**EARLY DETECTION OF HEART ATTACKS USING MACHINE LEARNING**

**By**

## DECLARATION

This thesis is our original work and has not been presented for a degree in any other University.

Signature…………………………………………… Date…………………………………………………

**William Bukachi**

Signature…………………………………………… Date………………………………………………….

**Alex Ochola**

This thesis has been submitted for examination with our approval as University Supervisor

Signature……………………………………………. Date………………………………………………….

Dr. Teresa Abuya Kisii University, Kenya

Table of Contents

[DECLARATION 2](#_Toc148534534)

[Abstract 5](#_Toc148534535)

[Keywords and Abbreviations 6](#_Toc148534536)

[Dedication 7](#_Toc148534537)

[CHAPTER 1 8](#_Toc148534538)

[1.0 Introduction 8](#_Toc148534539)

[1.1 Background 8](#_Toc148534540)

[1.2 Statement of the problem 9](#_Toc148534541)

[1.3 Objectives 10](#_Toc148534542)

[1.3.1 General objective 10](#_Toc148534543)

[1.3.2 Specific objectives 10](#_Toc148534544)

[1.4 Scope of Study 10](#_Toc148534545)

[1.5 Justification 11](#_Toc148534546)

[CHAPTER 2: LITERATURE REVIEW 12](#_Toc148534547)

[2.1 Introduction 12](#_Toc148534548)

[2.2 Related Works 12](#_Toc148534549)

[2.3 Proposed system 18](#_Toc148534550)

[2.4 Challenges with the existing systems 19](#_Toc148534551)

[CHAPTER 3: METHODOLOGY 20](#_Toc148534552)

[3.0 Introduction 20](#_Toc148534553)

[3.1 Dataset Description 20](#_Toc148534554)

[3.1.1 Dataset Location 20](#_Toc148534555)

[3.1.2 Dataset Definition 20](#_Toc148534556)

[3.1.3 Data Splitting 21](#_Toc148534557)

[3.2 Data Preprocessing 21](#_Toc148534558)

[3.2.1 Data Cleaning 21](#_Toc148534559)

[3.3 Support Vector Classifier (SVC) as the Chosen Algorithm 22](#_Toc148534560)

[3.4 Feature Engineering and selection 23](#_Toc148534561)

[3.5 System flowchart 24](#_Toc148534562)

[3.6Model Development, Hyperparameter Tuning, and Evaluation 24](#_Toc148534563)

[3.7 Building an Intuitive User Interface 25](#_Toc148534564)

[3.7.1 Input Fields 25](#_Toc148534565)

[3.7.2 Prediction Button 25](#_Toc148534566)

[3.7.3 Clear Labels and Instructions 25](#_Toc148534567)

[3.8 Model Integration 25](#_Toc148534568)

[3.8.1 Serialization 25](#_Toc148534569)

[3.8.2 Model Loading 26](#_Toc148534570)

[3.8.3 Prediction Logic 26](#_Toc148534571)

[3.9 Testing and Validation 26](#_Toc148534572)

[3.10 Conclusion 27](#_Toc148534573)

[CHAPTER FOUR: DISCUSSION AND ANALYSIS OF RESULTS 28](#_Toc148534574)

[4.1 Introduction 28](#_Toc148534575)

[4.2 The dataset 28](#_Toc148534576)

[4.3 Visualization and Interpretation of the dataset 29](#_Toc148534577)

[4.3.1 Gender prone to Heart Attack 29](#_Toc148534578)

[4.3.2 Age range prone to the heart attack 29](#_Toc148534579)

[4.3.3 possibility of heart attack based on heart rate 30](#_Toc148534580)

[4.3.4 Correlation Heatmap 30](#_Toc148534581)

[4.4 Accuracy of the SVM Model 31](#_Toc148534582)

[4.5 Integration with GUI 31](#_Toc148534583)

[4.5.1 User-Friendly Interface 31](#_Toc148534584)

[4.5.2 Input Validation 31](#_Toc148534585)

[4.5.3 Prediction Results 31](#_Toc148534586)

[4.4 Conclusion 32](#_Toc148534587)

[CHAPTER FIVE: RECOMMENDATIONS AND CONCLUSSIONS 33](#_Toc148534588)

[5.1 Conclusion 33](#_Toc148534589)

[5.1.1 Key Findings 33](#_Toc148534590)

[5.1.2 Contributions to Knowledge 33](#_Toc148534591)

[5.2 Recommendations 33](#_Toc148534592)

[5.3 Final Thoughts 34](#_Toc148534593)

[REFERENCES 35](#_Toc148534594)

[APPENDICES 38](#_Toc148534595)

## Abstract

Cardiovascular diseases are the most common cause of death worldwide over the last few decades in the developed as well as underdeveloped and developing countries. Early detection of cardiac diseases and continuous supervision of clinicians can reduce the mortality rate. However, accurate detection of heart diseases in all cases and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise. Here, a tentative design of a heart disease prediction system has been proposed to detect impending heart disease using Machine learning techniques.

In the pursuit of advancing healthcare outcomes and addressing the global burden of cardiovascular diseases, we introduce a pioneering machine learning system using the Support Vector machine classifier (SVC) designed for the early detection of heart attacks. This system represents a significant leap forward in cardiac care, offering the potential to transform the way we approach the diagnosis and treatment of this life-threatening condition. Our machine learning system is driven by a clear set of objectives, each aimed at optimizing its performance and impact. These objectives encompass the enhancement of prediction accuracy, the seamless integration of data collection processes, advanced feature engineering techniques, ethical compliance, scalable deployment, model interpretability, and a rigorous evaluation of real-world impact.

## Keywords and Abbreviations

* Heart attack - A medical emergency that occurs when blood flow to the heart muscle is blocked, often due to a blood clot, resulting in damage to the heart muscle.
* Heart cycle - The sequence of events in the cardiac system, including atrial and ventricular contractions, responsible for pumping blood throughout the body.
* ECG (Electrocardiogram) - A medical test that records the electrical activity of the heart, typically displayed as a graphical representation of heart rhythms to diagnose cardiac abnormalities.
* CAD - coronary artery disease
* CHF - Chronic heart failure
* ML – Machine Learning
* CHF - Congestive heart failure

## Dedication

We dedicate this work to our friends, lecturers, and parents whose guidance and wisdom have been the compass of this endeavour. Your relentless support and insightful feedback have made this work possible. It is with heartfelt gratitude that I acknowledge your pivotal roles in shaping the outcome.

# CHAPTER 1

## 1.0 Introduction

In the relentless pursuit of improved healthcare outcomes and the mitigation of cardiovascular diseases (including heart attacks), the development of advanced technology has played a pivotal role. Among these advancements, machine learning has emerged as a potent tool, holding the promise of transforming the landscape of cardiac care. Here we delve into the development of a machine learning system designed for the early detection of heart attacks, a mission with profound implications for patient well-being and healthcare efficiency.

## 1.1 Background

The heart is a kind of muscular organ which pumps blood into the body and is the central part of the body’s cardiovascular system which also contains lungs. Cardiovascular system also comprises a network of blood vessels, for example, veins, arteries, and capillaries. These blood vessels deliver blood all over the body. Abnormalities in normal blood flow from the heart cause several types of heart diseases which are commonly known as cardiovascular diseases (CVD). Heart diseases are the main reasons for death worldwide. According to the survey of the World Health Organization (WHO), 17.5 million total global deaths occur because of heart attacks and strokes. More than 75% of deaths from cardiovascular diseases occur mostly in middle-income and low-income countries. Also, 80% of the deaths that occur due to CVDs are because of stroke and heart attack Hazra (2017). Therefore, detection of cardiac abnormalities at the early stage and tools for the prediction of heart diseases can save a lot of life and help doctors to design an effective treatment plan which ultimately reduces the mortality rate due to cardiovascular diseases. Due to the development of advance healthcare systems, lots of patient data are nowadays available (i.e. Big Data in Electronic Health Record System) which can be used for designing predictive models for Cardiovascular diseases. Data mining or machine learning is a discovery method for analyzing big data from an assorted perspective and encapsulating it into useful information. “Data Mining is a non-trivial extraction of implicit, previously unknown and potentially useful information about data” Patel (2016). Nowadays, a huge amount of data pertaining to disease diagnosis, patients etc. are generated by healthcare industries. Data mining provides a number of techniques which discover hidden patterns or similarities from data. Therefore, in this paper, a machine learning algorithm is proposed for the implementation of a heart disease prediction system which was validated on two open access heart disease prediction datasets.

Electrocardiogram (ECG) is employed to diagnose CVD. However, visually identifying long-term ECG abnormalities takes time and effort. With the advent of machine learning (ML) applications in the medical domain, many researchers and practitioners found machine learning-based heart disease diagnosis (MLBHDD) systems as cheap and flexible approaches Benhar (2015). As a consequence, several studies proposed MLBHDD using different heart disease datasets (Rajesh) For instance, Bashir et al. used various machine learning approaches such as Naive Bayes (NB), Decision Tree (DT) based on Gini index, DT based information gain, instance-based learner, and Support Vector Machines (SVM) to develop an ensemble-based model to focus on prediction and analysis of heart patients and achieved an accuracy of 87.37% Sellami (2019). Awad (2017). presented an ML-based Myocardial infarction (MI) prediction model using J48 algorithms and reported 82.57% accuracy. Recently, Deep Learning (DL) added an additional layer and demonstrated the benefit of developing data-driven heart disease diagnosis approaches with an accuracy close to 100%. A Convolutional Neural Network (CNN) based coronary heart disease diagnosis model Romdhane (2020). Deep Neural Network (DNN) based model, named CraftNet by Li et al. are some of the examples out of many proposed DL based heart disease diagnosis model.

## 1.2 Statement of the problem

Heart disease, including myocardial infarction (commonly known as a heart attack), remains a leading global health concern, responsible for a substantial number of fatalities and a significant burden on healthcare systems. Despite advancements in medical science and technology, early detection of heart attacks remains a critical challenge, as timely intervention is pivotal in saving lives and minimizing long-term health complications.

Hasan and Bao (2020) carried out a study to identify the most efficient feature selection approach for anticipating cardiovascular illness. The findings demonstrated that the most accurate prediction results for cardiovascular illness were provided by the XGBoost classifier coupled with the wrapper technique. XGBoost delivered an accuracy of 73.74%, followed by SVC with 73.18% and ANN with 73.20%. Jagtap (2018) developed a web-based application for heart disease prediction using machine learning techniques. He used SVM but obtained an accuracy of 64.4%. The two SVM accuracies are so low as a result they may lead to inaccurate results.

We are coming in to develop a machine learning algorithm using the SVM with a higher accuracy for better results. By leveraging the power of data analytics and pattern recognition, machine learning models can potentially detect subtle indicators of impending heart attacks well before they manifest clinically. This model can analyse a wide range of patient data, including medical records, vital signs, and lifestyle factors, to predict the risk of a heart attack and enable proactive medical intervention.

By addressing this problem, we aim to contribute to the growing body of knowledge and best practices in leveraging machine learning for improving cardiovascular health outcomes, ultimately working towards a future where the timely identification of heart attacks becomes a reality for more individuals, leading to increased survival rates and improved quality of life.

## 1.3 Objectives

### 1.3.1 General objective

1. To develop an effective solution for the early detection of heart attacks using machine learning.

### 1.3.2 Specific objectives

1. To find out the early Major signs and symptoms of heart attacks.
2. To develop a Graphical User Interface to help the users in visualizing the results of the model.
3. To test and validate the developed model and the GUI.
4. To deploy the developed model and GUI.

## 1.4 Scope of Study

Early diagnosis of heart attacks can be achieved by applying machine learning methods to both current and historical patient data. This approach allows individuals to receive prompt and accurate diagnoses without the need for invasive tests or extensive screenings.

