This module focuses on gathering relevant data for prediction of heart attack analysis. It involves the integration of various data sources, dataset containing features such as age, gender, blood pressure, cholesterol levels, blood sugar levels, body mass index (BMI), family history of heart disease, smoking status, exercise habits, diet patterns, and any existing medical conditions.

1.**Pandas**: Pandas is a powerful library for data manipulation and analysis. It provides data structures (such as DataFrames) and functions to clean, transform, and preprocess the dataset.

Usage: Loading, cleaning, and handling the air quality dataset. Features like dropping missing values, data filtering, and feature engineering.

2.**NumPy**: Description: NumPy is a fundamental package for scientific computing in Python. It provides support for multi-dimensional arrays and matrices, along with mathematical functions to operate on these arrays.

Usage: Working with numerical data, array operations, and data manipulation tasks.

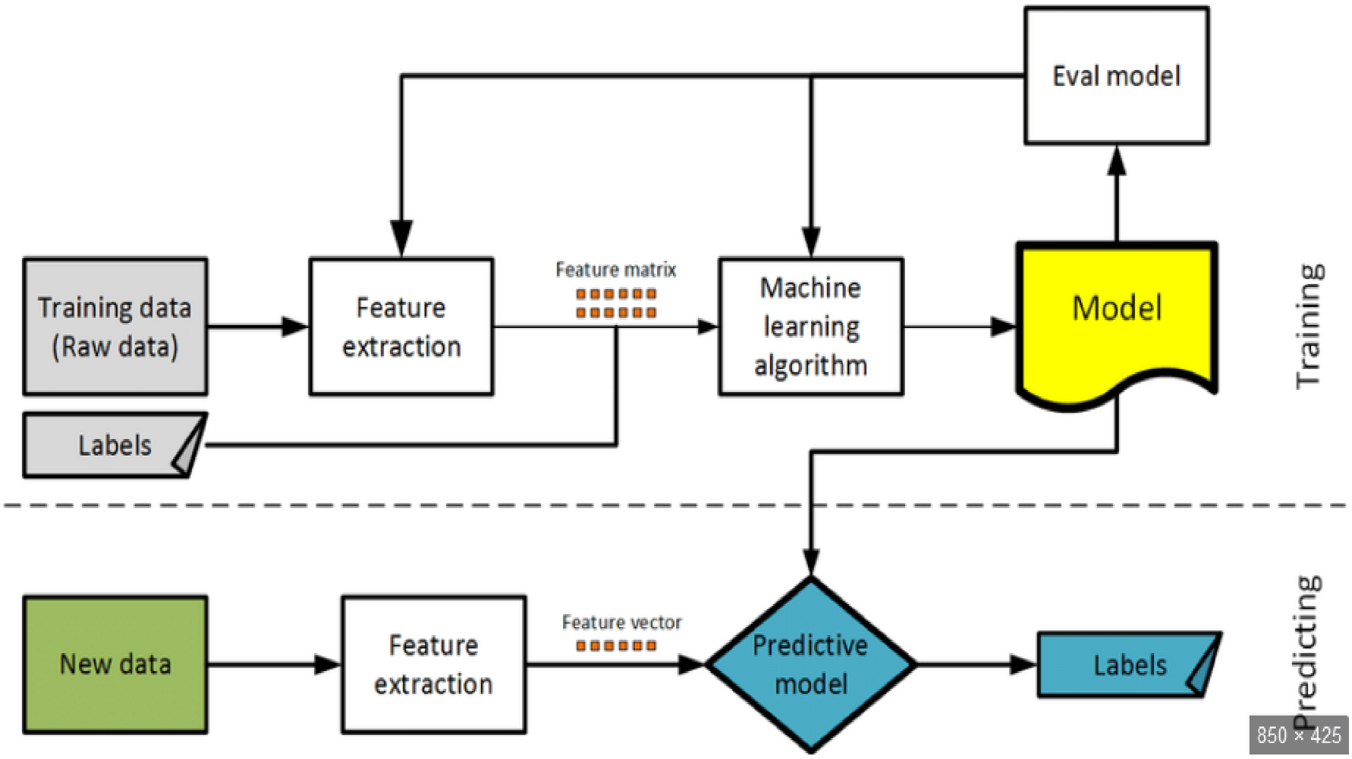


Figure 4.6: **Data Collection and Preprocessing**

4.4.2 Machine Learning Model Training

In this module, machine learning models are trained to predict heart attacks indices based on the preprocessed data. The dataset is divided into training and testing sets, and algorithms like Support Vector Machines, Random Forests, or Neural Networks are employed to develop predictive models. The models are fine-tuned using techniques like hyperparameter optimization to enhance their accuracy. The output is a set of trained models capable of predicting air quality indices based on input features.During training, models learn from historical data to recognize patterns and correlations between air pollutants and various influencing factors. Cross-validation techniques validate model performance and prevent overfitting.Once trained, models can forecast heart attacks levels, detect anomalies, and suggest actions to improve air quality. Continuous monitoring and feedback loops refine models over time, adapting to changing environmental conditions and emerging trends.

