A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset

by

Abdullah Al-Monsur, 170021015
Ishrak Jahan Ratul, 170021100
Abrar Mohammad Ar-Rafi, 170021114

A Dissertation
Submitted to the Department of Electrical & Electronic Engineering

In Partial Fulfillment of the Requirements
for the Degree of Bachelor of Science in Electrical & Electronic Engineering

Islamic University of Technology (IUT)
Gazipur, Bangladesh
May, 2022

Certificate of Approval

The thesis titled “A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset” submitted by Abdullah Al-Monsur (170021015), Ishrak Jahan Ratul (170021100) and Abrar Mohammad Ar-Rafi (170021114) has been found as satisfactory and accepted as partial fulfillment of the requirement for the degree of Bachelor of Science in Electrical and Electronic Engineering on 10th May, 2022.

Approved by:

(Signature of the Supervisor)
Fahim Faisal
Assistant Professor
Department of Electrical and Electronic Engineering
Islamic University of Technology

(Signature of the Co-Supervisor)
Mirza Muntasir Nishat
Assistant Professor
Department of Electrical and Electronic Engineering
Islamic University of Technology

Dedication

To

All the men and women of every land, who are not afraid of themselves, who trust so much in their own souls that the dare to stand up in the might of their own individuality to meet the tidal currents of the world.

Acknowlegements

First and foremost, we express our heartfelt gratitude to The Almighty Allah for providing us with the patience and strength to complete this thesis work smoothly. We are grateful to our honorable supervisor, **Fahim Faisal**, Assistant Professor, Department of Electrical and Electronic Engineering, IUT, for allowing us to work under him. He has shown us the proper way to conduct research and advised us to gather a strong foundation for our work. He has always motivated us to create new and innovative works. We would have lost track and become disoriented if it hadn't been for his help and advice.

We'd like to thank our co-supervisor, **Mr. Mirza Muntasir Nishat**, Assistant Professor, Department of EEE, IUT, for always giving us his time and effort when it was needed. He introduced us to the field of research when we had no prior knowledge of it. His continuous advice and insights have been extremely beneficial to us. He was always motivating us when we were stuck with different problems and always found a solution for us. We would struggle to complete this research without his clear guidance.

Abstract

The chronic cardiac condition myocardial infarction (heart failure) is characterized by decreased blood supply to the body as a result of the heart muscles' impaired contractile properties. Patients with heart failure, like those with any other cardiac disorder, have difficulty performing daily activities and have a shorter life expectancy, with the vast majority of cases resulting in death at some point during the patient's lifetime. Treatment outcomes and patient quality of life improve significantly when patients with heart failure are identified early and are likely to survive. As a result, machine learning techniques can be extremely beneficial in this situation because they can be used to predict the survival of heart failure patients in advance, allowing patients to receive the most appropriate treatment at the earliest possible stage. As a result, six supervised machine learning algorithms were applied to a dataset of 299 people from the University of California, Irvine Machine Learning Repository in order to predict their chances of surviving heart failure. There were a variety of algorithms used in this study including Decision Tree Classifier, Logistic Regression, Gaussian Nave Bayes, Random Forest Classifier, K-Nearest Neighbors, and Support Vector Machine, among others. Prior to scaling the data, a preprocessing step was carried out, and both the standard and min-max scaling methods were employed. When it came to optimizing the hyperparameters, the techniques grid-search cross validation and random search cross validation were combined. Data resampling techniques such as the edited nearest neighbor (SMOTE-ENN) and synthetic minority oversampling (SMOTE) data resampling are also employed (SMOTE-ENN). It has been thoroughly compared and analyzed the outcomes of all of the different approaches. As a result of these findings, the Random Forest Classifier (RFC) outperforms all other approaches, achieving a test accuracy of 90 percent when compared to the other approaches when SMOTE-ENN and the standard scaling technique are employed. With the help of an imbalanced dataset, this comprehensive investigation vividly illustrates the application and compatibility of several machine learning algorithms. Among the methods for improving the performance of machine learning algorithms discussed in this investigation are the SMOTE-ENN algorithm and hyperparameter optimization.

Contents

Chapter 1

Introduction

1.1 Heart Failure

Cardiac disease is a complicated condition that causes the heart to produce too little blood in comparison to the body's needs for blood. Heart failure is caused by many parts of the heart, including the pericardium, the myocardium, the endocardium, the heart valves, and the major blood vessels, which are all parts of the heart. Heart failure can happen to anyone who lives in any of these places. Blood flow to the body and lungs is cut off because of symptoms of cardiovascular disease, which show up as both forward and backward blood flow. Besides symptoms like shortness of breath, edema, and fatigue, heart failure shows up on a physical exam with rales and other things that can be seen. Ejection fraction can be used to group people with heart failure into two groups, one with a good heart function (heart failure with preserved ejection fraction: HFpEF) and one with a poor one (heart failure with reduced ejection fraction: HFrEF). The heart's left ventricular (LV) chamber changes in a way that causes HFpEF. This is the main idea behind the disease. There are now two different ways to think about how HFpEF affects the body. There was a lot of talk about diastolic dysfunction in the ventricles, left ventricular hypertrophy (a big heart), impaired relaxation, endothelial dysfunction, arterial stiffness, and how these things affect cardiac function in the old models. Diastolic dysfunction and left ventricular hypertrophy are both talked about in the new model of the heart. Model: The role of systemic microvascular endothelial inflammation caused by pre-existing conditions like diabetes and hypertension as well as obesity and smoking as well as ischemia was explained in great detail in the model that was just made. One of the main roles of HFrEF is to control how cells grow and how they use energy in the body, among other things. In both HFrEF and HFpEF, changes in excitation-contraction coupling, changes in epigenetics, and changes in sarcomeric coupling proteins make them different. People with heart failure also have more adrenaline released, more renin-angiotensin-aldosterone axis activity, less nitric oxide insensitivity, less ATP, reactive oxygen species, and a higher cell death rate because

of these other pathological processes. Alter Heart failure can be caused by two types of things: cardiac and noncardiac things. It's most common for heart problems to be the cause of death. There are many different things that can cause heart disease, like aortic regurgitation, heart valve problems, arrhythmias and heart rhythm problems. There are also things like heart aneurysm and congenital heart disease, as well as constrictive pericarditis and dilated cardiomyopathy, Eisenmenger syndrome and endocarditis that can cause problems. Congestive heart failure can cause breathing problems, fatigue, and edema. These are all signs and symptoms that should be looked out for. As many as 5 million people in the United States of America have heart failure, and about 500,000 people have it each year. Congestive heart failure costs a lot of money in the United States because so many people need to come back to the hospital within six months of leaving the hospital, and they stay in the hospital for an average of six days. This costs a lot of money in the country's healthcare budget. The number of people in the United States who had HIV in 2005 was about 5,300,000 people who were at least 20 years old, according to current estimates. It was estimated that 2,650,000 people were made up of both men and women. Heart failure is linked to big changes in both physical and mental health, and as a result, people who have it don't enjoy their lives as much as they used to. Heart failure caused by congestive heart failure has also been linked to a bad prognosis for people who have it. As a result of the complications that can happen if heart failure isn't treated, it can kill you if you don't get it taken care of. The heart failure caused by atherosclerotic coronary artery disease is linked to more fatal events than other types of cardiovascular disease. Chronic heart failure (CHF) is a disease that gets worse over time. It has a big negative impact on the patient's overall quality of life, so it needs to be treated. If you have heart failure caused by something that can be fixed, the condition tends to get worse over time. Cancer progresses, and even though some people live for a long time after being diagnosed with it, the disease gets worse and has a 10death rate every year. In the Framingham Heart Study, 80of women who had heart failure before the age of 65 died within eight years of being diagnosed with the condition. Multiple diagnostic criteria and algorithms are used to determine whether someone has heart failure. These include one developed by the European Society of Cardiology, the Framingham Heart Study, and the Boston Heart Study, to name a few. All of the symptoms of heart failure, like shortness of breath, fatigue, and water retention, should be kept in mind. Heart failure shows up in a lot of different ways in different people. Exercise isn't good for some people, but their bodies don't get crowded or swollen because of this. Other people have mild edema and bronchial obstruction in another part of the population. Ejection fraction: People who have systolic heart failure and have symptoms usually have an ejection fraction that is less than 35of what it should be. Those who have unexplained shortness of breath may benefit from BNP tests, which can help doctors find heart failure early on. People who have heart failure can have cardiomegaly (cardiac enlargement and congestive heart failure),