We focus on leveraging machine learning techniques to develop a robust and effective system for heart attack detection. By analysing a comprehensive dataset of patient information, including medical records, vital signs, and lifestyle factors, we aim to create a predictive model that can identify individuals at risk of a heart attack well in advance of symptom manifestation.

## 1.5 Justification

The development of this machine learning system is motivated by the pressing need for early detection of heart attacks, a condition where timely intervention can make a profound difference in patient outcomes. Traditional diagnostic methods often fall short in identifying at-risk individuals, leading to delayed treatment and increased mortality rates. By leveraging the power of data-driven insights and predictive analytics, our system aims to bridge this gap, ultimately saving lives and reducing the societal and economic burden associated with cardiovascular diseases.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Introduction

The human heart's electrical activity is recorded using an ECG, which involves attaching electrodes to the skin and produces various waveforms. This non-invasive technique helps assess heartbeat, heart rate, and overall cardiac health, crucial for identifying heart disease. Human cells do not directly connect with the external environment; instead, they rely on the cardiovascular system for transportation. The cardiovascular system comprises two fluid types: blood, which circulates through blood vessels and the heart, and lymph, involving lymph nodes and lymphatic vessels. The cardiovascular system can be seen as a combination of the vascular and lymphatic systems. In the context of the heartbeat, a heart cycle represents a sequence of actions Rahman (2019). Typically, a heart cycle involves the contraction of both atria, followed by the synchronized contraction of each ventricle a fraction of a second later. The heart is primarily composed of interconnected heart muscle cells, and when one of these cells contracts, it triggers neighbouring cells to become excited. During the cardiac cycle, the heart muscles also experience periods of rest between beats, which plays a crucial role in facilitating aerobic respiration.

## 2.2 Related Works

The application of artificial intelligence in cardiac disease detection, has shown significant improvements compared to existing widely used models, such as those provided by the American College of Cardiology/American Heart Association (ACC/AHA) models in cardiovascular disease (CVD) detection and prediction Weng (2017). Singh (2016) designed a heart disease prediction system based on Structural Equation Modelling (SEM) and Fuzzy Cognitive Map (FCM). They validated the data of the Canadian Community Health Survey (CCHS) 2012 dataset. They used twenty significant attributes. To generate the weight matrix for the FCM model, SEM was used which then predicted a possibility of cardiovascular diseases. An SEM model is defined with the correlation between CCC 121 along with 20 attributes; here CCC 121 is a variable which defines whether the respondent has heart disease. Ghadge (2016) researched an intelligent heart attack prediction system using big data. Heart attack needs to be diagnosed timely and effectively because of its high prevalence. The main objective of this research article was to find a prototype of an intelligent heart attack prediction system that uses big data and data mining modelling techniques. This system could gather hidden knowledge concerning heart disease from a given heart disease database.

Hasan and Bao (2020) conducted a study aimed at determining the most effective feature selection method for predicting cardiovascular disease by comparing multiple algorithms. They initially considered three well-known feature selection techniques: filter, wrapper, and embedding. Subsequently, a feature subset was derived from these methods through a Boolean process that identified common "True" conditions. This approach involved two stages to retrieve feature subsets. To evaluate and identify the most accurate predictive analytics, several models, including random forest, support vector classifier, k-nearest neighbours, naive Bayes, and XGBoost, were examined. An artificial neural network (ANN) served as a standard for comparison using all available features. The results revealed that the XGBoost classifier combined with the wrapper technique yielded the most accurate predictions for cardiovascular disease, achieving an accuracy rate of 73.74%. It was followed by the support vector classifier (SVC) with 73.18% accuracy and ANN with 73.20% accuracy.

Zhao, Nakahira (2018) analysed the possibilities and related aspects of enhancing human health management systems. They also outlined a research direction for medical technology in the context of the Internet of Things (IoT). They examined various health-related sensors and technologies, identifying key issues that require resolution. Chiuchisan, Geman (2014) designed a home monitoring system and decision support system, which proved beneficial for home-based monitoring, diagnosis, medical prescriptions, medical treatment, rehabilitation, and the overall care of patients with Parkinson's disease. Over the past decade, there has been considerable interest from both the research community and the industry in the Wireless Health Monitoring System (WHMS). Notably, there have been improvements in the performance of several machine learning algorithms and classifiers, such as the weighted associative classifier, for the detection of cardiac abnormalities Sharma (2011).

Sonawane (2017) introduced a real-time mobile healthcare system for monitoring elderly patients, both indoors and outdoors. The system primarily comprised a signal sensor and a smartphone. Bio-signal data from the sensor was transmitted to an intelligent server via the GPRS/UMTS network for data collection. This system was capable of monitoring the mobility, vital signs, location, and overall condition of elderly patients from a remote location. M.R (2010) proposed a fully functional wireless body area network (WBAN) system. This system utilized medical bands to gather physiological data from various sensors. The author carefully selected specific medical bands to minimize interference between the sensors and other existing devices. To extend the operational range, the multi-hopping technique was implemented, and a medical gateway wireless board was incorporated for this purpose.

Amin (2019) conducted a Research to classify the most relevant attributes of heart disease prediction. Seven classification algorithms are used, which consist of NB, KNN, LR, DT, NN, SVM, and Vote. The Cleveland datasets were obtained from the UCI repository of machine learning, which consists of 303 records and 76 attributes. The 10-fold cross-validation method is used for model training and testing. We used 10-fold cross-validation because, in the dataset, we have fewer training examples, and using data split, such as train-test split, will give us an underestimate of the model predictive performance because we will have fewer number of examples in the training set. However, using 10-fold validation, the model will have 90% of the data to learn from. The Vote Classifier achieved a higher accuracy of 87.4%.

Below is a table depicting each and every feature that he used;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
| S. no. | Feature name | Feature code | Description | Domain of values (min-max) |
|  | | | | |
| 1 | Age | AGE | Age in years | 30 < age < 77 |
| 2 | Sex | SEX | Male = 1 | 1 |
| Female = 0 | 0 |
| 3 | Type of chest pain | CPT | 1 = atypical angina | 1 |
| 2 = typical angina | 2 |
| 3 = asymptomatic | 3 |
| 4 = nonanginal pain | 4 |
| 4 | Resting blood pressure | RBP | mm Hg admitted at the hospital | 94–200 |
| 5 | Serum cholesterol | SCH | In mg/dl | 120–564 |
| 6 | Fasting blood sugar >120 mg/dl | FBS | Fasting blood sugar >120 mg/dl (1 = true; 0 = false) | 1 |
| 0 |
| 7 | Resting electrocardiographic results | RES | 0 = normal | 0 |
| 1 = having ST-T | 1 |
| 2 = hypertrophy | 2 |
| 8 | Maximum heart rate achieved | MHR | — | 71–202 |
| 9 | Exercise-induced angina | EIA | 1 = yes | 0 |
| 0 = no | 1 |
| 10 | Old peak = ST depression induced by exercise relative to rest | OPK | — | 0–6.2 |
| 11 | Slope of the peak exercise ST segment | PES | 1 = up sloping | 1 |
| 2 = flat | 2 |
| 3 = down sloping | 3 |
| 12 | Number of major vessels (0–3) coloured by fluoroscopy | VCA | — | 0 |
| 1 |
| 2 |
| 3 |
| 13 | Thallium scan | THA | 3 = normal | 3 |
| 6 = fixed defect | 6 |
| 7 = reversible defect | 7 |

Enriko (2016) used a KNN classifier with minimal parameters for heart disease prediction and had an accuracy rate of 81.85%. When using KNN, the performance drops as the number of parameters increases, and it uses 90% of the input for training, which is computationally expensive. Subhadra (2019) The used training algorithm is a multilayer perceptron neural network (MLP-NN) with backpropagation for heart disease prediction. To evaluate the system’s performance, recall, accuracy, precision, and F-measure are employed, and model training and testing are carried out using the UCI repository of machine learning Cleveland dataset, which consists of the records of 303 instances and has 76 attributes. Through preprocessing, missing values were removed from the data, which consisted of six records, and the 14 most relevant attributes of the heart disease were used. The results generated during the experiment showed that MLN-NN obtained a higher accuracy of 93.39%, with a running time of 3.86 seconds.

Khan (2019) used a comprehensive prediction of heart disease based on an analysis using some of the most popular machine learning classifiers. For training and testing, only 14 features are employed from the Cleveland (UCI) datasets, which consist of 303 records. There was a data preprocessing activity carried out, resulting in a dataset consisting of 296 records. The results of SVM classifiers achieved a higher accuracy of 90.00%. Tarawneh (2019) have conducted a study using the hybrid approaches of data mining classifiers to predict heart disease. The datasets were obtained from the UCI repository of machine learning, which consists of 303 records and has 76 attributes. Model training and testing were performed on 14 attributes. The data was pre-processed to minimize the features from 14 to 12. KNN, NN, SVM, GA, J48, RF, and NB are the classification algorithms used to assess the precision, recall, and accuracy of cardiac disease prediction. The accuracy obtained by SVM and NB was 89.2%, and they made better predictions of heart disease.

Anitha (2019) have conducted a study using learning vector quantization algorithms for the prediction of cardiac disease. The accuracy achieved by this algorithm is 85.55%. The datasets were taken from the University of California, Irvine’s (UCI), machine learning library, which consists of 303 records and has 76 attributes. The data were pre-processed because of missing values, resulting in a sample of 302 records, with only 14 features used for heart disease. The dataset is categorized into two sections: 70% for model training and 30% for model testing. Jagtap (2018) developed a web-based application for heart disease prediction using machine learning techniques. For the classification algorithms, LR, NB, and SVM are used for model training and testing. Using the UCI machine learning repository, the Cleveland datasets were divided into 75 percent and 25 percent for training and testing, respectively. The data were pre-processed to eliminate discrepancies and missing values, and SVM achieved a higher accuracy of 64.4%. The study’s limitation was its inability to detect the risk factors of human heart disease patients at an early stage. Dulhare (2018) in his study combined the common feature selection algorithms of particle swarm optimization (PSO) and Naïve Bayesian algorithms for an efficient prediction of heart disease. The model training and testing processes were conducted using the UCI repository of the machine learning dataset of VA Long Beach, which consists of 270 records and 14 attributes, however, only 7 attributes out of the 14 attributes of heart disease were used to predict it. When combined with PSO and NB, the performance accuracy of NB increases to 87.91%. It has been shown that accuracy improves by 8.79% as compared to NB accuracy. Kim (2017) studied using machine learning algorithms to predict heart disease. The datasets were collected from the repository of machine learning at the University of California, Irvine (UCI), which consists of 303 records and uses 14 attributes. For training and testing, the 10-fold cross-validation approach is utilized. The DT algorithm performs with a better accuracy of 93.19% prediction of heart disease.

Siontis (2021) describes the present and future condition of AI-enhanced electrocardiogram (ECG) in the diagnosis of heart disease in at-risk communities, summarize its consequences for healthcare decisions in patients with cardiovascular disease, and assess its potential drawbacks. Linda (2021) proposed a unique health information system for prescribing exercise to heart disease patients. According to their early findings, clinicians are confused about how to establish an exercise prescription for patients with numerous CVD risk factors. For patients, the supplied system is an easy-to-use, guided, and time-saving evidence-based method. Ali (2021) provided a three-phase PB-FARM approach for the assessment of disease-related risk factors. It was also used to analyze the factors that influence the incidence of this disease using the Z-Alizadeh Sani dataset. The findings revealed a clear link between the risk of coronary artery disease (CAD), elderly age, and normal chest pain. Rubini (2021) proposed a prediction model for heart disease prediction. Different classifiers, such as logistic regression, Nave Bayes, and SVM, were compared to the proposed algorithm. In the proposed article, random forest achieved the highest accuracy of 84.81%. Devansh (2020) utilized a dataset of 303 examples and 76 attributes, 14 of which were used in supervised learning algorithms, such as decision tree, Nave Bayes, K-NN, and random forest. The results show that K-NN has attained the maximum level of accuracy.