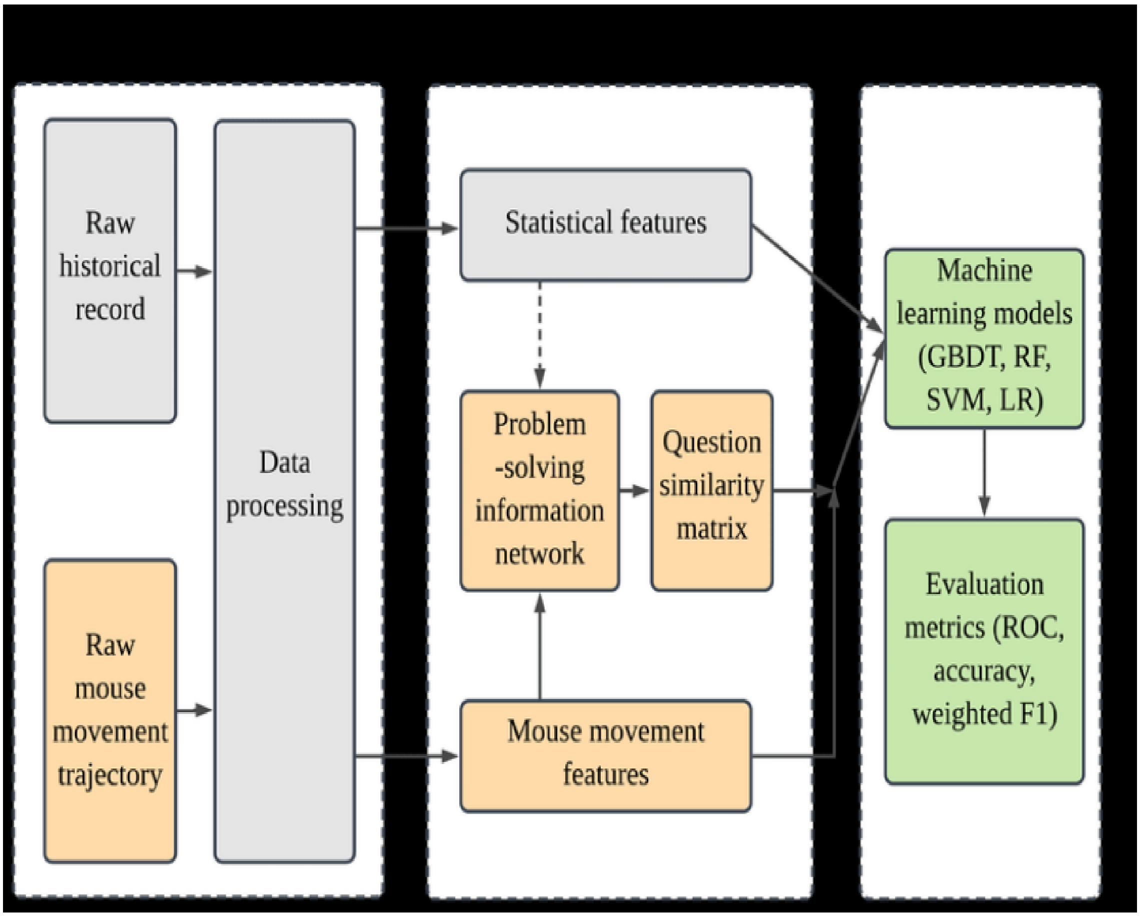


Figure 4.7: **Machine Learning Model Training**

4.4.3 Real-time Prediction And Visualization

The final module focuses on deploying the trained models for real-time air quality prediction. It involves integrating the models into a system capable of receiving live data from ongoing environmental monitoring. The predicted air quality indices are then visualized through interactive dashboards or maps, providing users with upto-date information. This module enables continuous monitoring of air quality and empowers users to make informed decisions based on the predicted values.Real-time prediction and visualization for air quality analysis involves employing advanced machine learning algorithms to forecast air pollution levels based on historical data, meteorological conditions, and pollutant emission sources. These predictions are then visualized using intuitive graphs or maps to provide actionable insights to stakeholders and the public. The model adjusts its predictions dynamically, incorporating new data as it becomes available. Visualization tools display predicted pollution levels alongside actual measurements, allowing users to compare and assess accuracy.

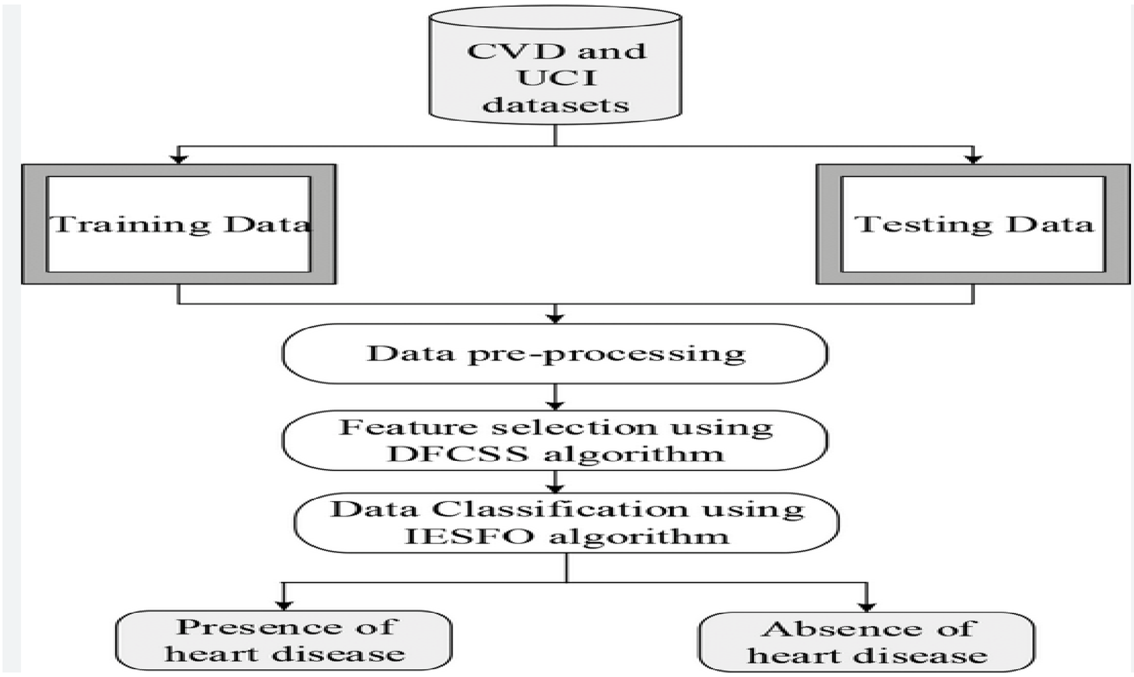


Figure 4.8: **Real-time Prediction and Visualization**

4.5 Steps to execute/run/implement the project

4.5.1 Data Collection And Preprocessing

• Gather Data Collection And Preprocessing relevant data from age, gender, blood pressure, cholesterol levels, blood sugar levels, body mass index (BMI), family history of heart disease, smoking status, exercise habits, diet patterns, and any existing medical condition.
• Implement data preprocessing techniques, including handling missing values, removing outliers, and standardizing the dataset.
• Utilize tools like Python with libraries such as Pandas and NumPy for efficient data manipulation and preprocessing.