which can be seen during a physical examination of them (Kerley B lines, and in some cases pleural effusion). To figure out why the left ventricular systolic or diastolic function doesn't work, a doctor can use CMR to look at the size and shape of the left and right ventricles, as well as the myocardial tissue in the heart. Echocardiography is a type of imaging technique that is used a lot to find out if someone has heart failure in both adults and kids. To figure out the stroke volume (SV), which is the amount of blood that leaves the heart's ventricles with each heartbeat, ultrasound can be used. End-diastolic volume (EDV) can also be found with ultrasound (EDV, the total volume of blood at the end of diastole). The ejection fraction is the difference between the SV and the EDV (EF). In children, the shortening fraction is the best way to figure out how well their systolic function is working. It is done on people who have heart failure and have a strong suspicion that their heart failure is caused by a condition called atherosclerosis, which is a disease of the blood vessels in the heart. Candidate for revascularization who show signs of an acute coronary syndrome should have coronary angiography as part of their evaluation. People who show signs of an acute coronary syndrome should have this done. Acute heart failure is a medical emergency that can happen if you have new-onset heart failure or if your chronic heart failure has worsened (also known as acute decompensated heart failure, flash pulmonary edema, ADHF). People who have ADHF have a shortness of breath that gets worse quickly. This is because the lungs have filled with fluid (the rapid accumulation of fluid in the lung). The condition ADHF is also marked by low blood pressure that affects the flow of blood and oxygen to different parts of the body. This causes kidney function to deteriorate and thought processes to change. Cold and clammy extremities are also common. A person who has ADHF is likely to have a bad outcome if the condition isn't treated quickly. Its main goal is the same as that of chronic heart failure therapy. However, unlike chronic heart failure therapy, it also tries to improve oxygenation and hemodynamic stability. During the treatment of acute decompensated congestive heart failure, oxygen, diuretics, and vasodilators are all used. Oxygen helps with hypoxia, diuretics reduce preload and intravascular volume, and vasodilators reduce afterload. The implementation of some important parts of chronic heart failure treatment is being held up (ACE inhibtors, beta blockers and digoxin). It's important for people who have congestive heart failure to have their comfort at the top of their end-of-life care, Patients may choose to have their defibrillators turned off if they think they need to avoid as many invasive procedures as possible. Palliative care or hospice care should be thought about because there are more people with Stage IV heart failure who are being treated (who experience intractable symptoms of fatigue, shortness of breath, or chest pain at rest despite optimal medical therapy).

1.2 What is machine learning?

Machine learning has become one of the most important and effective tools in engineering as a result of the rise of artificial intelligence. Artificial intelligence in the form of machine learning is becoming increasingly popular. Without having to be explicitly programmed, it allows computers to build on their prior experiences and learn new things.
Using machine learning, computers are able to learn on their own without any human intervention or help. As a side benefit, it allows computers to act based on their prior experiences. It is necessary to build a model from massive datasets in order to provide the previous experience. It is through these datasets that the model can better hone its abilities. A machine learning model relies on these datasets the most. The model takes all the necessary steps to ensure that the user receives the best possible results.
There are primarily three types of machine-learning approaches, depending on the model type.
Here are some real-world instances of machine learning:

Speech recognition: Voice recognition is the act of converting human speech into a written representation using natural language processing (NLP). The phrases automatic speech recognition (ASR), computer speech recognition (CSR), and voice-to-text all refer to the same thing. Many mobile devices employ speech recognition to execute voice searches (for example, Siri) and to improve messaging accessibility.

Customer service: Throughout the customer experience, online chatbots are gradually replacing human people. They provide personalized recommendations, cross-sell items, and size recommendations to customers, altering the way we think about customer engagement on websites and social media platforms. They give customers personalized advice, cross-sell products, and recommend sizes by responding to frequently asked questions (FAQs) about topics like shipping. Examples include virtual agent message bots on e-commerce sites, messaging apps like Slack and Facebook Messenger, and virtual assistant and voice assistant functionalities.

Computer vision: Computer vision is a type of artificial intelligence that allows computers and systems to extract useful information from digital images, videos, and other visual inputs and then act on it. It is distinct from other picture recognition jobs in that it may make suggestions. Computer vision, which has applications in social media photo tagging, healthcare radiological imaging, and self-driving automobiles, uses convolutional neural networks.

Recommendation engines: AI algorithms can help in the finding of data trends that can be used to create more effective cross-selling strategies by using past customer behavior

data. This is used to assist clients with appropriate add-on recommendations throughout the checkout process for online businesses.

Automated stock trading: Artificial intelligence-powered high-frequency trading platforms may execute hundreds or even millions of deals each day without the need for human intervention, allowing them to optimize stock portfolios.

1.2.1 Supervised Machine Learning:

Supervised Machine Learning is an algorithm that learns from labeled training data in order to assist us in predicting outcomes for unlabeled training data. Using data that has been ”labeled” correctly, you can train the machine using supervised learning. It indicates that some information has already been labeled with correct answers. When learning in the presence of a supervisor or a teacher, it is comparable to that experience.

1.2.2 Unsupervised Machine Learning:

Unsupervised Learning is a machine learning technique that eliminates the need for users to supervise the model during training. The model is then free to work autonomously, discovering previously unknown patterns and information. It is primarily concerned with unlabeled data.

It is not possible to apply unsupervised learning directly to a regression or classification problem because, unlike supervised learning, we do not have access to corresponding output data, whereas with supervised learning we do. Using unsupervised learning, the goal is to discover the fundamental structural properties of a dataset, classify that data into groups based on similarities, and represent that dataset in a compressed format.

1.2.3 Reinforcement Machine Learning:

The term ”reinforcement learning” refers to the process of instructing machine learning models to make a series of decisions over a long period of time. The agent learns how to complete a task in an uncertain and potentially complex environment by observing and learning from others. An artificial intelligence is presented with a situation that is similar to a game in order to learn through reinforcement. The computer works its way through the problem by trial and error. The artificial intelligence is rewarded or punished for its actions in order to persuade the machine to perform the task that the programmer desires. Its goal is to maximize the total amount of money earned.
A unique set of objectives exists in the context of reinforcement learning, as opposed to the objectives of unsupervised learning. Unsupervised learning has the goal of discovering similarities and differences between data points, whereas reinforcement learning has the goal of finding an appropriate action model that maximizes the total cumulative reward received by the agent. Action-reward feedback loop of a generic reinforcement learning model is depicted in the diagram below.

Unsupervised learning makes use of fresh data, whereas supervised learning makes use of data from an existing model. Sustained models help guide input to output, whereas unsupervised systems have no such guidance. Using a model that interacts with a dynamic model in which a specific task is fixed and must be completed by the computer, reinforcement learning is a method to be employed.

1.3 Machine Learning in Health Sector

Machine learning is now incorporated into nearly every part of our lives and is gaining popularity. Machine learning is being used to increase accuracy in a wide variety of applications, from everyday email checking to launching a rocket into outer space. Apart from healthcare, machine learning is employed in a wide variety of other disciplines. In the near future, it is likely to become one of the most commonly used branches of machine learning.

Machine learning is frequently used in the healthcare field, with a variety of benefits for patients and professionals. The most widely utilized applications of machine learning in healthcare are for automating medical billing, clinical decision support, and the formulation of clinical care recommendations. Clinical support and advice assist us in recognizing diseases effectively and implementing appropriate measures based on an accurate diagnosis. Machine learning has the potential to be a novel and effective method of diagnosing medical problems. It is entirely focused on locating and retrieving data from large databases.

Treatment for heart failure demands early detection. Machine learning has the ability to considerably aid health practitioners in recognizing heart failure in the symptom of symptoms. Researchers worldwide have already begun utilizing data science and machine learning algorithms to predict HF in its early and mid-stages, allowing medical professionals to provide a better cure to the general population before HF progresses to a critical stage and increases the risk of other fatal diseases.

Machine learning is an important and valuable approach in the medical field nowadays for detecting and forecasting various forms of disease and their stages. As machine learning is all about exploring enormous datasets and its related patterns, features, and modes, The massive volume of diagnostic data for various diseases can be included into various machine learning algorithms. This implementation of algorithms in medical databases can greatly benefit medical practitioners in making educated decisions regarding diseases, preventing errors, and ensuring the health of the general population. We studied eleven Machine Learning algorithms that performed well in predicting Heart Failure survival in this study.
In our study, we use six supervised machine learning algorithms along with SMOTE-ENN oversampling technique.

Chapter 2

Literature Review

Currently, machine learning (ML) is one of the most significant and successful techniques in the healthcare industry for detecting and forecasting a wide variety of diseases and their phases. The huge volume of diagnostic data for diverse diseases can be fed into various machine learning algorithms in order to find patterns and traits that are common to them. Including algorithms in medical databases can significantly aid medical practitioners in making educated decisions about diseases, minimizing the number of human-caused errors, and eventually guaranteeing that the vast majority of people live healthy lives.

The health sector is embracing machine learning for a variety of reasons, including disease prediction, medical imaging diagnostics, and personalized treatment. Numerous studies on the application of data mining techniques to the prediction of heart disease have been done in recent years. Numerous research publications have been published on the issue of employing machine learning techniques to predict heart failure patients' survival.

2.1 Related Works

1. Sanaa Elyassami and Achraf Kaddour developed a deep learning model iteratively using stochastic gradient descent. They incorporate the chi-square test and dropout regularization into the model to increase its performance in categorizing cardiovascular disease patients. The classification model obtains an accuracy of 91.43 percent on a balanced basis [1].