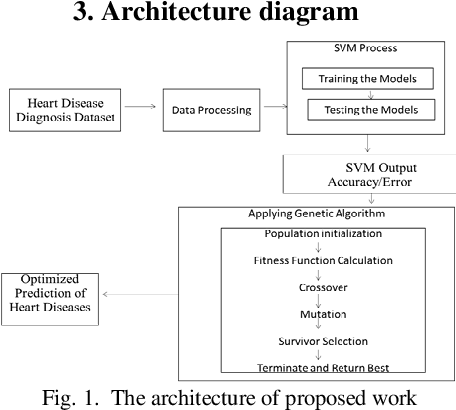
Archana (2020) developed a heart disease prediction model using machine learning classifiers. The UCI, Cleveland dataset uses 14 attributes to train and test their models to achieve maximum accuracy. The results achieved by the classifiers were as follows: linear regression 78%, decision trees 79%, support vector machines 79%, and K-NN 87%. The results revealed that K-NN had the highest accuracy. Khan (2020) uses SVM, logistic regression, artificial neural networks, KNN, Nave Bayes, and decision tree as classification techniques. When compared to previous models, the new model achieved an accuracy of 92.37%. Mohan (2019) studied how to uncover suitable features using machine learning techniques, such as decision trees, language models, SVM, random forests, Naive Bayes, neural networks, and KNN. The proposed hybrid HRFLM method was applied to merge the characteristics of random forests and linear techniques. This model’s accuracy was 88.4%. Kumar (2020) various machine learning algorithms were utilized to predict cardiovascular disease. When compared to other classifier techniques, the proposed model revealed that random forests had the greatest accuracy of 85.71%.

Summary table of related works is as below;

|  |  |  |  |
| --- | --- | --- | --- |
| Developer | System Developed | Algorithms Used | Accuracy (%) |
| Hasan and Bao (2020) | Feature selection method for predicting cardiovascular disease | XGBOOST  SVC  ANN | 73.74  73.18  73.20 |
| Enriko (2016) | Heart Disease prediction | KNN | 81.85 |
| Amin (2019) | Classification of most relevant attributes of heart disease prediction | NB, KNN, LR, DT, NN, SVM  VOTE | 87.4 |
| Subhadra (2019) | Backpropagation for Heart Disease prediction | MLP-NN | 93.39 |
| Jagtap (2018) | Web-based application for heart disease prediction using machine learning | SVM | 64.4 |
| Archana (2020) | Heart disease prediction model | SVM  KNN  Decision Trees  Linear Regression | 79  87  79  78 |

## 2.3 Proposed system

This work aims in developing a Machine learning System in heart disease detection that uses the data mining technique having best accuracy and performance, Support Vector Machine classifier is used in this case. By using several cardiovascular system parameters such as age, blood pressure, ECG results, sex, and blood sugar, it is possible to measure the possibility of getting affected by heart disease Turner, Stocker (2015). This algorithm takes the medical parameters such as age, blood pressure, heartbeat, sex, ECG results, blood sugar etc. as input and shows the probability of getting affected by heart disease as output. This proposed system comprises the scheme and design of a web-based application which uses an efficient machine learning technique to detect heart disease. It can serve as a very useful tool for doctors, patients, and medical students to diagnose heart disease. For diagnosis of fatal physiological conditions and symptoms such as heart attack requires 24 hours monitoring of patient’s health after transferring from hospital to home. Using this application, the patient can input the current parameters of heart disease from anywhere on the application interface and view the risk level of getting heart disease. All of the parameters that are not real-time like blood sugar, Serum cholesterol, ECG results will be available in the doctor’s prescribed report and there are some parameters like chest pain type and exercise-induced angina, which have to be self-measured periodically by the patient and input the values manually on the application interface. If any fatal situation is detected by the application, the patient can communicate with any doctor via video call and any registered doctor related to heart disease can be found by putting his phone number in the search bar.



## 2.4 Challenges with the existing systems

1. **Low Accuracy:** The SVM models used are of low accuracy i.e. 64.4%, 73.18%, and 79% which in return may give inaccurate results**.**
2. **Increased Complexity:** Managing and coordinating multiple models is more complex than working with a single model. Each model may have different hyperparameters and training requirements, leading to added complexity in the development and maintenance process.
3. **Overfitting:** Many systems as discussed above have employed the use of many models in testing and training the data. The more the number of models the more the Risk of Overfitting.
4. **Overhead:** Many developed systems have used multiple models, which in return have increase overhead in terms memory, storage and computational resources.

# CHAPTER 3: METHODOLOGY

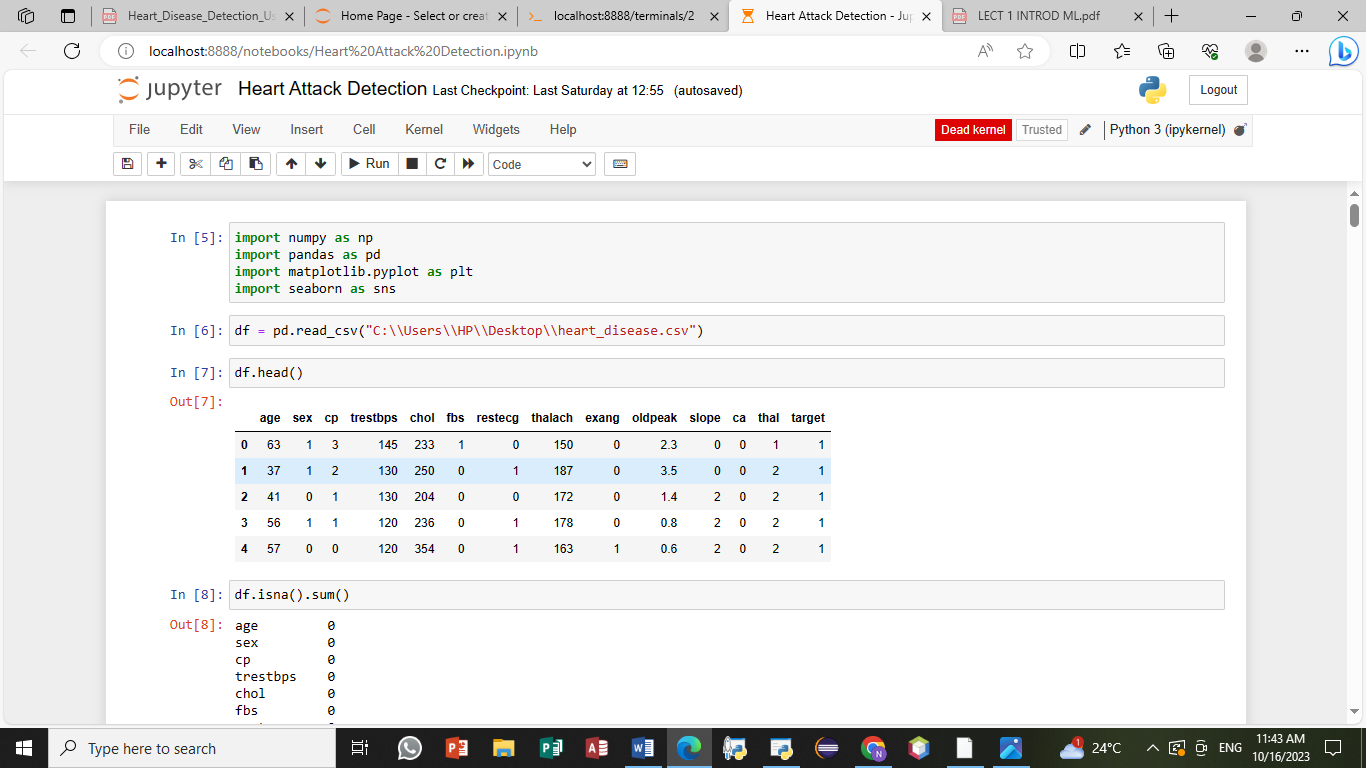
## 3.0 Introduction

In this chapter, we embark on the journey of developing a machine learning model for heart attack detection. We will begin by detailing our experimental approach, clearly defining the measurement metrics used for evaluation. Subsequently, we will provide insights into the dataset's location, its composition, and the strategy employed for splitting it into training and testing sets. Following this, we will delve into the critical steps of data preparation, highlighting the use of data pre-processing tools. The Support Vector Classifier (SVC) will be introduced as the chosen algorithm for this project, followed by an in-depth discussion on model development, hyperparameters tuning, and the comprehensive evaluation of the SVC-based heart attack detection model. Finally, we will emphasize the importance of model validation.

## 3.1 Dataset Description

### 3.1.1 Dataset Location

Our dataset for heart attack detection is sourced from Kaggle site provides various datasets for building and training machine learning models. It is publicly available and can be accessed through the official Kaggle website at (<https://www:Kaggle.com>).



### 3.1.2 Dataset Definition

This dataset is a comprehensive collection of features that encompasses patients' medical histories, demographic information, and clinical attributes. It includes both numerical and categorical data, covering aspects such as age, gender, cholesterol levels, blood pressure readings, and various other factors relevant to heart health. The target variable is binary, indicating whether a patient has experienced a heart attack (1) or not (0).

The table below clearly defines the attributes in the dataset;

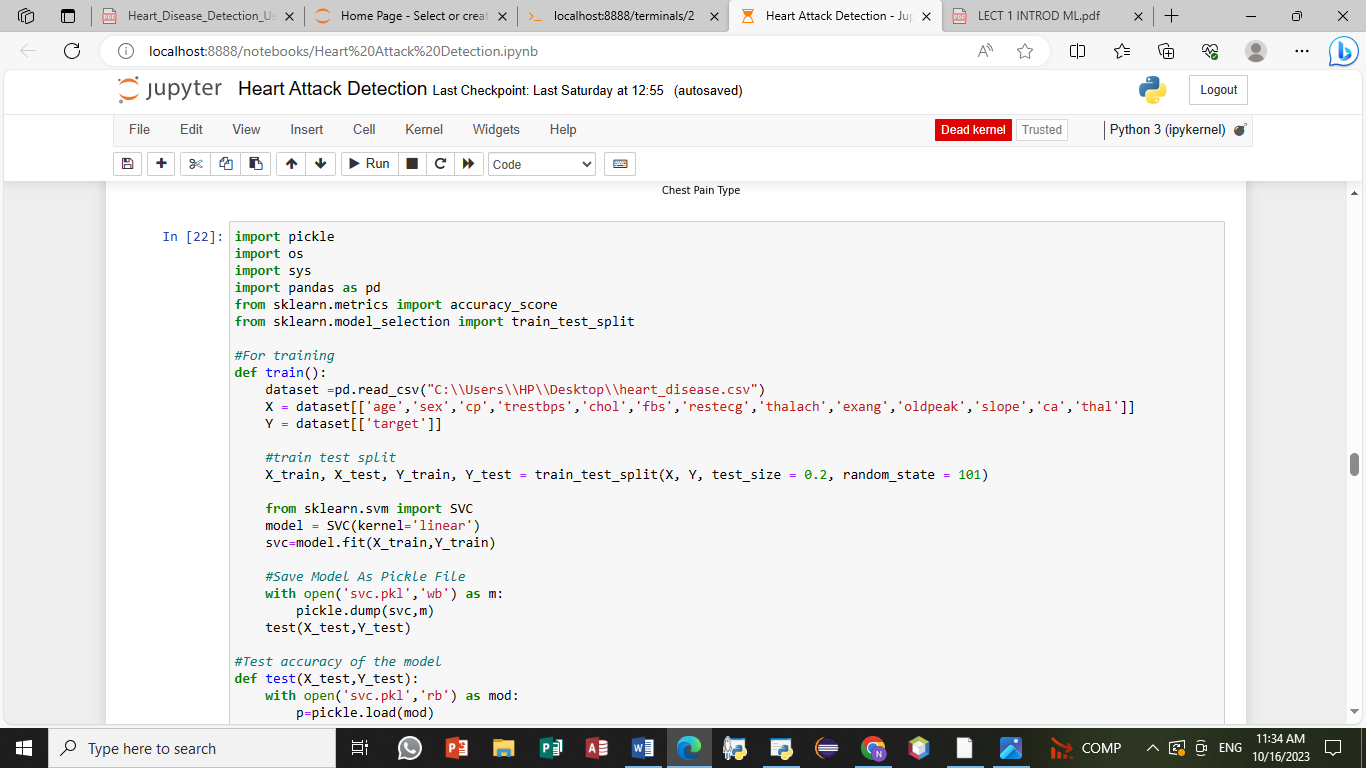
|  |  |  |
| --- | --- | --- |
| No | Attribute | Description |
| 1 | Age | Age in years |
| 2 | Sex | Male=1, Female=0 |
| 3 | Cp | Chest pain type (typical angina=1, atypical angina=2, non-anginal pain=3, asymptomatyic=4) |
| 4 | Trestbps | Resting blood sugar (in mm Hg in case of admission to hospital) |
| 5 | Chol | Serum Cholestral in mg/dl |
| 6 | Fbs | Fasting blood sugar>120 mg/dl (true=1, false=0) |
| 7 | Restecg | Resting electrocardiographic results (normal=0, Having ST-T wave abnormality=1, left ventricular hypertrophy=2) |
| 8 | Thalach | Maximum heart rate |
| 9 | Exang | Exercise-induced Angina |
| 10 | Old peak | ST depression induced by exercise comparative to rest |
| 11 | Slope | Slope of the peak Exercise ST segment (upsloping=1, flat=2, down sloping=3) |
| 12 | Ca | Number of major vessels which are coloured by fluoroscopy |
| 13 | Thal | Normal=0, fixed defect=2, reversible defect=3 |