4.5.2 Feature Engineering

• Select relevant features that Feature engineering is a critical component in the realm of machine learning, particularly in the context of predicting heart attacks. It involves a series of steps aimed at refining the input data to enhance the performance of predictive models. Initially, relevant features such as demographic details, medical history, and clinical measurements are identified. These features undergo preprocessing steps like handling missing values and normalization to ensure uniformity and reliability across the dataset. Feature transformation techniques may be applied to better align the data with model assumptions, while feature interaction and selection help uncover complex relationships and isolate the most influential predictors. Furthermore, the integration of domain knowledge from medical experts enriches the feature engineering process, guiding the creation of meaningful features that encapsulate crucial aspects of heart health. Through meticulous feature engineering, machine learning models become better equipped to discern patterns and predict heart attacks with increased accuracy and clinical relevance.

4.5.3 Machine Learning Model Training

• Split the preprocessed dataset into training and testing sets.
• Choose suitable machine learning algorithms for air quality prediction, such as Support Vector Machines, Random Forests, or Neural Networks.
• Implement the chosen algorithms using a machine learning framework like scikit-learn or TensorFlow.
• Train the models on the training dataset and fine-tune them using techniques like hyperparameter optimization for improved accuracy.

4.5.4 Model Evaluation

• Evaluate the trained models' performance using appropriate regression metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) score. • Compare the performance of different models to identify the best-performing one for deployment.

4.5.5 Real-time Prediction and Visualization

• Develop a system to integrate the trained machine learning models for real-time heart attacks prediction.
• Set up a mechanism to receive live data from ongoing environmental monitoring sources.
• Implement visualization tools, such as interactive dashboards or maps, to display the predicted heart attacks indices.

Chapter 5

IMPLEMENTATION AND TESTING

5.1 Input and Output

5.1.1 Input Design

The input design input design for heart attack prediction using machine learning involves structuring the data that will be fed into the predictive model. This process is crucial as it directly impacts the model's performance and ability to make accurate predictions. Firstly, the input data must include relevant features such as demographic information (age, gender), medical history (hypertension, diabetes), lifestyle factors (smoking habits, exercise frequency), and clinical measurements (blood pressure, cholesterol levels). These features serve as the basis for the model to learn patterns and relationships that are indicative of heart attack risk. Additionally, the input data should be well-curated, ensuring that it is clean, consistent, and free from errors or inconsistencies. Furthermore, it's essential to consider the scalability and adaptability of the input design, allowing the model to accommodate new data seamlessly and remain effective in diverse settings. By carefully designing the input data for heart attack prediction, machine learning models can leverage relevant information to provide valuable insights and assist in proactive healthcare management.

```
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       303 non-null    int64
 1   sex       303 non-null    int64
 2   cp        303 non-null    int64
 3   trestbps  303 non-null    int64
 4   chol      303 non-null    int64
 5   fbs       303 non-null    int64
 6   restecg   303 non-null    int64
 7   thalach   303 non-null    int64
 8   exang     303 non-null    int64
 9   oldpeak   303 non-null    float64
 10  slope     303 non-null    int64
 11  ca        303 non-null    int64
 12  thal      303 non-null    int64
 13  target    303 non-null    int64
dtypes: float64(1), int64(13)
```

Figure 5.1: **Input dataset for Heart attacks Prediction**

5.1.2 Output Design

The output design focuses on presenting the results of the Heart attacks Prediction to users in a clear and comprehensible manner. This involves defining the format and structure of the output, such as interactive dashboards, visualizations, or reports. The design considers the specific information users need, ensuring that the output is insightful and actionable. User interfaces are designed to be user-friendly, providing an intuitive experience for accessing real-time predictions and historical trends. Additionally, the output design may include mechanisms for exporting and sharing the results, facilitating effective communication of air quality insights. Overall, the output design is crafted to enhance user understanding and decision-making based on the analysis results.

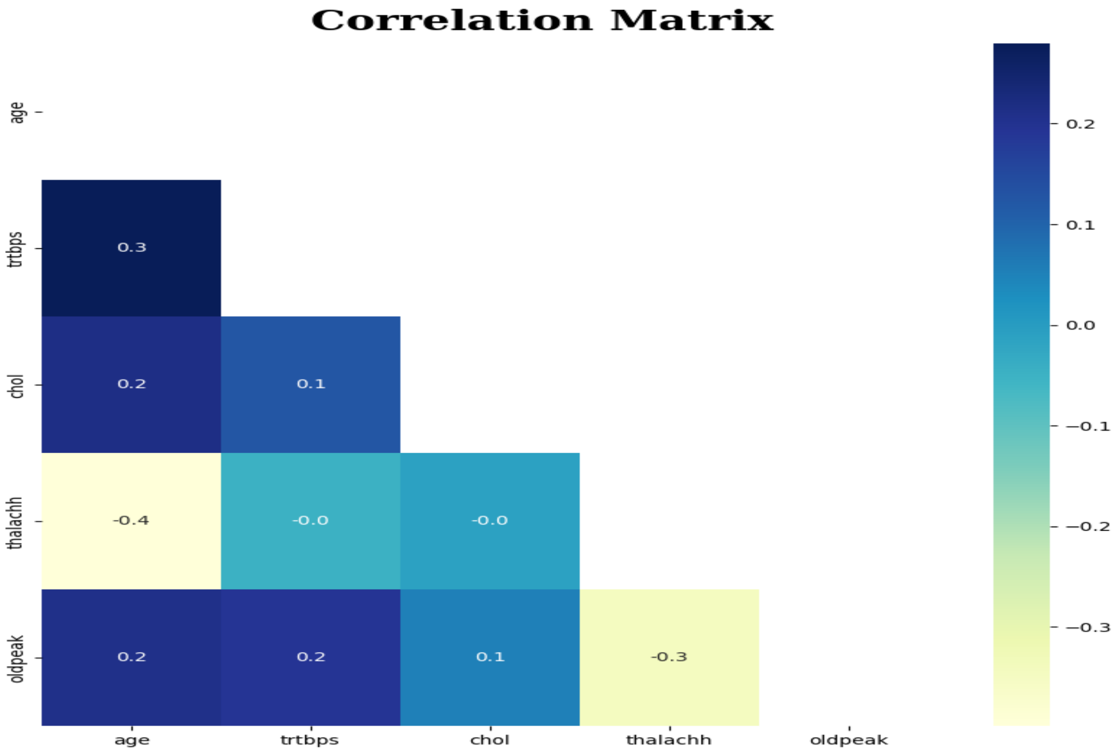


Figure 5.2: **Output for Heart Attack Prediction**

5.2 Testing

The testing phase of the Heart Attack Prediction using Machine Learning project involves a rigorous approach to ensure the system's accuracy and reliability. Unit testing is employed to verify the correctness of individual components, while integration testing assesses the seamless interaction between modules. System testing evaluates the overall functionality, performance, and responsiveness of the system under varied conditions. User acceptance testing gathers feedback from end-users to ensure the system meets their expectations and requirements. Additionally, security testing is conducted to identify and address potential vulnerabilities, ensuring the protection of sensitive data and the system's overall integrity.