2. Sri Rahayu and Jajang Jaya Purnama utilize 6 algorithms, i.e., the Random Forest algorithm, Decision Tree, , Artificial Neural Network, and 3 more algorithms on the heart failure clinical dataset, along with resample and Synthetic Minority Oversampling Technique (SMOTE) sampling techniques. Random forest gets 94.31 percent accuracy when employing resample sampling, but only 85.82 percent accuracy

when using the SOMTE technique [2].

3. Oladosu Oyebisi Oladimeji and Olayanju Oladimeji developed an integrated technique based on machine learning for the prediction of heart failure patient survival. The integrated strategy addresses class imbalances in the classification dataset by ranking key predictive variables. They attain the best accuracy of 83.1788 percent utilizing the random forest technique [3].

4. Remzi Gürfidan and Mevlüt Ersoy classify the heart failure dataset using unique classification approaches. They have success rates ranging from 73 percent to 83 percent. Among the approaches, the Support Vector Machine (SVM) algorithm has the highest accuracy of 90 percent [4]

5. Abid Ishaq, Saima Abid, and Seyedali Mirjalili, working with their collaborators, apply nine categorization models to predict the survival of heart failure patients. To fix the problems of class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is implemented. The experimental results reveal that the Extra Tree Classifier (ETC) surpasses other models and achieves a prediction accuracy of 92.62 percent when used in conjunction with SMOTE [5].

6. Davide Chicco and Giuseppe Jurman did the most detailed analysis on UCI heart failure dataset. They utilize a number of machine learning classifiers to predict survival of patients and evaluate the variables associated with the most significant risk factors. Additionally, they apply standard biostatistics methodologies and compare their findings to those obtained by machine learning systems. Serum Creatinine and Ejection Fraction are the most significant features for creating a prediction model, according to both feature rankings. They attain 74 percent accuracy when all features are included, and 83.8 percent accuracy when only two features are considered, namely Serum Creatinine and Ejection Fraction [6].

7. The heart failure data collection was collected in 2015 by Tanvir Ahmad, Assia Munir, and his colleagues. In order to estimate the likelihood of survival after a heart attack, Ahmad used statistical study (Cox regression and Kaplan Meier plots). According to their data, the strongest predictive markers for heart failure are age, ejection fraction (EF), serum creatinine, serum sodium, anemia, and blood pressure [7].

8. Dr. Rubini, PE, and their colleagues did a comparative analysis of machine learning algorithms such as Random Forest (RF), Logistic Regression, Support Vector Machines (SVM), and Naive Bayes for cardiovascular disease classification. Random Forest (84.81 percent) and Linear Regression (83.82 percent) had the highest accuracy, according to their comparison analysis [8].

9. Liaqat Ali and Shafqat Ullah Khan have built a two-stage feature-driven decision support system to boost the accuracy of cardiac prediction. The first stage employs a two-stage statistical model to rank 13 HF features. A set of features is obtained by forward best-first search. Gaussian Navie Bayes (GNB) classifier is utilized as a predictive model in the second stage. This research reveals that the proposed technique is capable of predicting with a 93.33 percent accuracy [9].

10. Dhomse Kanchan and Kishor anticipate heart disease and diabetes using the fewest number of attributes possible. They do it utilizing SVM Naive Bayes, Decision Trees, and Principal Component Analysis, among other techniques. This study examines a hospital repository dataset that contains 1865 instances of two types of tests: blood and urine. The data is analyzed using SAS software. SVM outperforms all other algorithms, according to their findings [10].

11. 11. According to rahman and his colleagues, a high level of highly sensitive C reactive protein signals an abrupt ST elevation myocardial infarction in the interior [11].

12. Maryam Aljanabi and her co-researchers prepare a review paper on machine learning classification algorithms for predicting cardiac disease. They overviewed a total of 21 pieces of literature on machine learning. They determined that due to the scarcity of clinical datasets, most authors chose the Cleveland dataset from UCI (13 out of 21). (13 out of 21). Among the examined article, Dangare and Apte acquired the highest accuracy (nearly 100 percent) using Artificial Neural Network (ANN) utilizing the WEKA tool. Isra'a Ahmed Zriqat, Ahmad Mousa Altamimi, and their other researchers attained the second-highest (99.01 percent) accuracy with the Decision Tree technique using MATLAB. Shouman et al. attain the third-highest accuracy (97.2 percent) using the KNN method. The lowest accuracy (56.76 percent) was attained by Patel et al. using the J48 algorithm [12].

13. Tejal Upadhyay and Dr. Samir Patel used three different algorithms: J48 with Reduced Error Pruning, Logistic Model Tree Algorithm, and Random Forest Algorithm. J48 is the best classification tree method with pruning mechanism, according to the study. Model accuracy for J48 is 56.76 percent, and model creation takes 0.04 seconds [13].

14. M. Dangare and D. Apte built the Heart Disease Prediction Model (HDPS) utilizing a multilayer perceptron neural network and a backpropagation algorithm. The Cleveland and Statlog (UCI) datasets were utilized to train the MLPNN. They got 99.25 percent accuracy with 13 attributes and raised it to 100 percent utilizing two new attributes (obesity and smoking) [14].

15. Zriqt and Altamimi utilize leave one out cross validation to examine five classification methods. With an accuracy rate of 99.0 , they discovered that the Decision Tree surpasses other classifiers, followed by Random Forest [15].

16. Shouman and Turner study if combining voting with KNN may boost the accuracy of heart disease diagnosis when compared to utilizing an ensemble of neural networks. According to their research, voting did not boost the accuracy of KNN in identifying heart disease. KNN, on the other hand, may independently obtain a high accuracy of 97.4 percent [16].

17. Shahadat Uddin and Arif Khan undertook review research to find major trends in the performance of numerous supervised machine learning algorithms and to establish how well each algorithm performs in terms of disease prediction. According to their findings, SVM outperforms other machine learning algorithms in predicting heart disease [17].

18. Mustaqeem and Anwar established a prediction model for arrhythmia categorization. The model is generated by first picking the best attributes using a wrapper algorithm around random forest, and then using different machine learning classifiers on the extracted features, i.e., SVM, KNN (7-layer percerptron), NB, Random Forest, Multi-Layer percerptron (MLP) (MLP). The experimental results reveal that MLP has a greater accuracy (78.26 percent) than KNN (76.6 percent) and SVM (74.4 percent) [18].

19. Nida Khateeb and Muhammad Usman conducted research to analyze different classifiers (DT, KNN, NB, and Bagging) and to apply them to the Cleve-land (UCI) data set in order to establish their accuracy using various feature reduction methodologies. DT surpasses Nave Bayes on seven of the fourteen features, but Nave Bayes outperforms DT on eleven of the fourteen features. However, the KNN obtains the maximum accuracy when all features (14) are integrated [19].

20. Sonam Nikhar and A.M. Karandikar discuss the Nave Bayes and decision tree classifiers, as well as the development of two algorithms for predicting heart disease. Although Bayesian classification is superior, Decision Tree classification is superior as well. They eliminate superfluous, irrelevant attributes and employ the Selective Naive Bayes classifier technique to improve performance in NB [20].

2.2 Problem Definition

So, while conducting our research, we discovered the following issues:

1. The vast majority of clinical data available in open-source repositories is comprised of imbalanced datasets, and feeding these datasets into machine learning algorithms results in erroneous outcomes.
2. When we use the machine learning algorithms with the default parameters, we do not get the results that we want them to give. In order to deal with this, we must fine-tune our model.
3. We must determine the most effective algorithm for predicting cardiac failure.

2.3 Research Objective

As heart failure is very serious and prevention is vital, patients must often speak with medical professionals. However, when advising or treating patients, healthcare practitioners should consider a range of situations and features. In many circumstances, however, diagnostic devices and qualified medical technologists are insufficient, making speedy and precise detection of the patient's condition relatively challenging. To aid healthcare practitioners, massive volumes of data are collected and processed for real-life patient scenarios. The main objectives of our study are:

1. Building trustworthy machine learning based heart failure survival models.
2. Achieving better performances with various algorithms and techniques.
3. Comparative analysis of various models to find the best model.
4. Analysis and comparison of existing works.
5. Feasibility and implementation in real world scenario.

2.4 Data Description

This dataset was gathered from the UCI machine learning repository, which contains the medical history of 299 patients from the Faisalabad Institute of Cardiology and the Allied Hospital in Faisalabad (Punjab, Pakistan) (Punjab, Pakistan) [21]. It comprises data on 105 women and 194 men with class 3 or 4 Left Ventricular Systolic Dysfunction (LVSD) according to the New York Heart Association (NYHA). The ages of the patients ranged from 40 to 95 years, and the duration of the follow-up extended from 4 to 285 days. The dataset contains thirteen attributes that were evaluated during hospital follow-ups with patients.

Table 2.1 provides a summary of the qualities. However, only seven of the thirteen characteristics are numeric, with the remaining six being Boolean. Table 2.2 presents the statistical information regarding the numerical qualities. The dataset was then imported into Jupyter Notebook and submitted to exploratory data analysis to establish its validity and general characteristics. As illustrated in Fig. 2.1, a correlation heatmap was produced to determine the degree of relationship between the attributes.