### 3.1.3 Data Splitting

To ensure robust model evaluation, we have meticulously split the dataset into two distinct subsets:

- **Training Set:** This subset constitutes 80% of the dataset and forms the basis for model training and hyperparameters optimization. The training set allows our model to learn underlying patterns and relationships within the data.

- **Testing Set:** The remaining 20% of the dataset is reserved exclusively for testing the model's performance. This separation ensures that our model's generalization to unseen data can be accurately assessed using the defined measurement metrics.



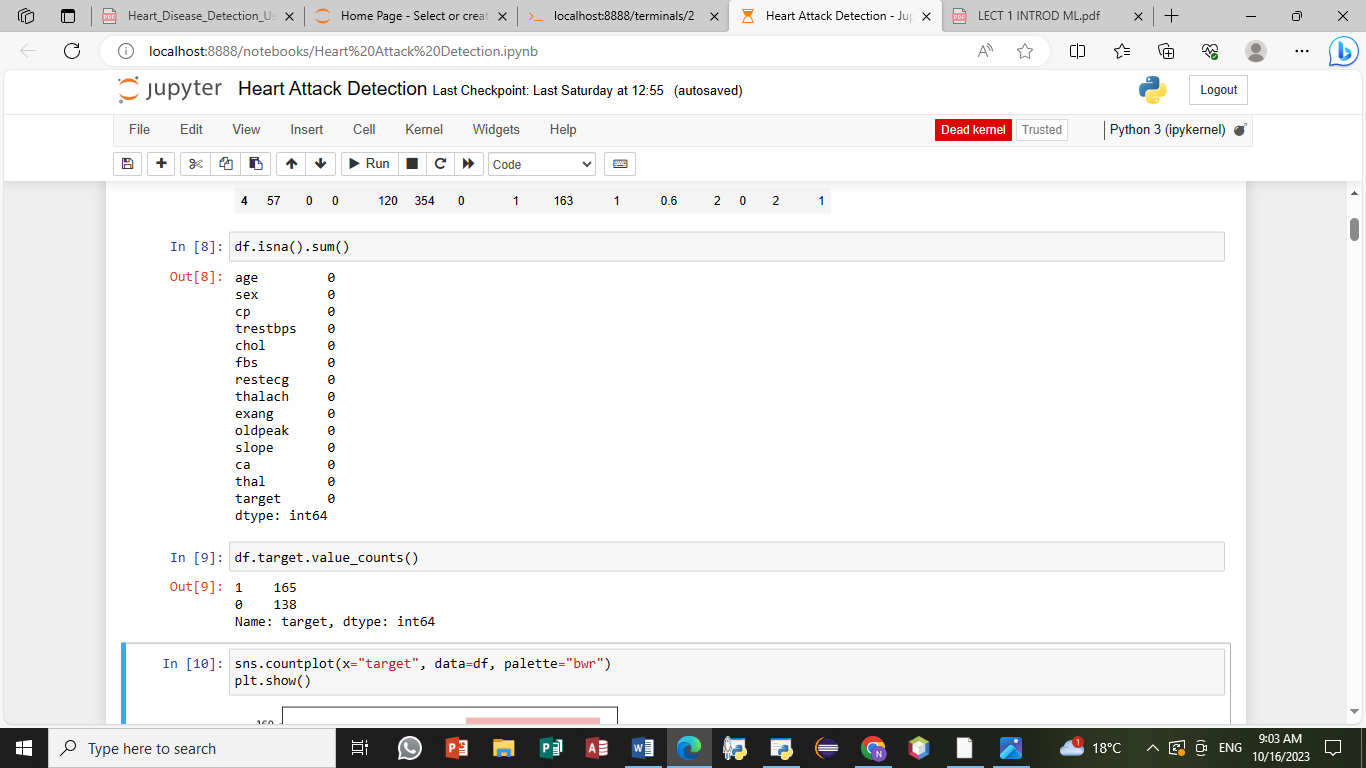
## 3.2 Data Preprocessing

Effective data preparation is pivotal in building a successful machine learning model. We employ various data preprocessing tools and techniques to ensure data quality and model readiness:

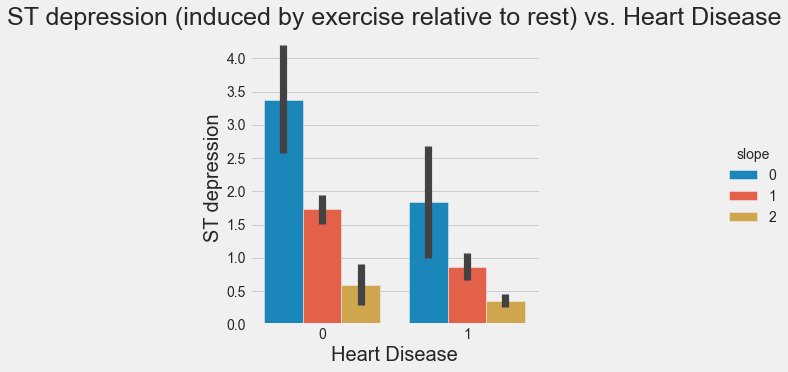
### 3.2.1 Data Cleaning

Raw healthcare data is often noisy and contains missing values, outliers, and inconsistencies. Therefore, data cleaning is a critical step to ensure the quality and reliability of our dataset. In this phase, we will perform the following tasks:

* **Handling missing data**: We have employed techniques such as imputation and removal of records with missing values, depending on the extent of missing data.



* **Outlier detection**: Outliers can significantly impact model performance. We have used statistical methods and domain knowledge to identify and handle outliers appropriately.

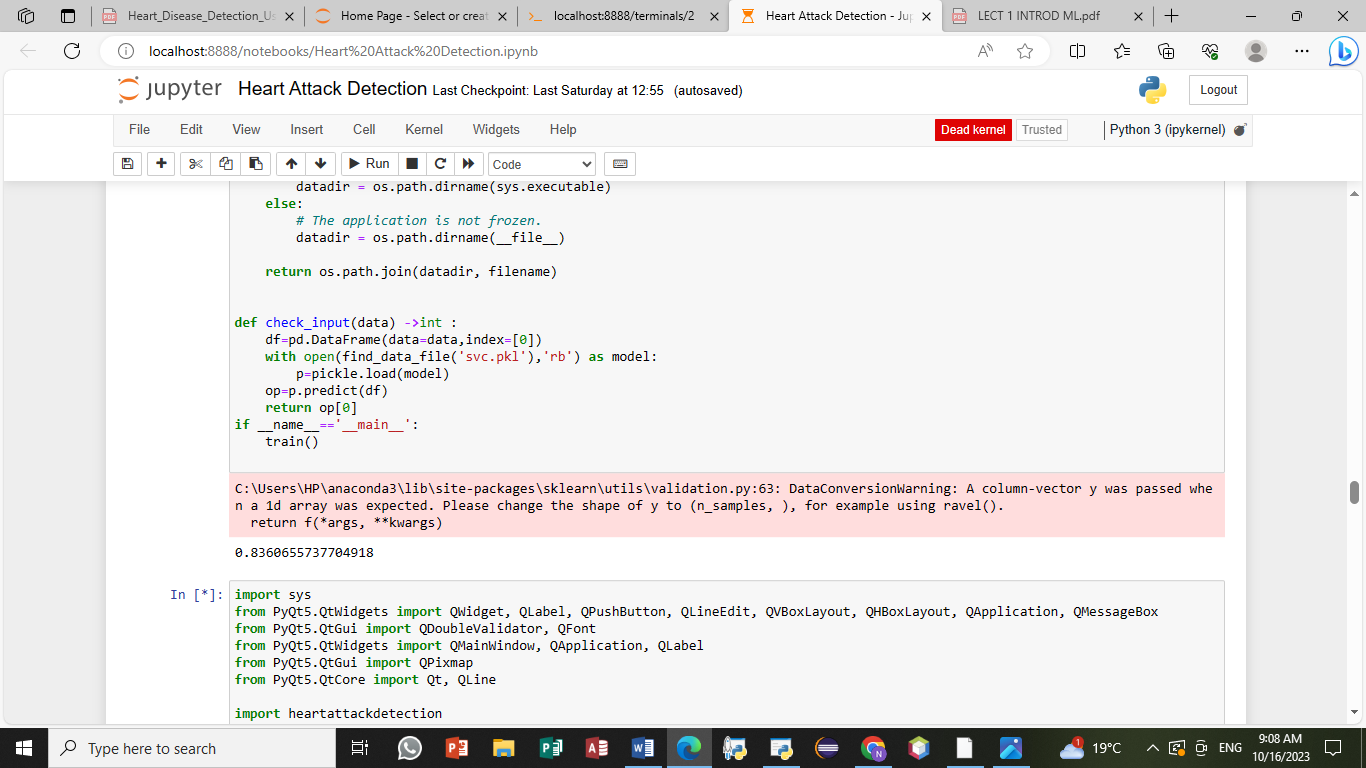


* **Data consistency**: Ensuring that the data adheres to a consistent format and scale is essential for model training. We have standardized and normalized features as needed.

After cleaning the data, we checked to ensure that the data is correct and can be used for prediction. All the attributes were checked to ensure that it doesn’t have missing values or incorrect information. The results obtained are shown below;

## 3.3 Support Vector Classifier (SVC) as the Chosen Algorithm

For the development of our heart attack detection model, we have chosen the Support Vector Classifier (SVC) as the primary algorithm. SVC is known for its ability to handle both linear and non-linear classification tasks effectively.