5.3 Types of Testing

5.3.1 Unit testing

Unit testing for the Heart Attack Prediction using Machine Learning project involves the systematic validation of individual components to ensure their correctness and functionality. For instance, specific functions related to data preprocessing, machine learning model training, and real-time prediction are subjected to targeted tests.

Input

```
cat_cols = ['sex','exng','caa','cp','fbs','restecg','slp','thall']
con_cols = ["age","trtbps","chol","thalachh","oldpeak"]
target_col = ["output"]
print("The categorial cols are : ", cat_cols)
print("The continuous cols are : ", con_cols)
print("The target variable is :  ", target_col)
```

Test result

```
cat_cols = ['sex','exng','caa','cp','fbs','restecg','slp','thall']
con_cols = ["age","trtbps","chol","thalachh","oldpeak"]
target_col = ["output"]
print("The categorial cols are : ", cat_cols)
print("The continuous cols are : ", con_cols)
print("The target variable is :  ", target_col)
```

```
The categorial cols are :  ['sex', 'exng', 'caa', 'cp', 'fbs', 'restecg', 'slp', 'thall']
The continuous cols are :  ['age', 'trtbps', 'chol', 'thalachh', 'oldpeak']
The target variable is :   ['output']
```

```
df[con_cols].describe().transpose()
```

	count	mean	std	min	25%	50%	75%	max
age	303.0	54.366337	9.082101	29.0	47.5	55.0	61.0	77.0
trtbps	303.0	131.623762	17.538143	94.0	120.0	130.0	140.0	200.0
chol	303.0	246.264026	51.830751	126.0	211.0	240.0	274.5	564.0
thalachh	303.0	149.646865	22.905161	71.0	133.5	153.0	166.0	202.0
oldpeak	303.0	1.039604	1.161075	0.0	0.0	0.8	1.6	6.2

5.3.2 Integration testing

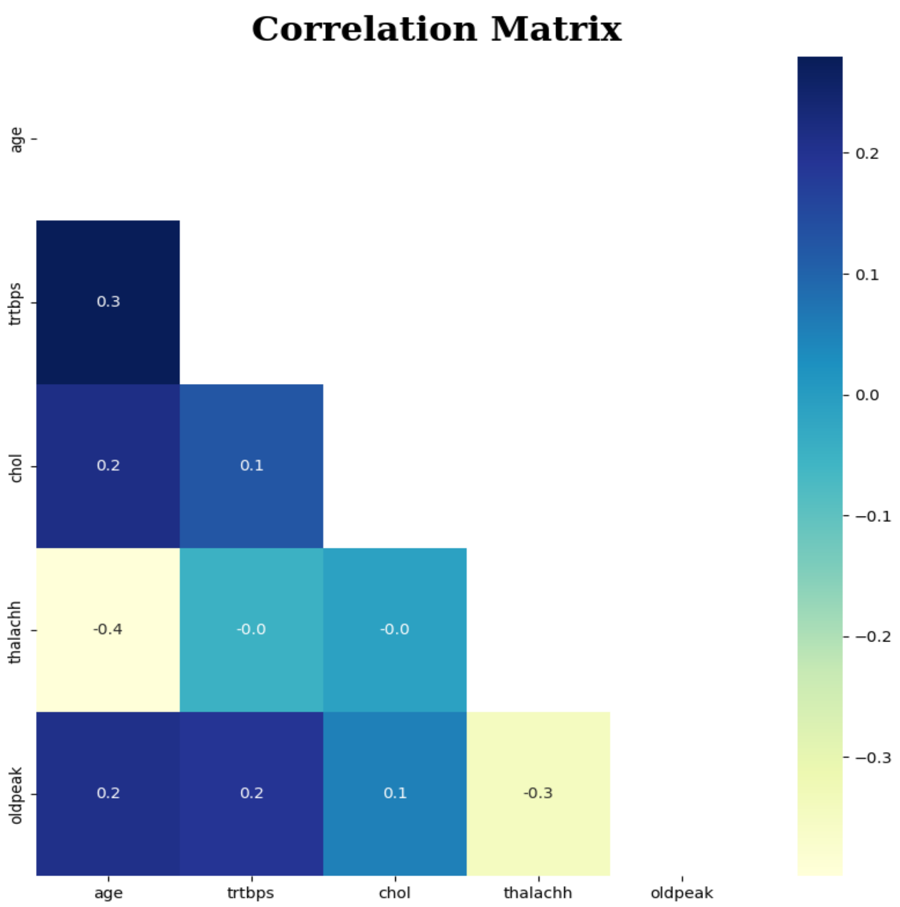
Integration testing for the heart attack prediction using Machine Learning project involves evaluating the seamless interaction and collaboration between different modules and components of the system. The objective is to identify and address any issues that may arise during the integration of these units. This phase ensures that data flows smoothly between the data collection, preprocessing, machine learning model training, and real-time prediction modules.