Table 2.1: Attribute of the dataset

Sl No.	Attribute	Measurement	Information
1	age	Years	Age of the patient
2	anaemia	Boolean	Decrease of red blood cells or hemoglobin
3	creatinine phosphokinase	Mcg/L	Level of creatine phosphokinase enzyme in blood
4	diabetes	Boolean	If the patient has diabetes
5	ejection fraction	Percentage	volume of blood ejected from left ventricle in each contraction.
6	high blood pressure	Boolean	If the patient has high blood pressure
7	platelets	Kiloplatelets/mL	Platelets count in the blood.
8	serum creatinine	Mg/dL	Level of creatinine in the blood
9	serum sodium	mEq/L	Level of sodium in the blood
10	sex	Binary	Man or woman
11	smoking	Boolean	If the patient has a smoking habit
12	time	Days	Follow up period
13	death event	Boolean	If the patient is died during follow up period

Table 2.2: Statistical information of numeric attributes

Sl No.	Numeric Attribute	Maximum	Minimum	Mean	Standard Deviation
1	Age	95.0	40.0	60.83	11.89
2	creatinine phosphokinase	7861.0	23.0	581.84	970.29
3	ejection fraction	80.0	14.0	38.084	11.83
4	Platelets	850000.0	25100.0	263358.03	97804.24
5	serum creatinine	9.4	0.5	1.40	1.035
6	serum sodium	148.0	113.0	136.63	4.41
7	Time	285.0	4.0	130.26	77.61

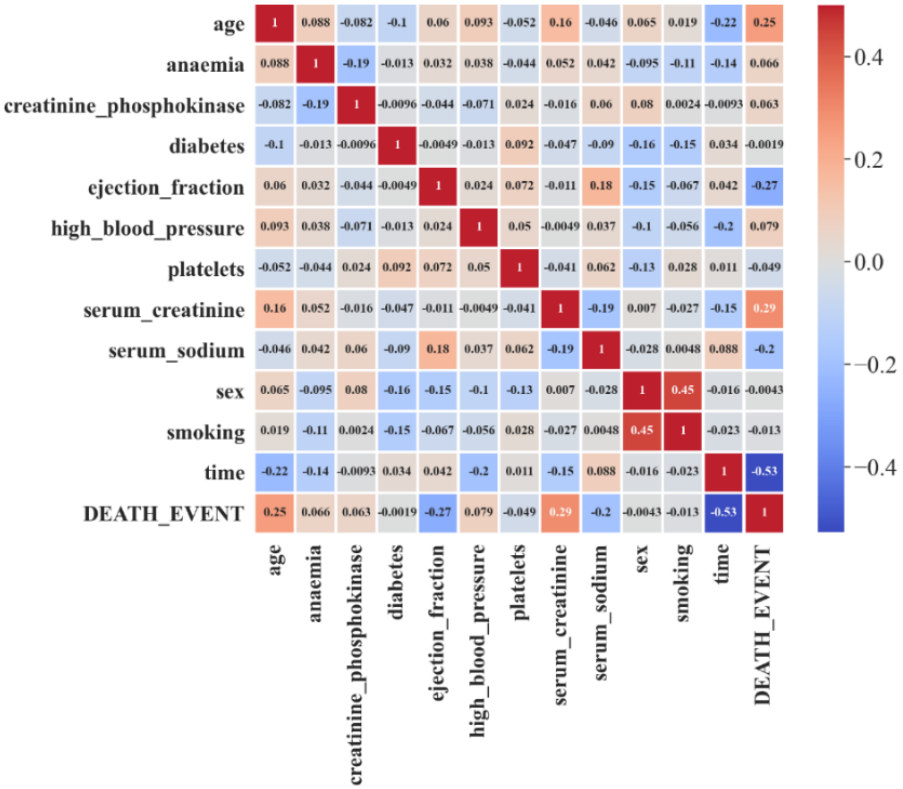


Figure 2.1: Correlation heatmap

Chapter 3

Study of Machine Learning Algorithms

Machine learning is a branch of Artificial Intelligence that investigates how algorithms perform and adapt to diverse settings over time. Experience is gained through training with a set of data known as training data. Machine learning algorithms predict or classify data without being specifically programmed to do so after going through the training process. Six supervised classification learning algorithms are chosen in this work to determine heart failure survival, and their outcomes are briefly evaluated using various criteria.

3.1 Decision Tree Classifier

The decision tree is a type of greedy algorithm that is often used in supervised learning. It can be used to solve classification and regression problems. It is a particular type of tree in which each internal node represents an attribute state. The branches reflect the condition test results, and each leaf node is pointed to as a class label. The route from root to leaf represents the classification laws. After calculating the entropy of each attribute in the data, the gain is determined. After that, the gain and entropy values are used to perform the splitting. This procedure is continued until the desired outcome is achieved.

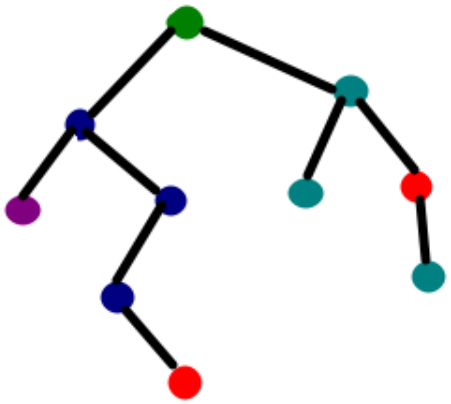


Figure 3.1: Decision Tree

3.2 Random forest Classifier

Random Forest is an algorithm based on ensemble learning trained using the bagging technique. Is a decision tree forest that can be used for classification and regression? Multiple decision trees are used to forecast the value, and the final decision is made using an averaging or voting procedure, respectively, for regression and classification problems. It is capable of processing large quantities of training data without feature scaling. It is, however, more precise and needs less training time. The accuracy is dependent on the number of trees or forest density in this algorithm. When the forest is thick, the prediction accuracy is higher.

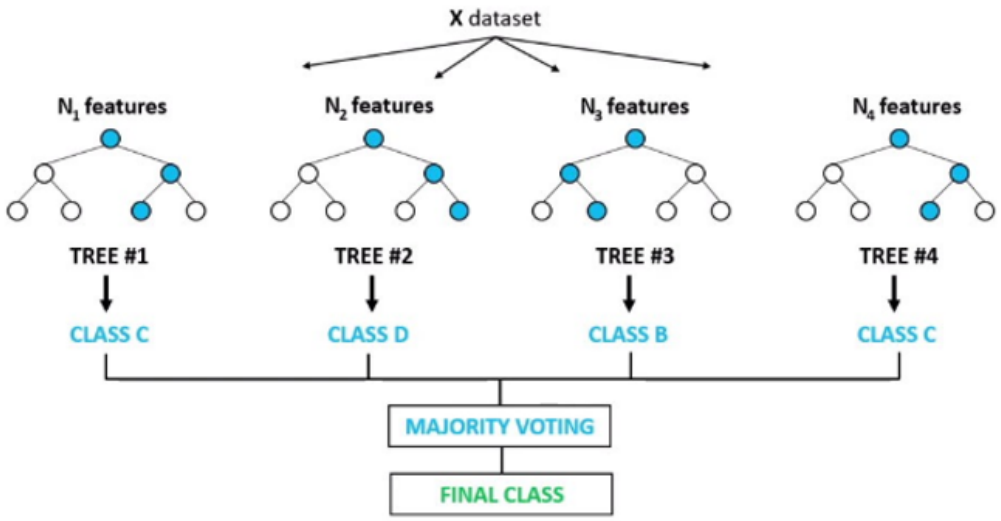


Figure 3.2: Random Forest classifier

3.3 Logistic Regression

Logistic Regression is a robust and commonly used binary classification algorithm due to its simplicity. It is similar to linear regression, except that it uses a logistic function instead of a linear function. The logistic function is represented by an S-shaped curve. The curve is fitted to the data points using maximum likelihood calculation. According to the trained model, the output is predicted to be either 0 or 1.The ROC (Receiver Operation Characteristics) curve is used to determine the optimal threshold value for logistic regression. This value decides the classification outcomes.