## 3.4 Feature Engineering and selection

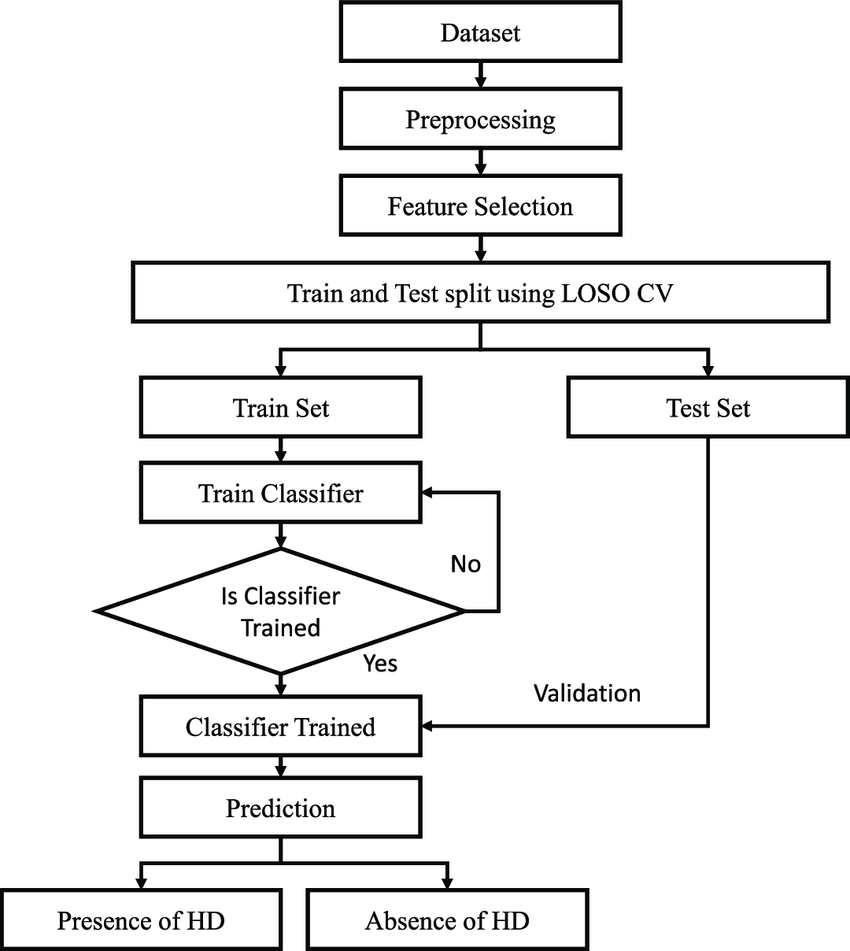
Feature engineering is a crucial step in the process of model development in machine learning. It involves selecting, creating, or transforming the input variables (features) used in your model to improve its predictive performance. The goal of feature engineering is to provide the model with the most relevant and informative input data, which can lead to better accuracy, interpretability, and generalization.

Here are some key aspects of feature engineering and how it is achieved in model development:

1. **Feature Selection:** Feature selection is the process of choosing a subset of the most relevant features from the original set of input variables. This is often done to reduce dimensionality and eliminate noise in the data. Feature selection techniques can include statistical tests, feature importance rankings, and domain knowledge.
2. **Feature Creation:** Sometimes, existing features may not capture the relationship between the input data and the target variable effectively. In such cases, you can create new features by combining or transforming existing ones. For example, you might calculate the body mass index (BMI) from height and weight, which could be a more informative feature in certain health-related predictions.
3. **Normalization and Scaling:** Ensuring that all features have the same scale or distribution can be crucial for some machine learning algorithms. Common techniques include z-score normalization (scaling features to have a mean of 0 and a standard deviation of 1) and min-max scaling (scaling features to a specific range, typically [0, 1]).
4. **Handling Categorical Data:** If your dataset includes categorical features (e.g., "gender" with values "male" and "female"), you need to encode them into numerical values. Common approaches include one-hot encoding (creating binary columns for each category) or label encoding (assigning unique numerical labels to categories).
5. **Dealing with Missing Data:** Missing data can be a common issue in real-world datasets. Depending on the nature of the data and the extent of missing values, you may choose to impute missing values using techniques like mean imputation, median imputation, or more advanced methods such as regression imputation.
6. **Feature Scaling for Algorithms:** Some machine learning algorithms are sensitive to the scale of the input features. For instance, support vector machines and k-nearest neighbours often require feature scaling to perform well. Feature scaling ensures that no single feature dominates the learning process.

Achieving effective feature engineering involves a combination of domain expertise, exploratory data analysis, and experimentation. It requires a deep understanding of the data and the problem you are trying to solve. Properly engineered features can significantly enhance the performance of your machine learning models, making them more accurate and valuable for real-world applications

## 3.5 System flowchart



## 3.6Model Development, Hyperparameter Tuning, and Evaluation

The next phase of our project involves the following steps:

1. Model Development- We have trained the Support Vector Classifier (SVC) on the training dataset, allowing it to learn the patterns and relationships within the data.

2. Hyperparameter Tuning- To optimize the model's performance, we will conduct hyperparameter tuning, experimenting with different parameter configurations and selecting the best-performing ones.

3. Model Evaluation- We have rigorously evaluated the SVC-based heart attack detection model using the defined measurement metrics, including; Accuracy, Recall and Precision

**Accuracy**= *tp+tn*

*n*

**Recal**l= *tp*

*tp+fn*

**Precision**= *tp*

*tp+fp*

where,

n = Total number of instances, tp = true positive, tn = true negative,

fp = false positive, fn = false negative.

4. Model Validation- The final step involves validating the model's effectiveness and generalization on an independent testing dataset. This ensures that our model performs well on unseen data and is ready for deployment.

## 3.7 Building an Intuitive User Interface

The foundation of our machine learning system's user interface is a well-designed form that enables users to input data for prediction. When creating this interface, several considerations come into play:

### 3.7.1 Input Fields

We’ve designed a user-friendly form with input fields corresponding to the features our model relies on for predictions. For instance, in a healthcare context, users may input attributes such as age, gender, blood pressure, and cholesterol levels.

### 3.7.2 Prediction Button

We’ve incorporated a button, labelled "Predict" which users can click to initiate the prediction process.

### 3.7.3 Clear Labels and Instructions

To ensure a seamless user experience, we have provided clear labels for each input field and instructions on how to input data correctly.

## 3.8 Model Integration

To Integrate our trained SVC model into the GUI is a pivotal step. We have performed the following;

### 3.8.1 Serialization

We have Serialized the trained model into a format that can be easily loaded and used within the application. We have employed the Pickle serialization format.

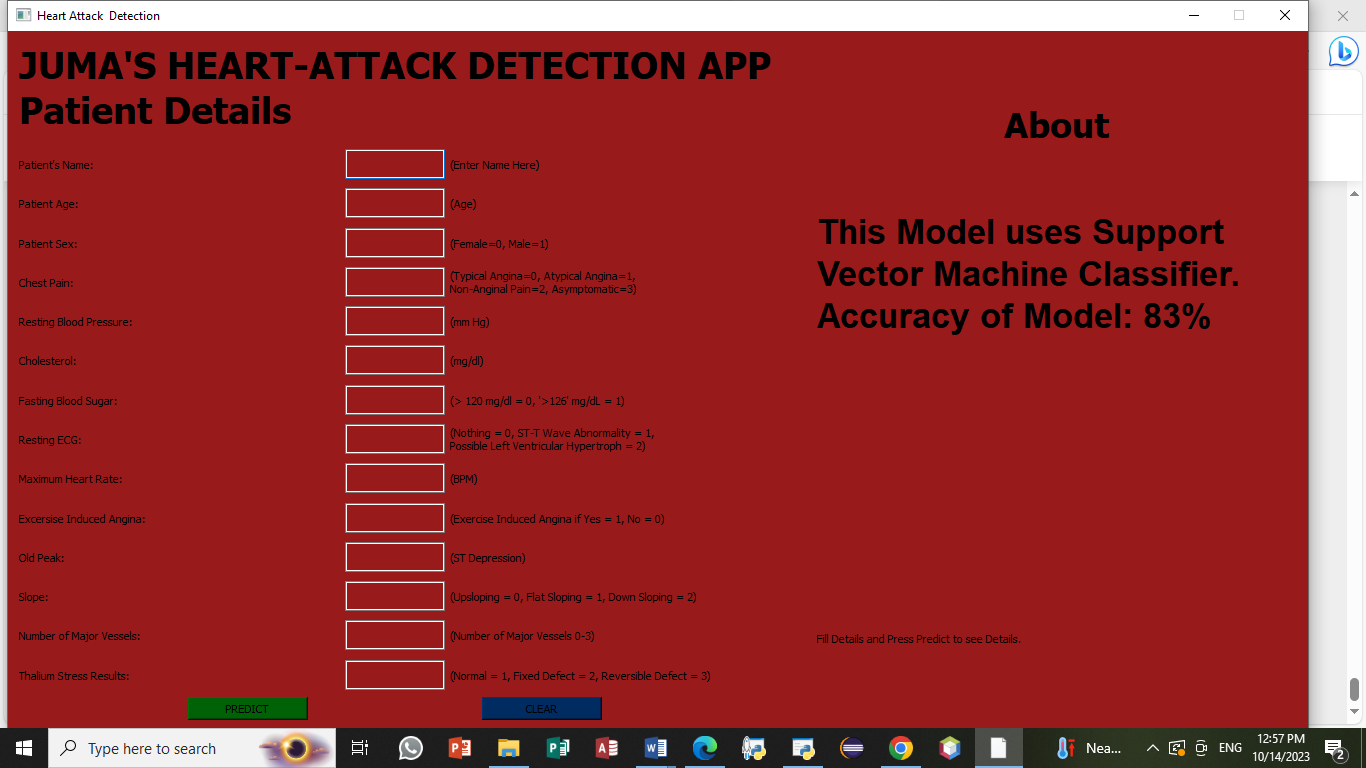
### 3.8.2 Model Loading

When the GUI is launched, we will load the serialized model so that it's ready to make predictions.

### 3.8.3 Prediction Logic

Upon receiving user input via the GUI, we have pre-processed the data as necessary (e.g., scaling or encoding) and pass it through the loaded model to obtain predictions.

Below is the developed user interface developed;



## 3.9 Testing and Validation

Here are some of the steps used to test and validate our model with an accuracy of 83.6%:

1. **Confusion Matrix**: Start by creating a confusion matrix, which breaks down the model's predictions into four categories: true positives, true negatives, false positives, and false negatives. This helps you understand where the model is making errors.
2. **Cross-Validation**: Use cross-validation techniques, such as k-fold cross-validation, to assess how well the model generalizes to new data. This helps you ensure that your model's performance is not just a result of overfitting.
3. **Hyperparameter Tuning**: Revisit the model's hyperparameters to see if you can optimize its performance further. Grid search or random search can be used to find the best hyperparameters.
4. **External Validation**: we have validated our model on an external dataset to check its robustness and performance in different settings.
5. **Deployment Testing**: since our model is for deployment, we have conducted testing in a real-world setting to ensure it performs as expected.