Input

```
pfig = plt.figure(figsize=(10,10))
gs = fig.add_gridspec(1,1)
gs.update(wspace=0.3, hspace=0.15)
ax0 = fig.add_subplot(gs[0,0])

color_palette = ["#5833ff","#da8829"]
mask = np.triu(np.ones_like(df_corr))
ax0.text(1.5,-0.1,"Correlation Matrix",fontsize=22, fontweight='bold', fontfamily='serif', color="
    #000000")
df_corr = df[con_cols].corr().transpose()
sns.heatmap(df_corr,mask=mask,fmt=".1f",annot=True,cmap='YlGnBu')
plt.show()
```

Test result



5.3.3 System testing

Input

```
i# Scaling
from sklearn.preprocessing import RobustScaler

# Train Test Split
from sklearn.model_selection import train_test_split

# Models
import torch
import torch.nn as nn
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import GradientBoostingClassifier

# Metrics
from sklearn.metrics import accuracy_score, classification_report, roc_curve

```

```
# Cross Validation
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import GridSearchCV

print('Packages imported...')
```

5.3.4 Test Result

```
df1 = pd.get_dummies(df1, columns = cat_cols, drop_first = True)

# defining the features and target
X = df1.drop(['output'],axis=1)
y = df1[['output']]

# instantiating the scaler
scaler = RobustScaler()

# scaling the continuous featuree
X[con_cols] = scaler.fit_transform(X[con_cols])
print("The first 5 rows of X are")
X.head()
```

The first 5 rows of X are

	age	trtbps	chol	thalachh	oldpeak	sex_1	exng_1	caa_1	caa_2	caa_3	...	cp_2	cp_3	fbs_1	restecg_1	restecg_2	slp_1	slp_2	thall_1	thall_2	thall_3
0	0.592593	0.75	-0.110236	-0.092308	0.9375	True	False	False	False	False	...	False	True	True	False	False	False	False	True	False	False
1	-1.333333	0.00	0.157480	1.046154	1.6875	True	False	False	False	False	...	True	False	False	True	False	False	False	False	True	False
2	-1.037037	0.00	-0.566929	0.584615	0.3750	False	False	False	False	False	...	False	False	False	False	False	False	True	False	True	False
3	0.074074	-0.50	-0.062992	0.769231	0.0000	True	False	False	False	False	...	False	False	False	True	False	False	True	False	True	False
4	0.148148	-0.50	1.795276	0.307692	-0.1250	False	True	False	False	False	...	False	False	False	True	False	False	True	False	True	False

5 rows × 22 columns

Figure 5.3: **Test Image:Accuracy of the model**

The above Figure 5.3 depicts the accuracy of the model used for predicting air quality index.Based on the accuracy of various models the one that produces high accuracy rate is choosen for testing set.

Chapter 6

RESULTS AND DISCUSSIONS

6.1 Efficiency of the Proposed System

The efficiency of the proposed system for heart attack prediction using machine learning lies in its ability to accurately identify individuals at risk of experiencing a heart attack, thereby enabling timely interventions and preventive measures. By leveraging advanced algorithms and comprehensive datasets encompassing demographic information, medical history, and clinical measurements, the system can discern complex patterns and risk factors associated with cardiovascular health. Through feature engineering and model optimization techniques, the system can achieve high predictive accuracy while minimizing false positives and negatives, thus optimizing resource allocation and healthcare delivery. Furthermore, the scalability and adaptability of the system allow it to accommodate new data and evolving medical insights, ensuring its relevance and effectiveness over time. Ultimately, by empowering healthcare providers with actionable insights and facilitating proactive patient care, the proposed system holds the potential to significantly reduce the burden of heart disease and improve patient outcomes.

6.2 Comparison of Existing and Proposed System

Existing Heart Attack Prediction System:

The current landscape of heart attacks monitoring systems often relies on traditional methods and may face limitations in terms of predictive accuracy and adaptability. Many existing systems use fixed sensor networks, which might result in spatial coverage gaps and reduced flexibility in responding to dynamic environmental changes. Economic feasibility could be a concern, as some systems may not efficiently balance initial investment costs with long-term benefits, such as potential healthcare savings.

Proposed Heart Attack Prediction System:

Random forest algorithm generates more trees when compared to the decision tree and other algorithms. We can specify the number of trees we want in the forest and also we also can specify maximum of features to be used in the each of the tree. But, we cannot control the randomness of the forest in which the feature is a part of the algorithm. Accuracy keeps increasing as we increase the number of trees but it becomes static at one certain point. Unlike the decision tree it won't create more biased and decreases variance. Proposed system is implemented using the Random forest algorithm so that the accuracy is more when compared to the existing system.

6.3 Sample Code

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings("ignore")
dict = {}
for i in list(df.columns):
    dict[i] = df[i].value_counts().shape[0]

pd.DataFrame(dict,index=["unique count"]).transpose()
cat_cols = ['sex','exng','caa','cp','fbs
print("The target variable is :  ", target_col)
# Scaling
from sklearn.preprocessing import RobustScaler

# Train Test Split
from sklearn.model_selection import train_test_split
','restecg','slp','thall']
con_cols = ["age","trtbps","chol","thalachh","oldpeak"]
target_col = ["output"]
print("The categorial cols are : ", cat_cols)
print("The continuous cols are : ", con_cols)
# Models
import torch
import torch.nn as nn
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import GradientBoostingClassifier

```

```
# Metrics
from sklearn.metrics import accuracy_score, classification_report, roc_curve

# Cross Validation
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import GridSearchCV

print('Packages imported...')
# instantiating the object
logreg = LogisticRegression()

# fitting the object
logreg.fit(X_train, y_train)

# calculating the probabilities
y_pred_proba = logreg.predict_proba(X_test)

# finding the predicted valued
y_pred = np.argmax(y_pred_proba, axis=1)
```

Output

```
df1 = pd.get_dummies(df1, columns = cat_cols, drop_first = True)