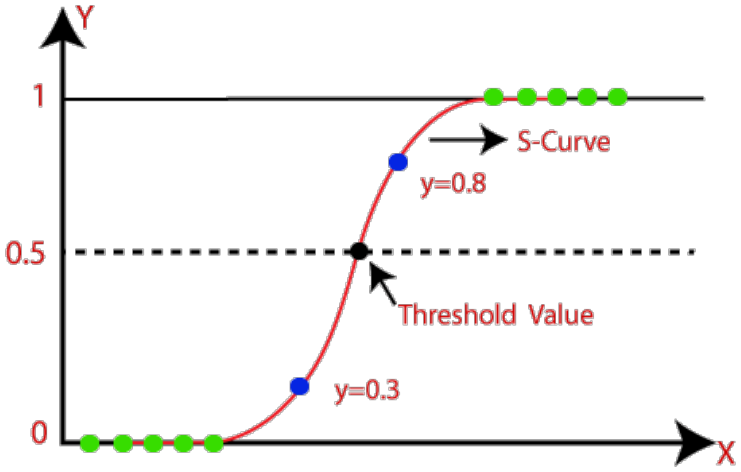


Figure 3.3: Logistic function

3.4 K Nearest Neighbors

K-Nearest Neighbor is a simple, non-parametric algorithm used to solve classification and regression problems. It computes the similarity between test and training data. It does this by calculating the distance between the given test point and the neighbors using the value of K, also known as the number of neighbors. Following that, the algorithm classifies the test data into a group with a more significant number of neighbors. This algorithm needs more time to predict since no training is performed before providing the test point. That is why the algorithm is referred to as a lazy learner algorithm. There is no set rule for determining the value of K. It can be selected by trial and error or intuitively. The lower value of K, on the other hand, could be a noisier solution for this algorithm.

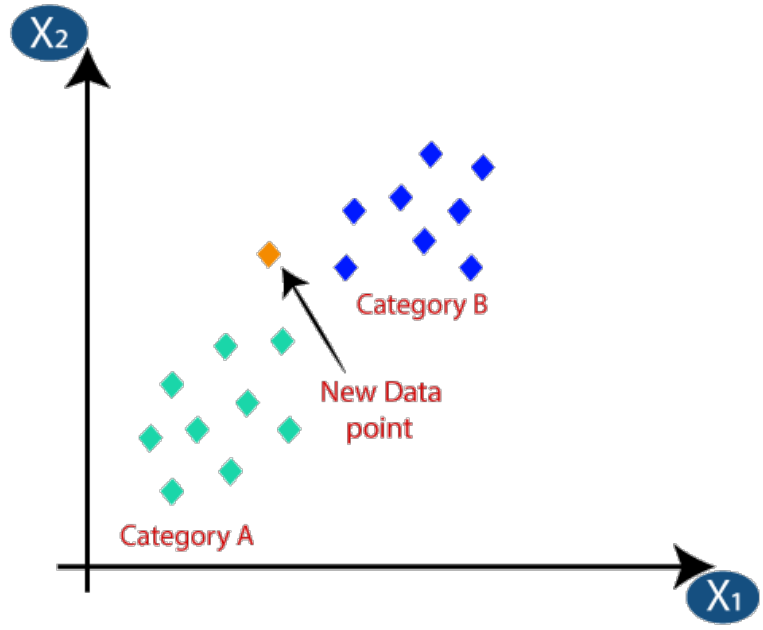


Figure 3.4: K Nearest Neighbors

3.5 Gaussian Naïve Bayes

Naïve Bayes is a classification algorithm family based on Bayes' theorem. All features are treated as independent and contributing equally to the outcome in this algorithm. It is more convenient to use because it requires little data to train the model. This probabilistic method begins with an initial guess based on the training data and then calculates the probability based on that initial guess.Gaussian Naïve Bayes classifier is a subclass of Naïve Bayes in which the attributes are continuous-valued with a Gaussian or Natural distribution. When data is plotted, a bell-shaped symmetrical curve is created.

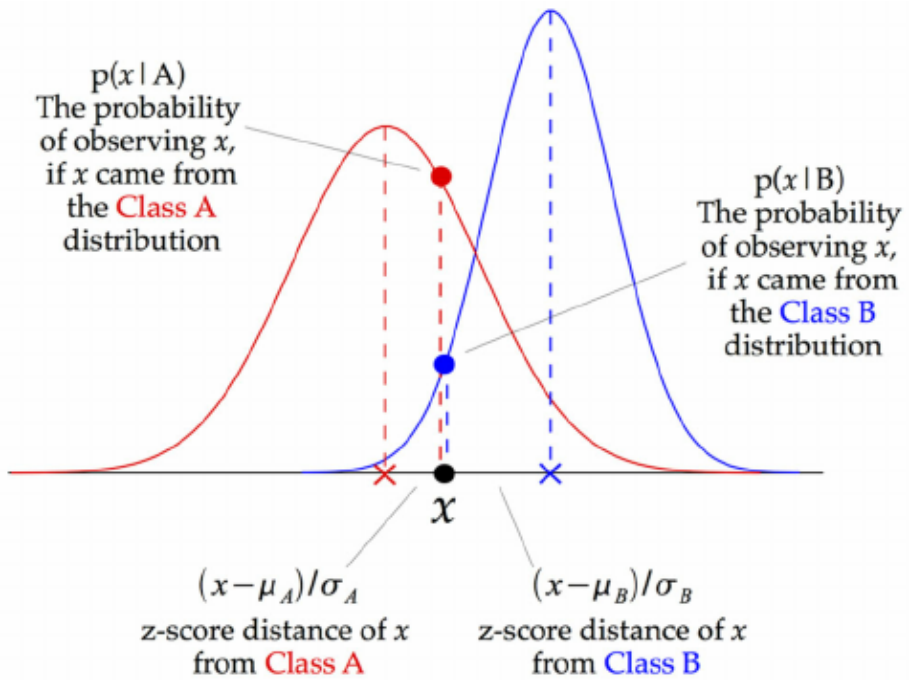


Figure 3.5: Gaussian distribution function

3.6 Support Vector Machine

Support Vector Machine (SVM) are classification and regression machine learning algorithms trained on labeled training data. The data points are classified based on their degree of similarity. This algorithm is capable of working with either linear or non-linear models. In linear SVM, a dividing plane called the hyperplane is used to distinguish between similar groups by assigning a line. The closest point to the hyperplane is the support vector, which is also the algorithm's name. The hyperplane classifies the test data into a subclass, making this a non-probabilistic binary classifier. This algorithm's primary objective is to find the optimal hyperplane for the given training data. The optimal hyperplane is accomplished by selecting the hyperplane with the most significant margin. SVM is often implemented using the kernel. The kernel simplifies the intractable problem and makes determining the hyperplane simpler.

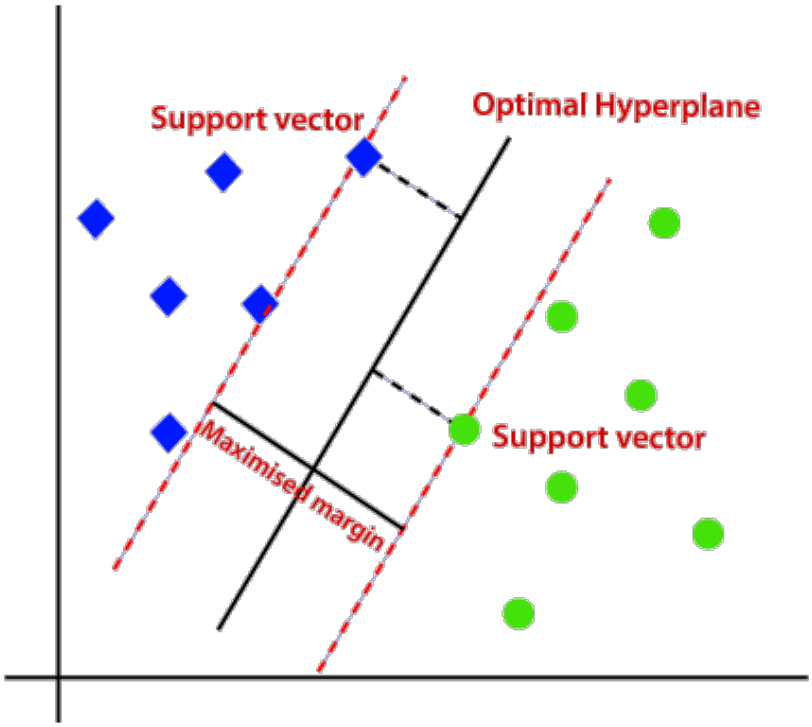


Figure 3.6: Support Vector

3.7 Hyperparameter Optimization (HPO)

Hyper-parameters are a collection of parameters that control how a machine learning algorithm learns. Optimization of hyper-parameters has the potential to significantly impact the outcome and performance of machine learning algorithms. This study employs Randomized Search Cross-Validation (RSCV) and Grid Search Cross-Validation (GSCV) to determine the optimal hyper-parameter combination. Grid search is a parameter sweep technique that evaluates all possible combinations of given parameters and returns the optimal result based on previously defined performance metrics. It takes longer and consumes more resources. On the other hand, random search chooses random combinations rather than attempting all possible combinations. It is more time and resource-efficient and is used when parameter influences on outcomes are minimal.