## 3.10 Conclusion

In this chapter, we have achieved all the activities involved in the development of a machine learning model. We've also explored the creation of a user interface that empowers users to interact with our SVC model, inputting values for prediction and exploring the results it generates. By building an accessible and user-friendly GUI, we extend the reach and applicability of our machine learning system

# CHAPTER FOUR: DISCUSSION AND ANALYSIS OF RESULTS

## 4.1 Introduction

In this chapter we indulge ourselves in the discussion of the results that have been achieved during the implementation of this system in the previous chapters. We will discuss various aspects such as visualization results from the dataset, accuracy of the model, integration with the GUI, as discussed below;

## 4.2 The dataset

The dataset used in the development of the model consists of 304 records with each having 13 attributes. The table below shows the 13 attributes in the dataset;

|  |  |  |
| --- | --- | --- |
| No | Attribute | Description |
| 1 | Age | Age in years |
| 2 | Sex | Male=1, Female=0 |
| 3 | Cp | Chest pain type (typical angina=1, atypical angina=2, non-anginal pain=3, asymptomatyic=4) |
| 4 | Trestbps | Resting blood sugar (in mm Hg in case of admission to hospital) |
| 5 | Chol | Serum Cholestral in mg/dl |
| 6 | Fbs | Fasting blood sugar>120 mg/dl (true=1, false=0) |
| 7 | Restecg | Resting electrocardiographic results (normal=0, Having ST-T wave abnormality=1, left ventricular hypertrophy=2) |
| 8 | Thalach | Maximum heart rate |
| 9 | Exang | Exercise-induced Angina |
| 10 | Old peak | ST depression induced by exercise comparative to rest |
| 11 | Slope | Slope of the peak Exercise ST segment (upsloping=1, flat=2, down sloping=3) |
| 12 | Ca | Number of major vessels which are coloured by fluoroscopy |
| 13 | Thal | Normal=0, fixed defect=2, reversible defect=3 |

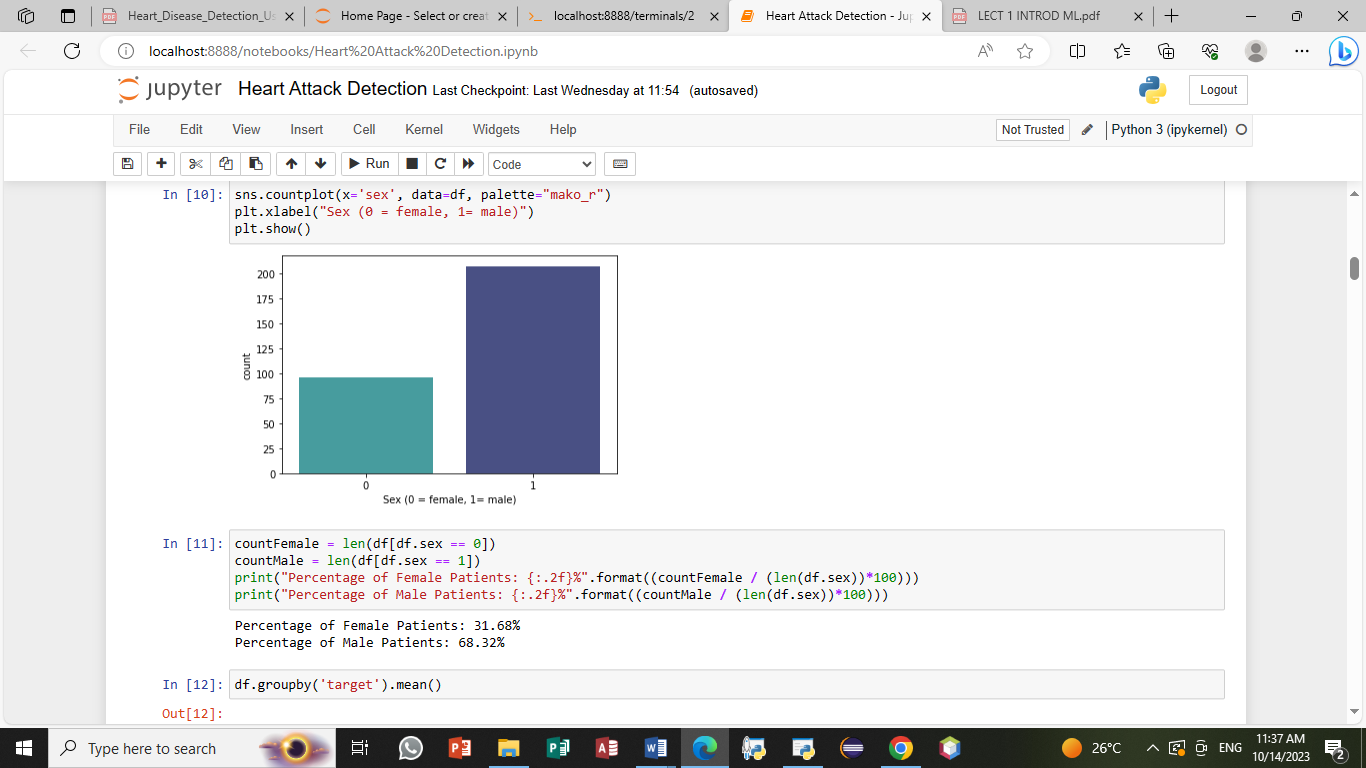
The age is written in numerical; it represents the age of the heart attack suspect. The sex represents the gender of the patient. Label encoding has been used to handle the Gender where 1 has mean been used to represent male and 0 to represent female gender. The chest pain type has also employed the label encoding where typical angina=1, atypical angina =2, non–anginal pain=3 and asymptomatic=4. Trestbps represents the Resting blood sugar in mmHg. Chol represents the Cholesterol level in the body. Fbs represents the Fasting blood sugar where blood sugar greater than 120 is represented by 1 and that which is below 120 is represented by 0. Restecg represents the electrocardiographic results (normal=0, having ST-T wave abnormality=1, left ventricular hypertrophy=2). Thalach represents the maximum heart rate. Exang represents the exercise-induced Angina. Old peak represents the ST depression induced by exercise comparative to rest. Slope represents the slope of the peak Exercise ST segment where (Upsloping=1, flat=2, down sloping=3). Ca represents the number of major vessels which are coloured by fluoroscopy. Finally, Thal has employed label Encoding where Normal=0, fixed Defect =2, reversible defect=3.

## 4.3 Visualization and Interpretation of the dataset

The heart disease dataset contains so main features which can be viewed in terms of graphs and charts to provide easier understanding of the Dataset.

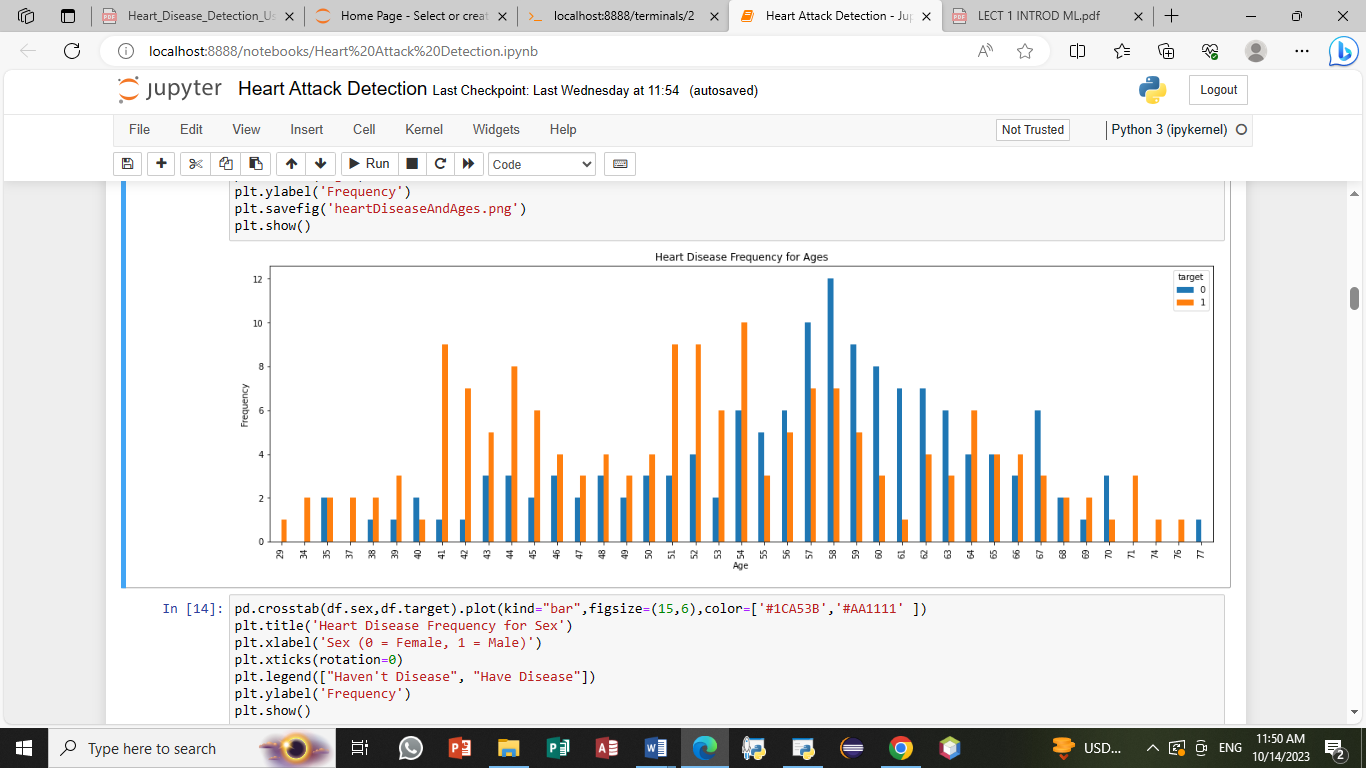
### 4.3.1 Gender prone to Heart Attack

From the dataset we deduce that more male are prone to heart attack as compared to female. The male gender are contributing 68.32% and female constitute of 31.68% as shown below;



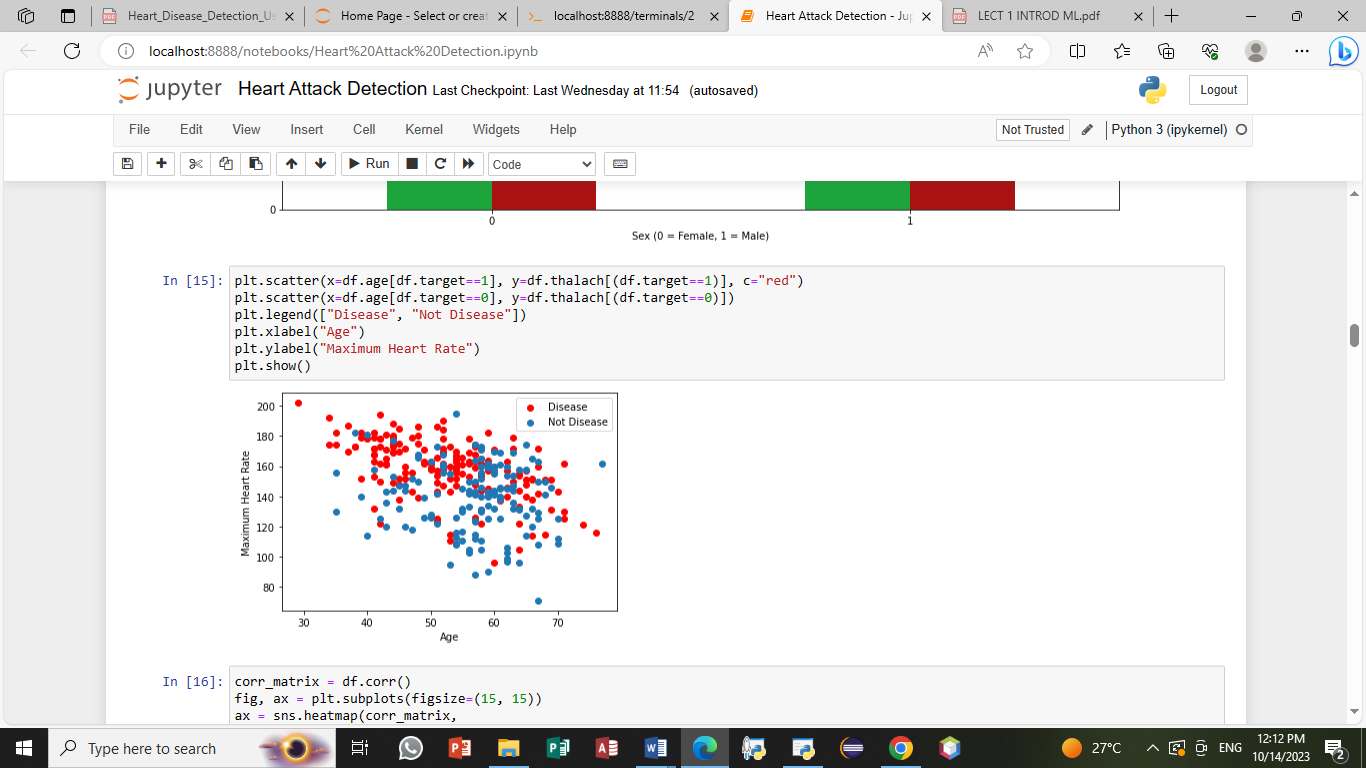
### 4.3.2 Age range prone to the heart attack