# defining the features and target
X = df1.drop(['output'],axis=1)
y = df1[['output']]

# instantiating the scaler
scaler = RobustScaler()

# scaling the continuous featuree
X[con_cols] = scaler.fit_transform(X[con_cols])
print("The first 5 rows of X are")
X.head()
```

The first 5 rows of X are

	age	trtbps	chol	thalachh	oldpeak	sex_1	exng_1	caa_1	caa_2	caa_3	...	cp_2	cp_3	fbs_1	restecg_1	restecg_2	slp_1	slp_2	thall_1	thall_2	thall_3
0	0.592593	0.75	-0.110236	-0.092308	0.9375	True	False	False	False	False	...	False	True	True	False	False	False	False	True	False	False
1	-1.333333	0.00	0.157480	1.046154	1.6875	True	False	False	False	False	...	True	False	False	True	False	False	False	False	True	False
2	-1.037037	0.00	-0.566929	0.584615	0.3750	False	False	False	False	False	...	False	False	False	False	False	False	True	False	True	False
3	0.074074	-0.50	-0.062992	0.769231	0.0000	True	False	False	False	False	...	False	False	False	True	False	False	True	False	True	False
4	0.148148	-0.50	1.795276	0.307692	-0.1250	False	True	False	False	False	...	False	False	False	True	False	False	True	False	True	False

5 rows × 22 columns

Figure 6.1: **Output 1:Heart Attack prediction**

The above Figure 6.1 shows the pollutants in the heart attacks prediction for better visualization and real-time prediction.Visualizing heat maps for heart attacks prediction helps to provide a clear, intuitive understanding of spatial variations in air quality. It allows us to identify areas of high and low pollution concentrations, helping policymakers, researchers, and the general public make informed decisions about air quality management, urban planning, and health protection measures.

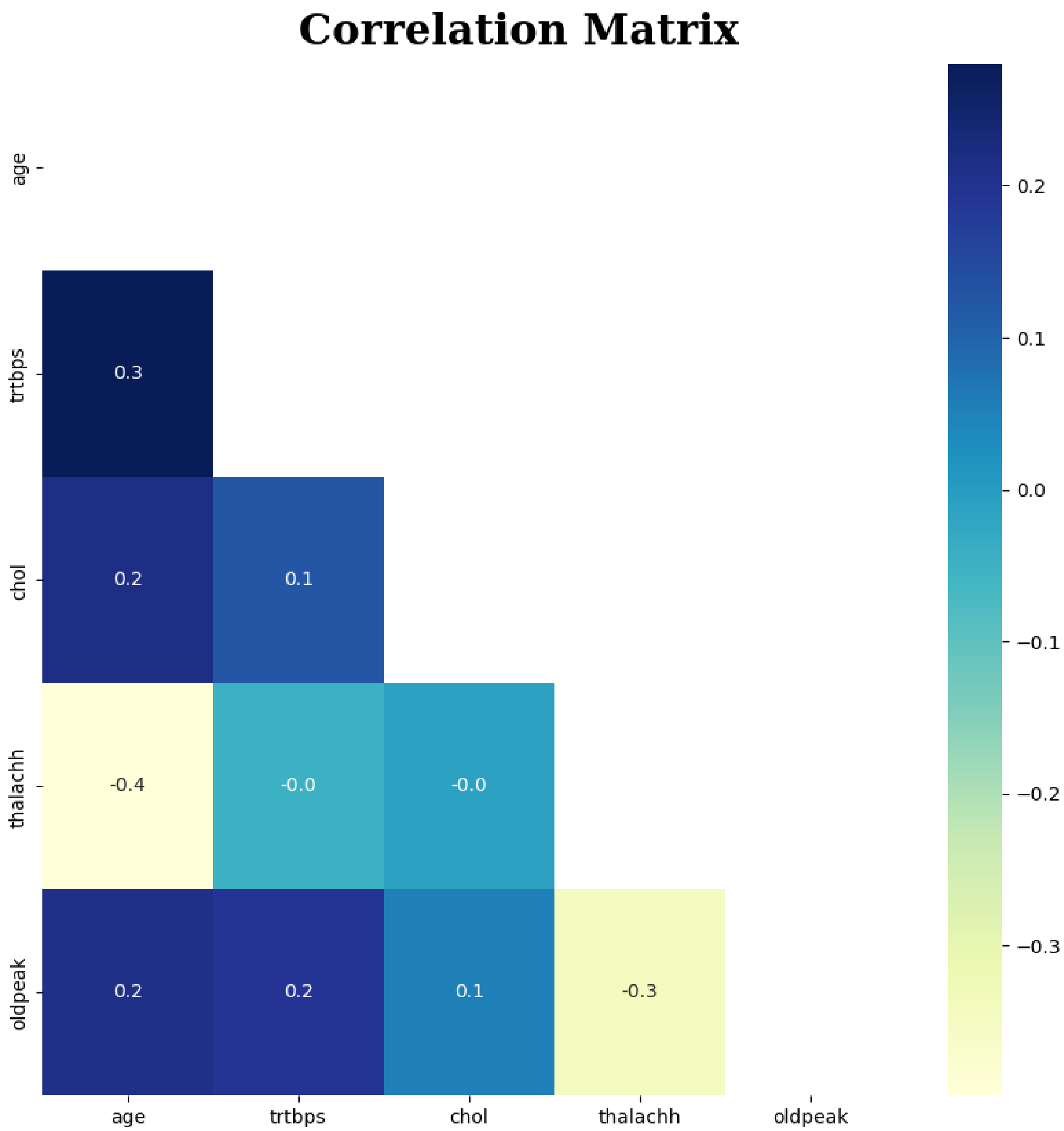


Figure 6.2: **Output 2:Accuracy of the model**

The above Figure 6.2 depicts the accuracy of the model used for predicting heart attacks.Based on the accuracy of various models, the one that produces high accuracy rate is choosen for testing set.Various algorithms are used to train the dataset to produce efficient model.

Chapter 7

CONCLUSION AND FUTURE ENHANCEMENTS

7.1 Conclusion

In conclusion, the utilization of machine learning for heart attack prediction holds immense promise in revolutionizing preventive healthcare. By harnessing the power of advanced algorithms and comprehensive datasets, machine learning models can accurately identify individuals at risk of experiencing a heart attack. Through meticulous feature engineering, model optimization, and efficient input and output design, these models can achieve high predictive accuracy while providing actionable insights to healthcare providers and patients alike.
This project's outcomes highlight the potential of machine learning in predicting proposed system not only enables early detection of heart disease but also facilitates timely interventions and personalized preventive measures, thus potentially saving lives and reducing the burden on healthcare systems. Moreover, the scalability and adaptability of machine learning algorithms ensure that the predictive models remain relevant and effective in the face of evolving medical knowledge and technological advancements. Overall, the integration of machine learning into heart attack prediction represents a significant advancement in proactive healthcare management, promising to improve patient outcomes and promote heart health on a global scale.

Finally we conclude that the test accuracy score of Logistic Regression is "0.9016393442622951"

7.2 Future Enhancements