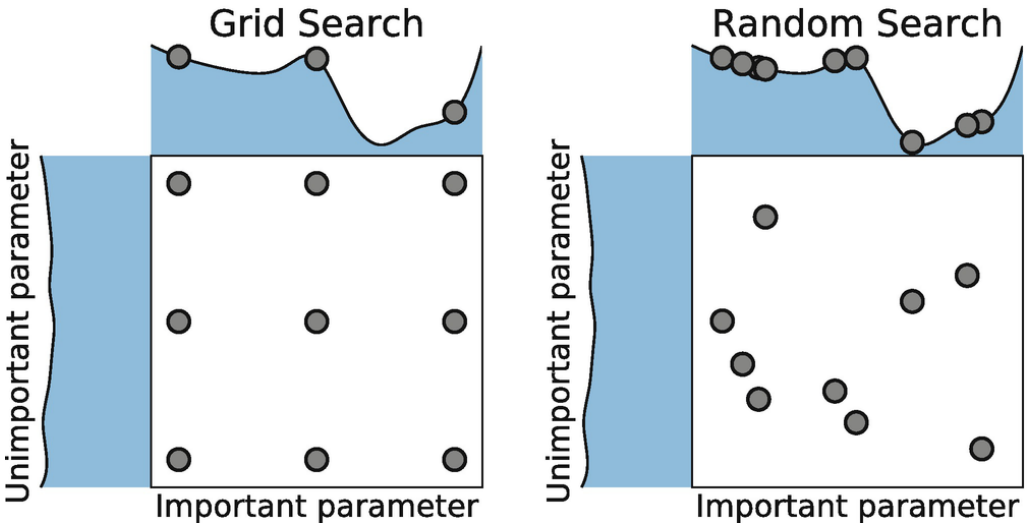


Figure 3.7: GSCV and RSCV

3.7.1 Grid Search Cross- Validation

GSCV is a parameter sweep technique that tries all the possible combinations of hyperparameters. The problem with this is that it required very high computational time. Thus, it appears to be more expensive and uses more resources

3.7.2 Random Search Cross-Validation

RSCV used only a handful of combinations of hyperparameters to find out the best among them. It is a much more efficient technique compared to GSCV. However, this approach might not find the absolute best combination of hyperparameters. Still, in real life applications, this is a widely used approach especially when the performance is less dependent on hyperparameters

3.8 Feature Scaling

The process of normalizing or standardizing independent features or variables is referred to as ”feature scaling.” This is because, regardless of their units, machine learning algorithms can give higher values more weight and lower values less weight. Standardization guarantees that specific attribute values have a mean of zero and a variance of one.

3.8.1 Min-Max Scaler

In this method, scaling is done by using the maximum and minimum value of an attribute.

$$Min - MaxScaling, x = \frac{x - min(x)}{max(x) - min(x)} \tag{3.1}$$

3.8.2 Standard Scaler

In this technique, the mean and variance of the values of the attributes are used for scaling.

$$Standardization, x = \frac{x - \bar{x}}{\sigma} \tag{3.2}$$

3.9 Sampling Method

SMOTE-ENN

Synthetic Minority Over-sampling Technique and Edited Nearest Neighbor is abbreviated as SMOTE-ENN. In an unbalanced dataset, it is a sampling technique that combines techniques of over- and under-sampling minority classes. When there is an imbalance in the distribution of classes, it is an extremely effective method, as machine learning algorithms can be biased in favor of the majority class. SMOTE-ENN uses interpolation to over-sample the minority class, then uses the ENN method to remove redundant samples. Finally, it generates data that is balanced across classes and can be used with machine learning algorithms to achieve the desired results.

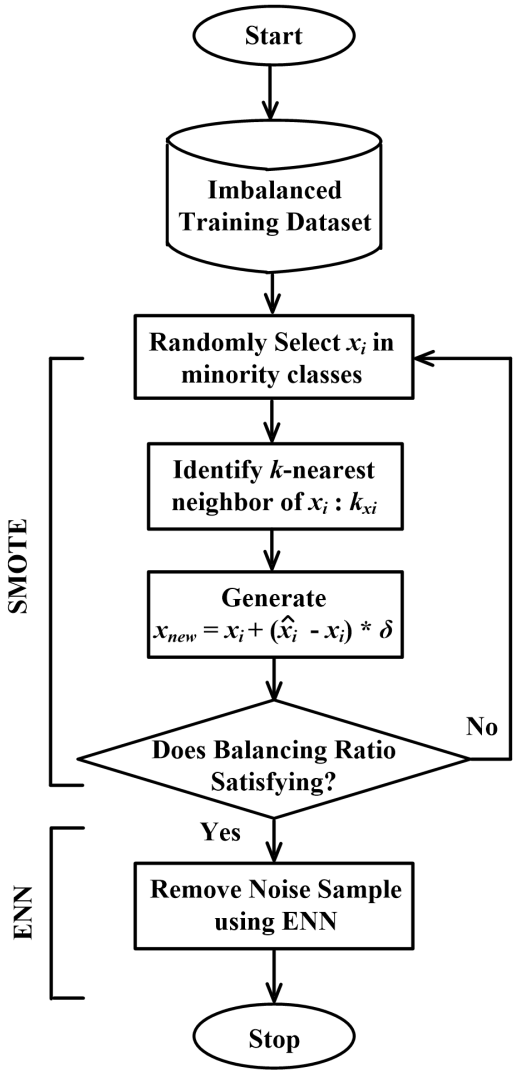


Figure 3.8: SMOTE-ENN algorithm

Chapter 4

Methodology

Three different approaches are taken into consideration in order to examine the performance of six widely used supervised machine learning models, namely the Decision Tree Classifier (DTC), Logistic Regression (LR), Gaussian Naive Bayes (GNB), Random Forest Classifier (RFC), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) (SVM). The following methodologies are described in detail, along with a flowchart depicting the related workflow.

4.1 Workflow

4.1.1 Approach A

Firstly, six algorithms are used to build machine learning models, which are then trained and validated using the default data distribution with no preprocessing. K-fold cross validation was used to evaluate the performance metrics. All default hyperparameters are used in this method. The workflow diagram is illustrated in Fig 4.1

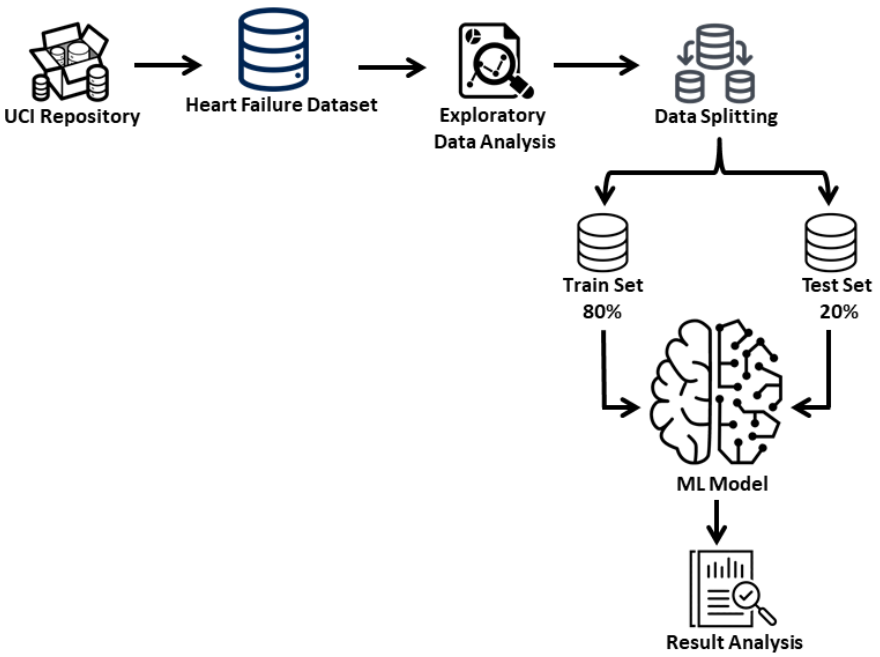


Figure 4.1: Workflow diagram of Approach A

4.1.2 Approach B

Secondly, Grid Search Cross-Validation (GSCV) and Random Search Cross-Validation (RSCV) are used to perform hyperparameter optimization (HPO) (RSCV). The dataset is not class balanced, and data scaling is accomplished with the use of Min-Max and Standard Scalar methods. The dataset was then five and tenfold cross-validated, and the best hyperparameters were found and used to evaluate our model. The workflow is illustrated in Fig 4.2.