We also see clearly that the age bracket of the people affected most by the heart attack ranges from 41 -71 years. And the age that is so risky to be affected by the heart attack is 59 years, as illustrated in the diagram below;



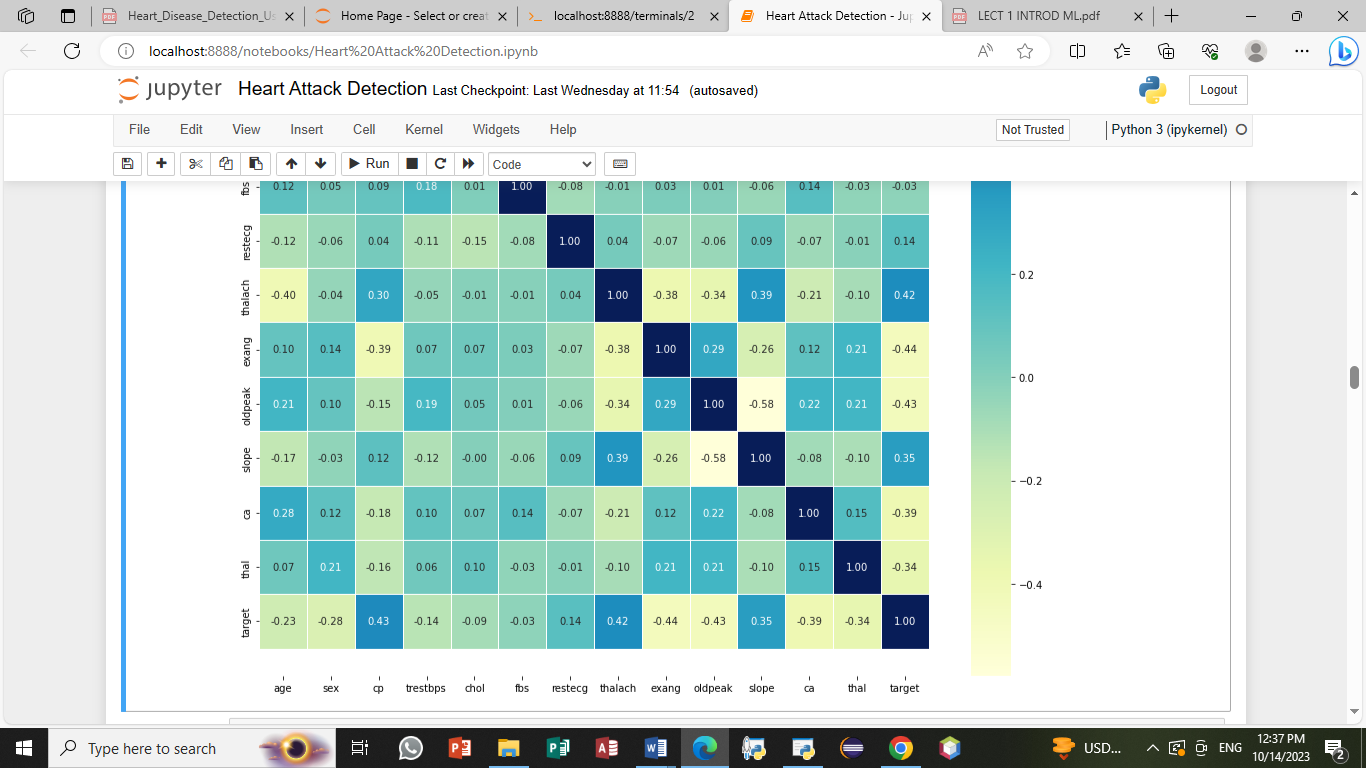
### 4.3.3 possibility of heart attack based on heart rate

While analysing the dataset keenly, we also discovered that many people affected by the Heart attack have a maximum heart rate of between 150 to 180. the figure below clearly explains the information above;



### 4.3.4 Correlation Heatmap

To uncover potential relationships between features, a correlation heatmap was generated. Figure 4.2 demonstrates the correlations between attributes, highlighting patterns that may influence heart attack likelihood.



## 4.4 Accuracy of the SVM Model

The heart attack detection model utilized in this project is based on a Support Vector Machine (SVM) classifier. The model was trained and evaluated on the provided dataset, resulting in an accuracy of 83%.

This impressive accuracy showcases the model's capability to predict heart attack risk, utilizing the patient's attributes and the power of SVM classification.

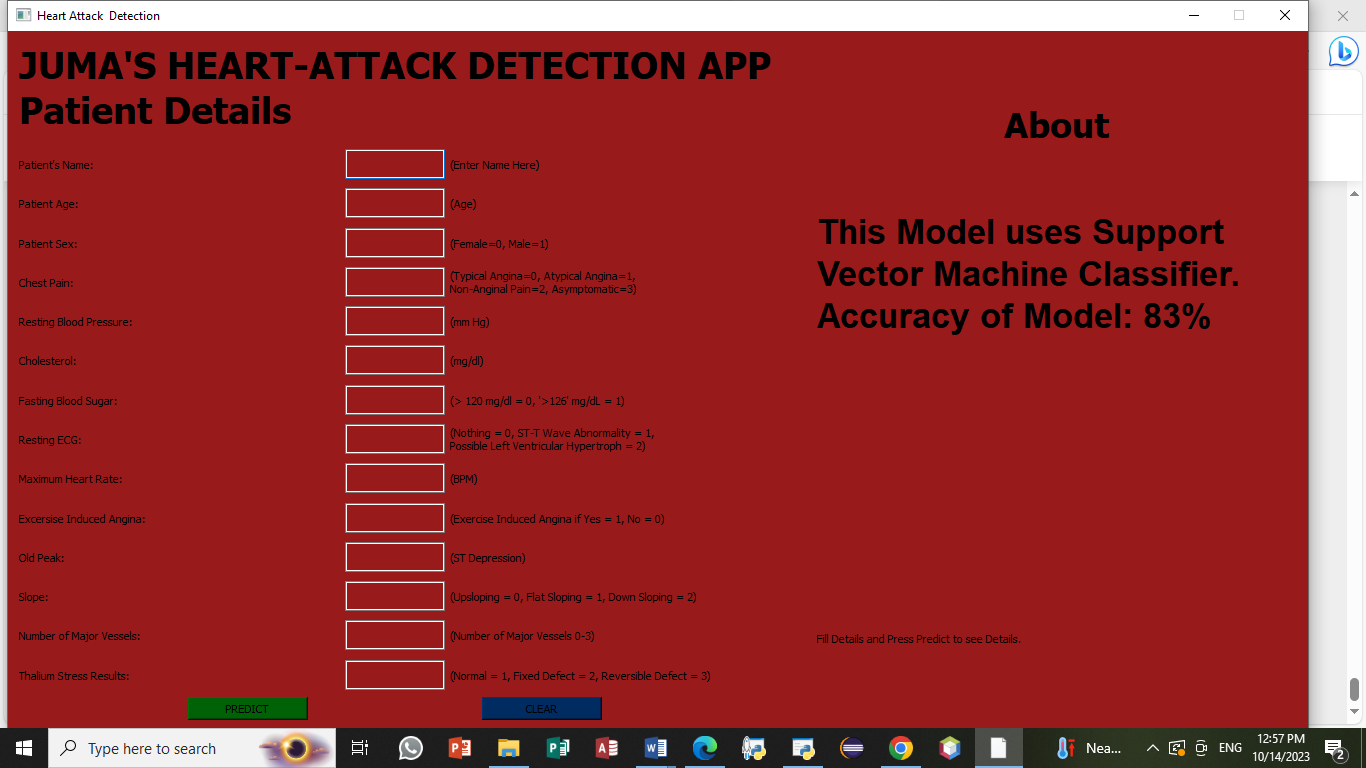
## 4.5 Integration with GUI

### 4.5.1 User-Friendly Interface

Incorporating the heart attack detection model into a graphical user interface (GUI) enhances the usability and accessibility of the prediction system. The GUI allows users to input their details and obtain predictions with ease. The interface is designed with a clean and intuitive layout, ensuring a user-friendly experience.

### 4.5.2 Input Validation

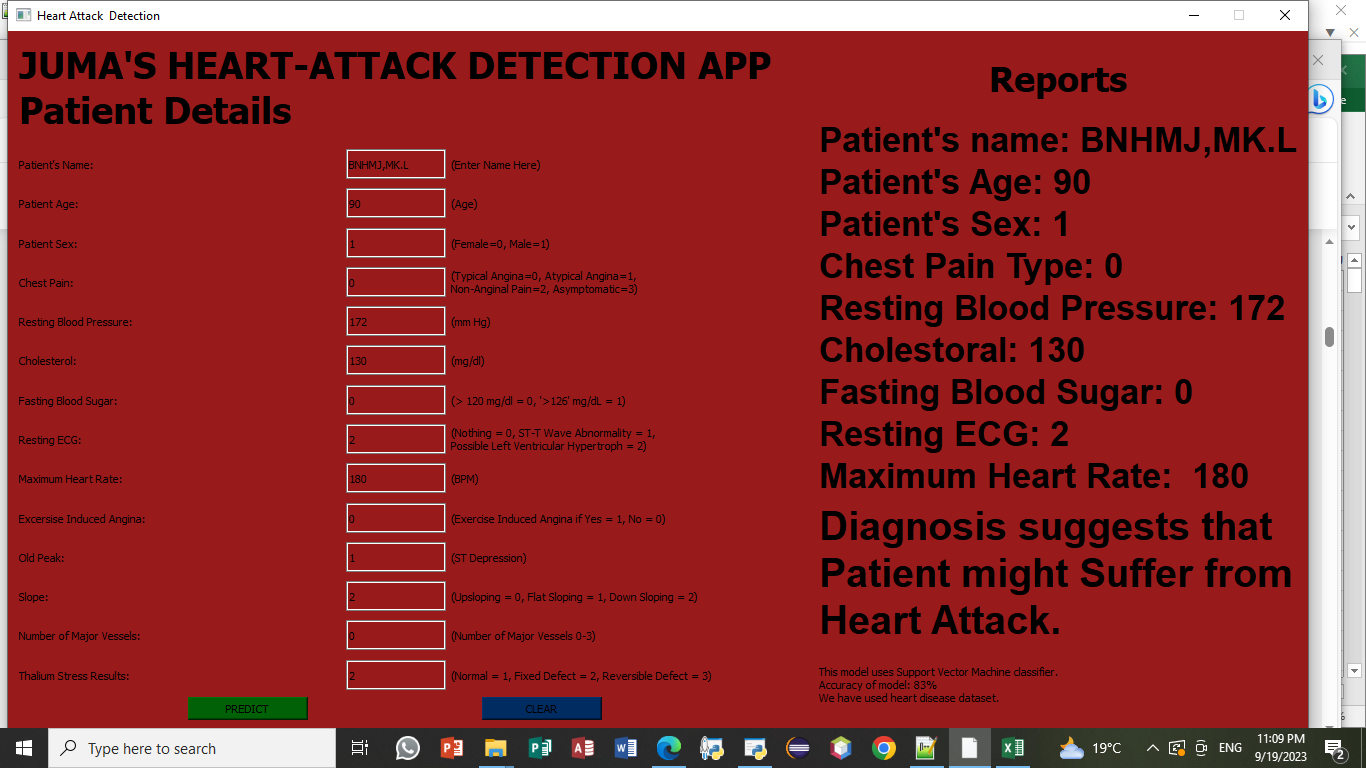
To maintain data integrity and user input consistency, we have provided clear guidelines of what the user should input at any given place. This feature ensures that users provide valid data, reducing errors and enhancing prediction accuracy.