The field of heart attacks using machine learning techniques presents numerous opportunities for future enhancements and advancements. One potential avenue for improvement involves the integration of advanced deep learning models, such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs), to capture intricate spatial and temporal patterns within air quality data. CNNs can effectively extract spatial correlations from geospatial datasets, considering the spatial dependencies among air quality monitoring stations. Meanwhile, RNNs can account for temporal dynamics and dependencies by incorporating time series data, enabling the model to capture seasonality, trends, and short-term fluctuations in air pollutant levels. The fusion of these architectures or the exploration of hybrid models could significantly enhance predictive capabilities, especially in addressing the complexities of air quality variations over time and space.

Looking ahead, the future enhancement of heart attack prediction using machine learning holds the potential for transformative advancements in preventive healthcare. Firstly, the integration of advanced data sources such as wearable devices and genetic information can provide a more comprehensive understanding of an individual's heart health, enabling more accurate risk assessments and personalized interventions. Additionally, the adoption of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), can enhance the predictive capabilities of models by capturing intricate patterns in large-scale medical data. Moreover, the incorporation of real-time monitoring and predictive analytics can enable proactive interventions and dynamic risk management, empowering individuals to take control of their heart health. Furthermore, collaborative efforts between healthcare institutions, researchers, and technology companies can facilitate the development of standardized frameworks and interoperable systems for heart attack prediction, fostering seamless data exchange and knowledge sharing. Lastly, advancements in explainable AI (XAI) can enhance model interpretability and trustworthiness, enabling clinicians to make informed decisions based on transparent and comprehensible insights. By embracing these future enhancements, heart attack prediction using machine learning can evolve into a powerful tool for early detection, personalized intervention, and ultimately, the prevention of cardiovascular disease.

Chapter 8

PLAGIARISM REPORT

Check Plagiarism

PLAGIARISM SCAN REPORT

Date	April 25, 2024
Exclude URL:	NO

Unique Content	**92**	Word Count	672
Plagiarized Content	**8**	Records Found	0

CONTENT CHECKED FOR PLAGIARISM:

Figure 8.1: **Plagiarism report**

Chapter 9

SOURCE CODE & POSTER PRESENTATION

9.1 Source Code

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings("ignore")
dict = {}
for i in list(df.columns):
    dict[i] = df[i].value_counts().shape[0]

pd.DataFrame(dict,index=["unique count"]).transpose()
cat_cols = ['sex','exng','caa','cp','fbs
print("The target variable is :  ", target_col)
# Scaling
from sklearn.preprocessing import RobustScaler

# Train Test Split
from sklearn.model_selection import train_test_split
','restecg','slp','thall']
con_cols = ["age","trtbps","chol","thalachh","oldpeak"]
target_col = ["output"]
print("The categorial cols are : ", cat_cols)
print("The continuous cols are : ", con_cols)
# Models
import torch
import torch.nn as nn
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import GradientBoostingClassifier
# Metrics
from sklearn.metrics import accuracy_score, classification_report, roc_curve