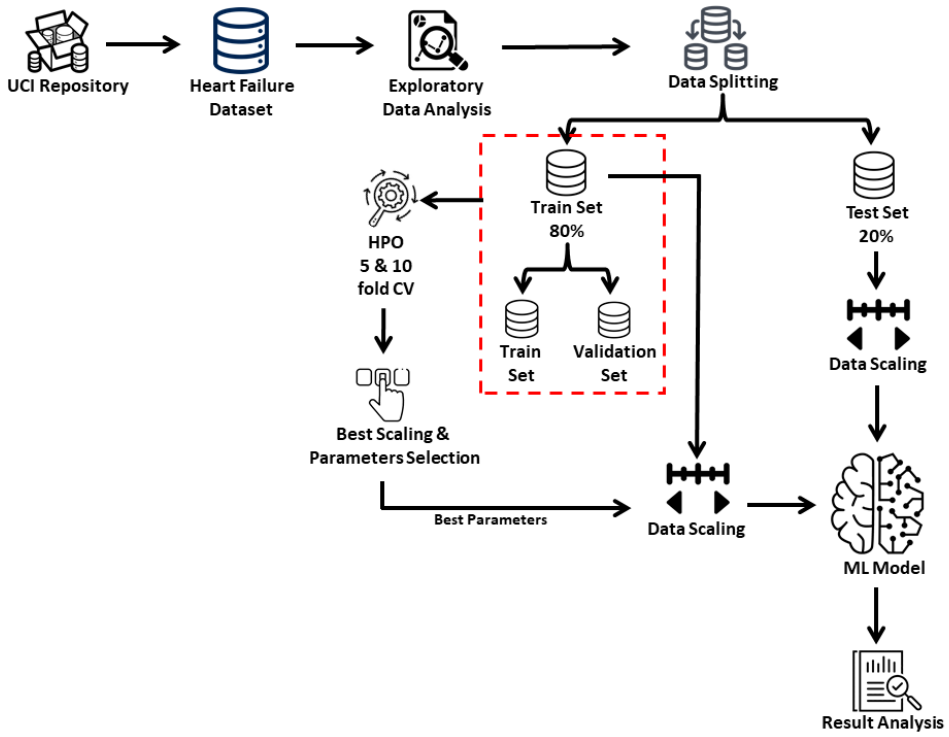


Figure 4.2: Workflow diagram of Approach B

4.1.3 Approach C

Finally, To balance the class distributions, the dataset was re-sampled after data scaling using SMOTE-ENN (a technique that combines up- and down-sampling of minority classes). The dataset was then split into test and train sets using the k-fold method in five and ten-fold, resulting in an 80:20 and 90:10 (test/train) split, respectively. Randomized Search Cross-Validation (RSCV) and Grid Search Cross-Validation (GSCRV) were used to find the best combination of algorithms' hyperparameters, which were then used to evaluate the algorithms' performance metrics. Following the completion of the experiments using three different methodologies, a qualitative and quantitative analysis was conducted, and the procedures were compared in order to choose the best model for this project.The workflow is illustrated in Fig 4.3.

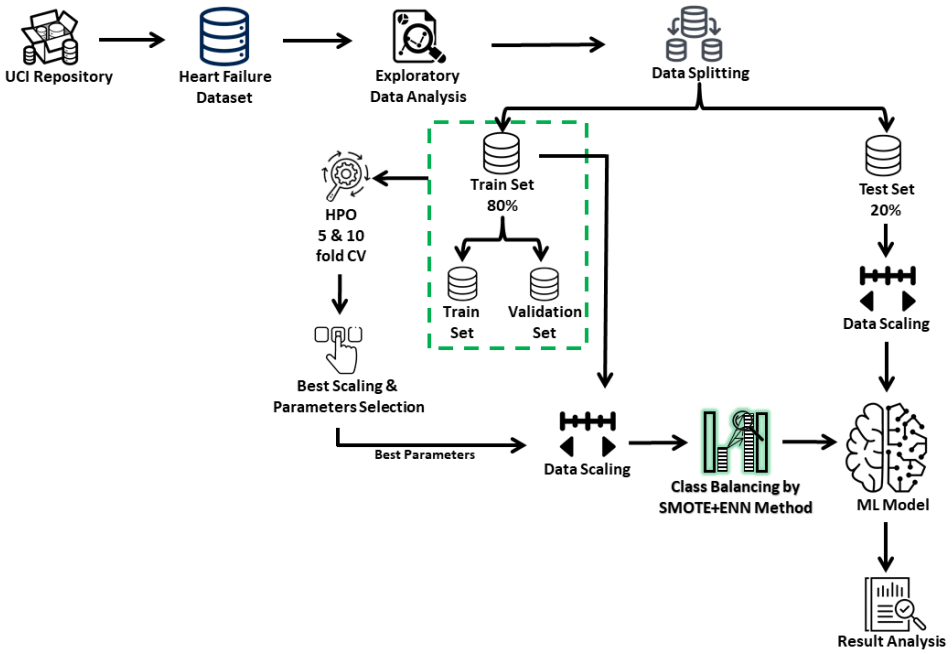


Figure 4.3: Workflow diagram of Approach

4.2 Experimental Environment

Jupyter Notebook v6.1.4 (python 3.8.5) and Anaconda distribution v4.10.3 on Intel Core i5-8300H CPU running at 2.30GHz, 8GB of RAM, and an NVIDIA GTX 1050 Ti graphics unit with 4GB of dedicated memory for the experiment.

Chapter 5

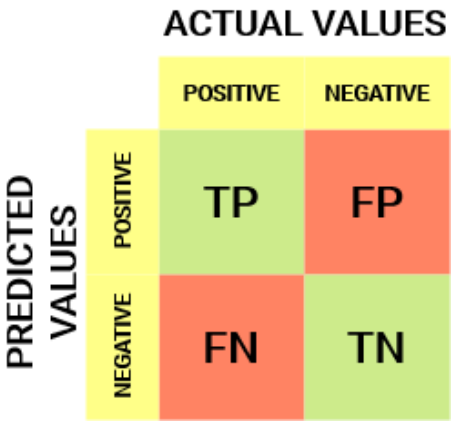
Result Analysis

5.1 Approach A

In this approach, the Random Forest Classifier (RFC) outperforms in the majority of performance metrics, as shown in Table 5.7, with accuracy, recall, and ROC AUC values of 0.8,0.8537, and 0.7689, respectively. The GNB and SVM, on the other hand, maximize precision, while the LR algorithm measures F1 score the best. The Decision Tree Classifier (DTC) came in second in terms of accuracy with a score of 0.7333, while LR and RFC came in first with a score of 0.8. In Fig. 5.1, bar charts show how algorithms compare in terms of performance metrics. As seen in Fig. 5.2 and the ROC curve, the LR and RFC algorithms work well in this manner.

5.1.1 Confusion Matrices

A confusion matrix can be used to summarize how well a classification algorithm performs. When each class has a different number of observations or a dataset has more than two classes, classification accuracy alone can be misleading. A confusion matrix can aid in determining where your classification model succeeds and fails.
A confusion matrix is a N x N matrix that is used to evaluate the performance of a classification model, with N denoting the number of target classes. The matrix compares the target values to the predictions of the machine learning model. This demonstrates how well our classification model performs as well as the types of errors it makes.



True Positive (TP)

The total number of correct results or predictions when the actual class was positive is known as the true positive rate (TP).

True Negative (TN)

The total number of correct predictions or results when the actual class was negative is known as the true negative rate (TN).

False Positive (FP) – Type 1 error

The total number of incorrect results or predictions when the actual class was positive is known as the false positive rate (FP).

False Negative (FN) – Type 2 error

The total number of incorrect results or predictions when the actual class was negative is referred to as the false negative rate (FN).

Table 5.1: Confusion matrix of Decision Tree Classifier

Decision Tree Classifier (DTC)		Predicted Positive	Predicted Negative
Actual	True	31	13
	False	10	6

Table 5.2: Confusion matrix of Logistic Regression

Logistic Regression (LR)		Predicted Positive	Predicted Negative
Actual	True	36	12
	False	5	7

Table 5.3: Confusion matrix of Gaussian Naive Bayes

Gaussian Naive Bayes (GNB)		Predicted Positive	Predicted Negative
Actual	True	41	0
	False	0	19

Table 5.4: Confusion matrix of Random Forest Classifier

Random Forest Classifier (RFC)		Predicted Positive	Predicted Negative
Actual	True	35	13
	False	6	6

Table 5.5: Confusion matrix of K-Nearest Neighbors

Random Forest Classifier (KNN)		Predicted Positive	Predicted Negative
Actual	True	37	3
	False	4	16

Table 5.6: Confusion matrix of Support Vector Machine

Support Vector Machine (SVM)		Predicted	
		Positive	Negative
Actual	True	41	0
	False	0	19

5.1.2 Performance Metrics

The majority of model-performance metrics are calculated by comparing a model's predictions to the (known) values of a dataset's dependent variable. An ideal model would have the same predictions and dependent-variable values.
Machine Learning algorithms are evaluated using a variety of performance metrics. We can use classification performance metrics like Precision, Accuracy, AUC, Recall and others to distinguish between images of different objects. The root mean squared error (RMSE) can be used to assess the efficacy of a machine learning model attempting to forecast a stock's price. Precision recall, or NDCG, is another metric for evaluating machine learning algorithms. It can be used for sorting algorithms commonly used by search engines. As a result, different metrics are required to evaluate the efficiency of various algorithms depending on the dataset.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.3}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{5.4}$$

Table 5.7: Performance metrics of ML algorithms by Approach A

Algorithms	Accuracy	Precision	F1	Recall	ROC_AUC
DTC	0.733	0.756	0.795	0.838	0.720
LR	0.800	0.878	0.857	0.837	0.755
GNB	0.683	1.000	0.812	0.683	0.500
RFC	0.800	0.854	0.854	0.854	0.769
KNN	0.667	0.902	0.787	0.698	0.530
SVM	0.683	1.000	0.812	0.683	0.500