### 4.5.3 Prediction Results

Upon clicking the "PREDICT" button, users receive prompt predictions based on their input. The result is displayed clearly, making it easy for users to interpret the outcome.

Below is a sample of the prediction after one input the required values, From the values obtained the model predicts that the patient is likely to be affected by heart attack and should be diagnosed



## 4.4 Conclusion

In Chapter Four, we presented the results of the heart attack detection model that has achieved all the stated objectives in Chapter one. The SVM classifier achieved an impressive accuracy of 83%, demonstrating its capability to predict heart attack risk accurately. Additionally, we discussed the integration of the model with a user-friendly GUI, providing a convenient platform for users to input their details and receive prompt predictions.

This integrated system serves as a valuable tool for individuals to assess their heart attack risk, ultimately contributing to improved health awareness and preventive measures.

# CHAPTER FIVE: RECOMMENDATIONS AND CONCLUSSIONS

## 5.1 Conclusion

The development of the heart attack prediction system, which seamlessly integrates a Support Vector Machine (SVM) model with a user-friendly graphical user interface (GUI), has been a significant achievement. In this chapter, we provide a comprehensive conclusion to summarize the project's key findings, contributions to knowledge, and offer recommendations for future work.

### 5.1.1 Key Findings

* **Model Accuracy**: The SVM-based heart attack detection model has demonstrated a commendable accuracy of 83% in predicting heart attack risk. This level of accuracy is a testament to the effectiveness of the model in assessing a patient's risk based on their attributes.
* **User-Friendly GUI**: The GUI developed for this project offers an intuitive platform for users to input their information and obtain prompt predictions. The interface is designed with user experience in mind, making it accessible to a wide range of users.
* **Input Validation**: The integration of input validation ensures data integrity and consistency in user input. This feature reduces the likelihood of errors and inaccuracies in the predictions.
* **Data Visualization**: Data visualization techniques, including feature distribution histograms and correlation heatmaps, have provided insights into the dataset's characteristics, helping identify relationships and patterns among attributes.

### 5.1.2 Contributions to Knowledge

This project contributes to knowledge in several ways:

* **Healthcare Informatics**: The development of a heart attack prediction system provides a practical application of machine learning in healthcare. It enhances our understanding of how machine learning models can be employed for early risk assessment and preventive healthcare.
* **User-Friendly Design**: The user-friendly GUI design offers a template for creating accessible healthcare prediction tools. The project's GUI design principles can be applied to various health-related applications.
* **Data Visualization in Healthcare**: The use of data visualization techniques to understand healthcare data is an important contribution. It emphasizes the significance of data visualization in interpreting medical datasets.

## 5.2 Recommendations

While the heart attack prediction system has achieved a high degree of accuracy and usability, we recommend the following;

1. **Expanded Dataset**: To enhance the model's accuracy, it is recommended to collect a larger and more diverse dataset. Additional data sources can provide a more comprehensive understanding of heart attack risk factors.
2. **Feature Engineering**: Further investigation into feature engineering and selection methods can lead to improved model performance. Identifying the most influential attributes for prediction can help refine the model.
3. **Validation and Clinical Trials**: Conducting clinical trials and medical validation of the model is crucial to assess its real-world applicability. Collaboration with healthcare professionals can help validate the system's predictions.
4. **Mobile Application**: Developing a mobile version of the application can enhance accessibility, allowing users to access the heart attack prediction system from their smartphones.

## 5.3 Final Thoughts

In conclusion, the heart attack prediction system, comprising an accurate SVM model and a user-friendly GUI, is a significant step towards early risk assessment and healthcare informatics. The project's contributions to knowledge, coupled with recommendations for future work, lay the foundation for further research and development in the field of healthcare prediction systems. With continuous improvement and validation, this system has the potential to make a meaningful impact on healthcare and empower individuals to take proactive steps in managing their heart health.

The journey towards more accurate and accessible healthcare prediction tools continues, and the heart attack prediction system represents a promising stride in this direction.

# REFERENCES

1. Ali L., Rahman A., Khan A., Zhou M., Javeed A., Khan J. A.( 2019) An automated diagnostic system for heart disease prediction based on statistical model and optimally configured deep neural network. IEEE Access .
2. Ali Sellami et al. (2019) **[A robust deep convolutional neural network with batch-weighted loss for heartbeat classification](https://www.sciencedirect.com/science/article/pii/S0957417418308054" \t "_self)**
3. Aya Awad et al. (2017) **[Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach](https://www.sciencedirect.com/science/article/pii/S1386505617303581" \t "_self)**
4. Behnam Fahimnia et al. (2015) **[Green supply chain management: a review and bibliometric analysis](https://www.sciencedirect.com/science/article/pii/S0925527315000067" \t "_self)**.
5. Chiuchisan, I., & Geman, O. (2014). An Approach of a Decision Support and Home Monitoring System for Patients with Neurological Disorders Using Internet of Things Concepts. WSEAS Transactions on Systems, 13, 460-469.
6. S. Anitha and N. Sridevi, (2015) “Heart disease prediction using data mining techniques,” *Journal of Analysis and Computation*, vol. 8, no. 2, pp. 48–55.
7. A. Jagtap, P. Malewadkar, O. Baswat, and H. Rambade, (2019) “Heart disease prediction using machine learning,” *International Journal of Research in Engineering, Science and Management*, vol. 2, no. 2, pp. 352–355.
8. Dhiman, C.; Vishwakarma, D.K. (2019) A review of state-of-the-art techniques for abnormal human activity recognition.
9. K. Subhadra and B. Vikas, (2019) “Neural network based intelligent system for predicting heart disease,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 5, pp. 484–487.
10. S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim (, 2013) “Innovative artificial neural networks-based decision support system for heart diseases diagnosis,” *Journal of Intelligent Learning Systems and Applications*, vol. 5, no. 3, pp. 176–183.
11. F. M. J. M. Shamrat, M. A. Raihan, A. K. M. S. Rahman, I. Mahmud, and R. Akter, (, 2020) “An analysis on breast disease prediction using machine learning approaches,”’*’ Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 2450–2455.
12. M. S. Amin, Y. K. Chiam, and K. D. Varathan, (2019) “Identification of significant features and data mining techniques in predicting heart disease,” *Telematics and Informatics*, vol. 36, pp. 82–93.
13. H. Wu, T. T. Huynh, and R. Souvenir, (2014) “Motion factorization for echocardiogram classification,” in 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), pp. 445–448, Beijing, China.
14. Hazra, A., Mandal, S., Gupta, A., & Mukherjee, A. (2017). Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review. Advances in Computational Sciences and Technology, 10, 2137-2159.
15. Houda Benhar (2020) **[Data preprocessing for heart disease classification: a systematic literature review](https://www.sciencedirect.com/science/article/pii/S0169260720314681" \t "_self)**. Computer Methods and Programs in Biomedicine
16. K. T. Spencer, B. J. Kimura, C. E. Korcarz, P. A. Pellikka, P. S. Rahko, and R. J. Siegel, (,2013)“Focused cardiac ultrasound: recommendations from the American society of echocardiography,” Journal of the American Society of Echocardiography, vol. 26, no. 6, pp. 567–581.
17. Kandala N.V.P.S. Rajesh (2018). **[Classification of imbalanced ecg beats using re-sampling techniques and adaboost ensemble classifier](https://www.sciencedirect.com/science/article/pii/S1746809417302872" \t "_self)**
18. Mshali, H.; Lemlouma, T.; Moloney, M.; Magoni, D. (2020) A survey on health monitoring systems for health smart homes. .
19. Patel, A.; Fang, J.; Gillespie, C.; Odom, E.; Luncheon, C.; Ayala, C.(2018.) Awareness of heart attack signs and symptoms and calling 9-1-1 among U.S. adults. J. Am. Coll. Cardiol.
20. Patel, J., Upadhyay, P., & Patel, D. (2016). Heart Disease Prediction Using Machine learning and Data Mining Technique. Journals of Computer Science & Electronics, 7, 129-137.
21. Shouman, M., Turner, T., & Stocker, R. (2012). Using Data Mining Techniques in Heart Disease Diagnosis and Treatment. \*Electronics, Communications, and Computers, Alexandria, 173-177.
22. Singh, M., Martins, L.M., Joanis, P., & Mago, V.K. (2016). Building a Cardiovascular Disease Predictive Model Using Structural Equation Model and Fuzzy Cognitive Map. IEEE International Conference on Fuzzy Systems (FUZZ), Vancouver, 24-29 July 2016, 1377-1382. **[https://doi.org/10.1109/FUZZ-IEEE.2016.7737850](https://doi.org/10.1109/FUZZ-IEEE.2016.7737850" \t "_new)**
23. Soni, J., Ansari, U., & Sharma, D. (2011). Intelligent and Effective Heart Disease Prediction System Using Weighted Associative Classifiers. International Journal on Computer Science and Engineering (IJCSE), 3, 2385-2392.
24. Taissir Fekih Romdhane (2020)**[Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss](https://www.sciencedirect.com/science/article/pii/S0010482520302237" \t "_self)** Comput Biol Med
25. U. Rajendra Acharya (2017)**[A deep convolutional neural network model to classify heartbeats](https://www.sciencedirect.com/science/article/pii/S0010482517302810" \t "_self)**. Comput Biol Med
26. W. S. Colucci, (2021)“Patient education: Heart failure (beyond the basics),”.
27. Weiyi Yang (2018) **[Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine](https://www.sciencedirect.com/science/article/pii/S001048251830221X" \t "_self)**.
28. Weng, S.F., Reps, J., Kai, J., Garibaldi, J.M., & Qureshi, N. (2017). Can Machine-Learning Improve Cardiovascular Risk Prediction Using Routine Clinical Data? PLoS ONE, 12, e0174944. **[https://doi.org/10.1371/journal.pone.0174944](https://doi.org/10.1371/journal.pone.0174944" \t "_new)**
29. World Health Organization (2018.) The Top 10 Causes of Death; WHO: Geneva, Switzerland,
30. Yahaya, S.W.; Lotfi, A.; Mahmud, M. (2019) A Consensus Novelty Detection Ensemble Approach for Anomaly Detection in Activities of Daily Living. Appl. Soft Comput.
31. Yuce, M.R. (2010). Implementation of Wireless Body Area Networks for Healthcare Systems. Sensor and Actuators A: Physical, 162, 116-129. **[https://doi.org/10.1016/j.sna.2010.06.004](https://doi.org/10.1016/j.sna.2010.06.004" \t "_new)**
32. Zhao, W., Wang, C., & Nakahira, Y. (2011). Medical Application on Internet of Things. IET International Conference on Communication Technology and Application (ICCTA 2011), Beijing, 14-16 October 2011, 660-665.

## APPENDICES

Below is the Code snippet for the development of the SVM model with an Accuracy of 83%.