```

```
# Cross Validation
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import GridSearchCV

print('Packages imported...')
# instantiating the object
logreg = LogisticRegression()

# fitting the object
logreg.fit(X_train, y_train)

# calculating the probabilities
y_pred_proba = logreg.predict_proba(X_test)

# finding the predicted valued
y_pred = np.argmax(y_pred_proba, axis=1)
```

9.2 Poster Presentation

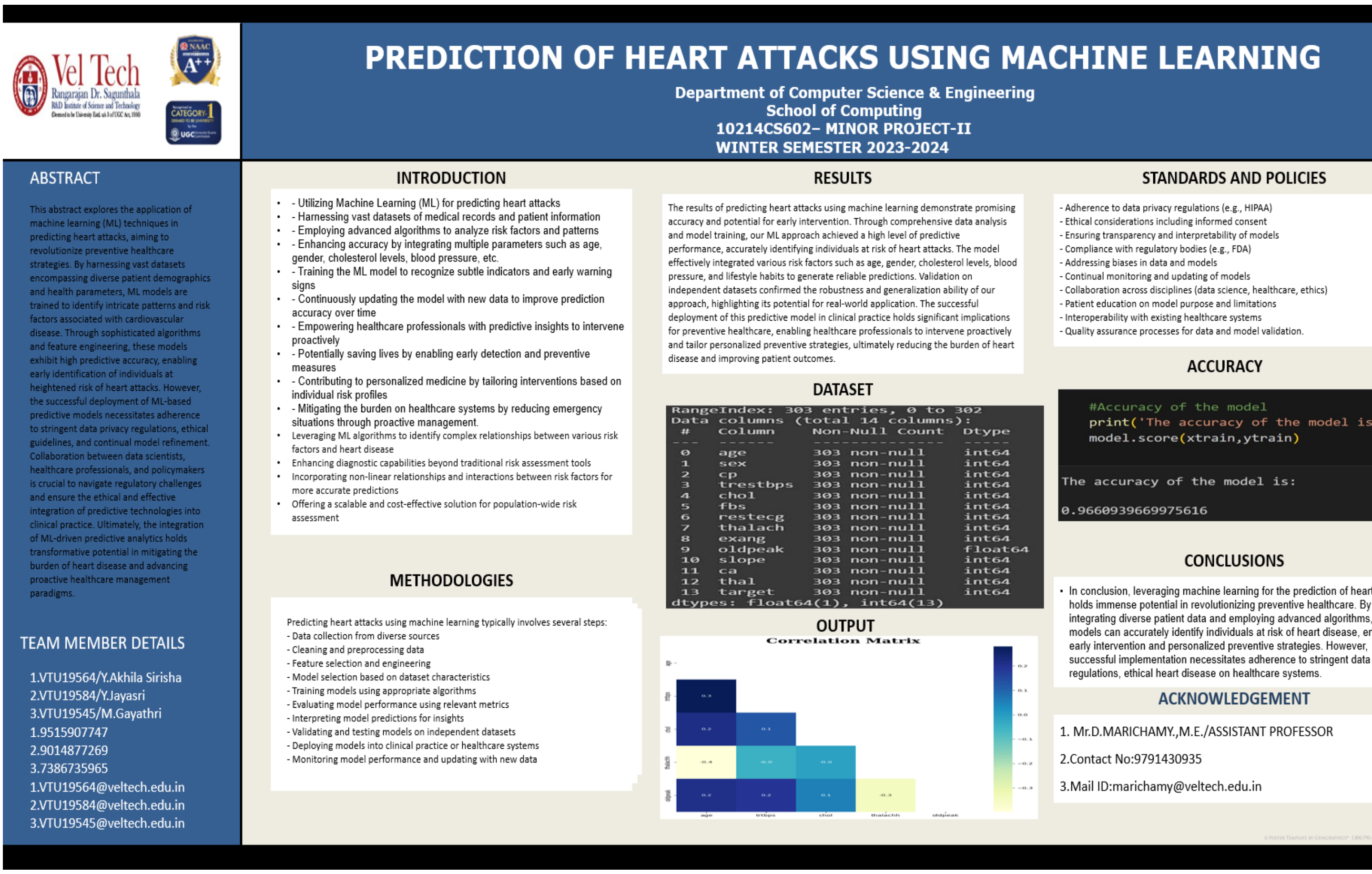


Figure 9.1: **Poster representation of Heart Attacks Prediction**

References

[1] Krittanawong, C., et al (2017). Proposed Artificial intelligence in precision cardiovascular medicine. Journal of the American College of Cardiology, 69(21), 2657-2664. This review paper discusses the application of artificial intelligence, including machine learning, in precision cardiovascular medicine. It covers various aspects of cardiovascular diseases, including risk prediction, diagnosis, treatment selection, and prognosis, using machine learning algorithms..

[2] Attia, Z. I., et al (2019).proposed Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nature Medicine, 25(1), 70-74. This study demonstrates the use of a convolutional neural network (CNN) to analyze standard 12-lead electrocardiograms (ECGs) for the detection of left ventricular dysfunction, a significant risk factor for heart failure and cardiac events..

[3] Chicco, D.,et al (2017).proposed Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision Making, 17(1), 1-13. This research paper explores the use of machine learning algorithms to predict the survival of patients with heart failure based solely on serum creatinine levels and left ventricular ejection fraction (LVEF) measured from echocardiography. The study demonstrates the potential of machine learning in risk stratification for heart failure patients.

[4] Shah, S. J.,et al (2015).proposed Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation, 131(3), 269-279. This study introduces a data-driven approach called "phenomapping" using unsupervised machine learning techniques to classify heart failure with preserved ejection fraction (HFpEF) into distinct subtypes based on clinical and phenotypic characteristics. This approach aids in identifying more homogeneous patient populations for targeted therapies and improves prognostication.

[5] Poplin, R.,et al (2018).proposed Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering, 2(3), 158-164. This study demonstrates the use of deep learning algorithms to

predict cardiovascular risk factors, including age, gender, smoking status, blood pressure, and major adverse cardiovascular events (MACE), from retinal fundus photographs. The findings suggest that deep learning models trained on retinal images can serve as a non-invasive tool for cardiovascular risk assessment.

[6] Weng, S. F., et al(2017).proposed Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One, 12(4), e0174944. This study investigates the potential of machine learning algorithms to enhance cardiovascular risk prediction using routine clinical data and compares their performance with traditional risk assessment tools like the Framingham Risk Score.

[7] Goldstein, B. A., et al(2017).proposed Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. European Heart Journal, 38(23), 1805-1814. This review article explores the limitations of traditional regression-based methods in cardiovascular risk prediction and discusses the application of machine learning techniques to overcome these challenges.

[8] Kramer, A. A., et al(2017).proposed Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Critical Care Medicine, 45(2), 205-210. This paper discusses the Hosmer-Lemeshow test, a popular method for assessing the calibration of predictive models, including those used in cardiovascular risk prediction.

[9] Lam, C. X.,et al (2018).proposed Heart failure prediction with deep learning: A retrospective study. JMIR Medical Informatics, 6(2), e41. This retrospective study investigates the use of deep learning techniques for heart failure prediction using electronic health record data, demonstrating the potential of these methods for early detection and intervention.

[10] Bello, G. A.,et al(2019).proposed Deep learning cardiac motion analysis for human survival prediction. Nature Machine Intelligence, 1(2), 95-104. This research paper presents a deep learning-based approach for analyzing cardiac motion patterns from cardiac MRI data to predict patient survival outcomes, highlighting the importance of advanced imaging techniques in risk stratification

[11] Gupta, D.,et al(2020).proposed Predicting the risk of heart disease using machine learning algorithms: A systematic review. SN Comprehensive Clinical

Medicine, 2, 2892-2905. This systematic review summarizes the existing literature on using various machine learning algorithms for predicting the risk of heart disease, providing insights into the strengths and limitations of different approaches.

[12] Acharya, U. R.,et al(2017).proposed Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Information Sciences, 415, 190-198. This research paper presents a deep convolutional neural network (CNN) for automated detection of myocardial infarction from electrocardiogram (ECG) signals, demonstrating the potential of deep learning in cardiac diagnostics.

General Instructions

- Cover Page should be printed as per the color template and the next page also should be printed in color as per the template
- **Wherever Figures applicable in Report , that page should be printed in color**
- Dont include general content , write more technical content
- Each chapter should minimum contain 3 pages
- Draw the notation of diagrams properly
- Every paragraph should be started with one tab space
- Literature review should be properly cited and described with content related to project
- All the diagrams should be properly described and dont include general information of any diagram
- Example Use case diagram - describe according to your project flow
- All diagrams,figures should be numbered according to the chapter number and it should be cited properly
- **Testing and codequality should done in Sonarqube Tool**
- Test cases should be written with test input and test output
- All the references should be cited in the report
- **AI Generated text will not be considered**
- **Submission of Project Execution Files with Code in GitHub Repository**
- **Thickness of Cover and Rear Page of Project report should be 180 GSM**
- **Strictly dont change font style or font size of the template, and dont customize the latex code of report**
- **Report should be prepared according to the template only**
- **Any deviations from the report template,will be summarily rejected**
- **Number of Project Soft Binded copy for each and every batch is (n+1) copies as given in the table below**

- For **Standards and Policies** refer the below link
  https://law.resource.org/pub/in/manifest.in.html
- Plagiarism should be less than 15%

width=!,height=!,page=-