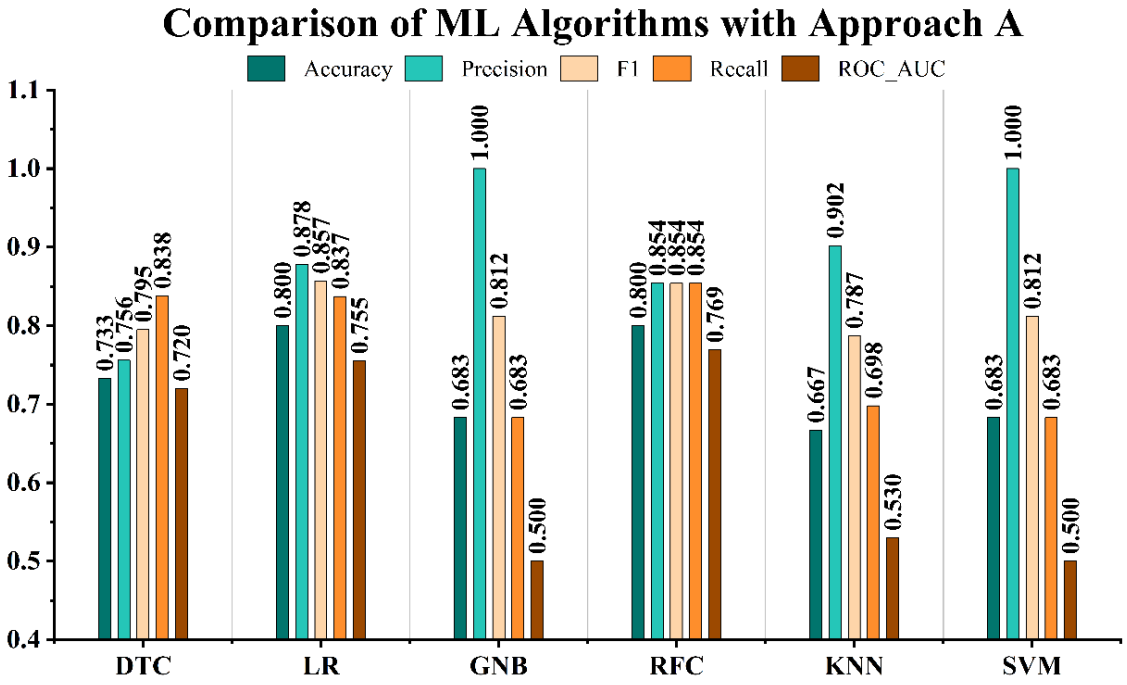


Figure 5.1: Comparative analysis of the performance metrics by Approach A

5.1.3 Receiver Operating Characteristic Curve (ROC)

The Receiver Operating Characteristics Curve, or ROC curve, is a metric for evaluating the performance of a classifier. The ROC curve illustrates the ratio of true positives to false positives, emphasizing the classifier model's sensitivity. Because it compares two operating characteristics, the True Positive Rate and the False Positive Rate, as the criterion changes, the ROC is also referred to as a relative operating characteristic curve. A perfect classifier would have a ROC with a true positive rate of 100and no false positives. When the rate of false positives rises, we usually track how many correct positive classifications are gained.

One of the most widely used model evaluation metrics is the Area Under Curve (AUC). It's frequently used to solve binary classification problems. The AUC is the two-dimensional area beneath the ROC curve. The probability that a randomly selected positive example will be ranked higher than a randomly selected negative example is the AUC of a classifier. The AUC is a summary of a classifier's ROC curve, which is used to distinguish between classes. The AUC measures the model's ability to distinguish between positive and negative classes. The better, the higher the AUC.

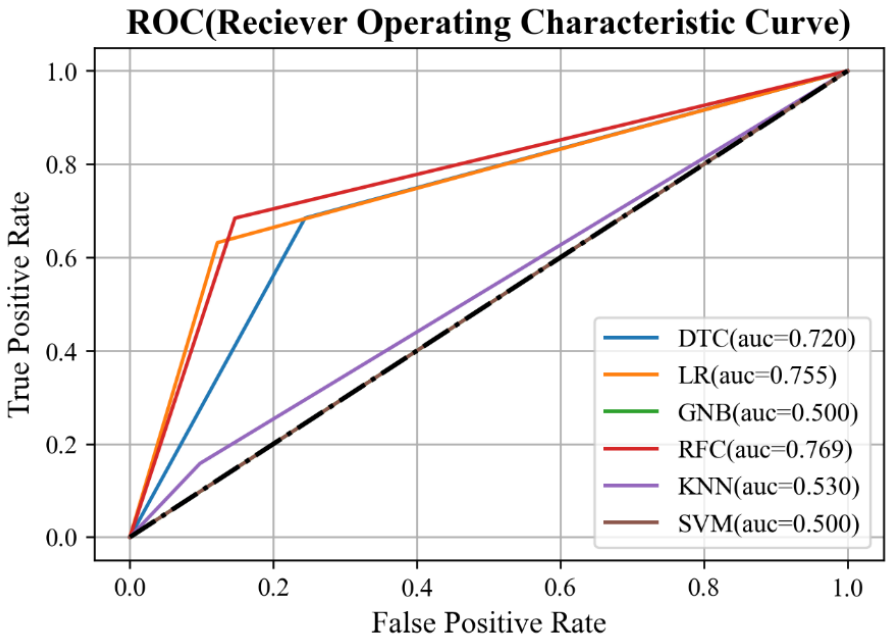


Figure 5.2: ROC curve for all the ML models for Approach A

5.2 Approach B

The high variation between the numeric values of the various attributes can lead ML systems to bias toward large values, as seen in table 2.2. Because it tries random combinations of hyper-parameters rather than all options, RSCV takes substantially less time than GSCV, as seen in Table 5.2.3. However, GSCV outperformed RSCV in terms of accuracy. A bar plot in Figure 5.3 depicts the contrast in computational time. The results of our eight studies are summarized in Table 5.2.2. The classifier with the highest accuracy was determined after eight different combinations of scaling and cross-validation methods were used. The GSCV and standard scaling procedures have clearly delivered the best results, as seen in the table 5.2.2. According to a 10-fold GSCV with standard scaling, RFC delivers the best result, with an accuracy of 0.8703. However, when using a 10-fold GSCV with standard scaling, the best overall performance was observed. As a result, Table 5.7 uses this combination to evaluate all performance measures. RFC outperforms all other algorithms in terms of accuracy, recall, and ROC AUC in this scenario. However, LR has the greatest F1 score as well as the most accuracy. In this method, GNB maximizes the precision. RFC and LR determined that the best accuracy is 0.8333, and SVM determined that the second-best accuracy is 0.8. A comparison of all algorithms based on performance metrics is shown in Figure 5.4. This technique yields a considerably better result than Approach A.

5.2.1 Confusion Matrices

Table 5.8: Confusion matrix of Decision Tree Classifier

Decision Tree Classifier (DTC)		Predicted	
		Positive	Negative
Actual	True	37	9
	False	4	10

Table 5.9: Confusion matrix of Logistic Regression

Logistic Regression (LR)		Predicted Positive	Predicted Negative
Actual	True	38	12
	False	3	7

Table 5.10: Confusion matrix of Gaussian Naive Bayes

Gaussian Naive Bayes (GNB)		Predicted Positive	Predicted Negative
Actual	True	41	0
	False	0	19

Table 5.11: Confusion matrix of Random Forest Classifier

Random Forest Classifier (RFC)		Predicted Positive	Predicted Negative
Actual	True	37	13
	False	4	6

Table 5.12: Confusion matrix of K-Nearest Neighbors

Random Forest Classifier (KNN)		Predicted Positive	Predicted Negative
Actual	True	33	9
	False	8	10

Table 5.13: Confusion matrix of Support Vector Machine

Support Vector Machine (SVM)		Predicted Positive	Predicted Negative
Actual	True	36	12
	False	5	7

5.2.2 Hyperparameter Optimization and Data Preprocessing

Table 5.14: Highest accuracies of classifiers in conducted eight experiments

Experiment no.	Preprocessing method	K fold CV	HPO method	Highest accuracy classifier	Highest accuracy
1	Standard scalar	5-fold	Random search	LR	0.845
2	Standard scalar	10-fold	Random search	RFC	0.854
3	Min–max scalar	5-fold	Random search	RFC	0.850
4	Min–max scalar	10-fold	Random search	RFC	0.858
5	Standard scalar	5-fold	Grid search	RFC	0.866
6	Standard scalar	10-fold	Grid search	RFC	0.870
7	Min–max scalar	5-fold	Grid search	RFC	0.862
8	Min–max scalar	10-fold	Grid search	RFC	0.866

5.2.3 Computational Time

Table 5.15: Comparison of computational time

Algorithms	Computation time Grid search CV (sec)	Computation time Random search CV (sec)
DTC	11.763	0.285
LR	4.207	0.244
GNB	0.140	0.140
RFC	88.745	6.175
KNN	4.160	0.598
SVM	3.955	1.011

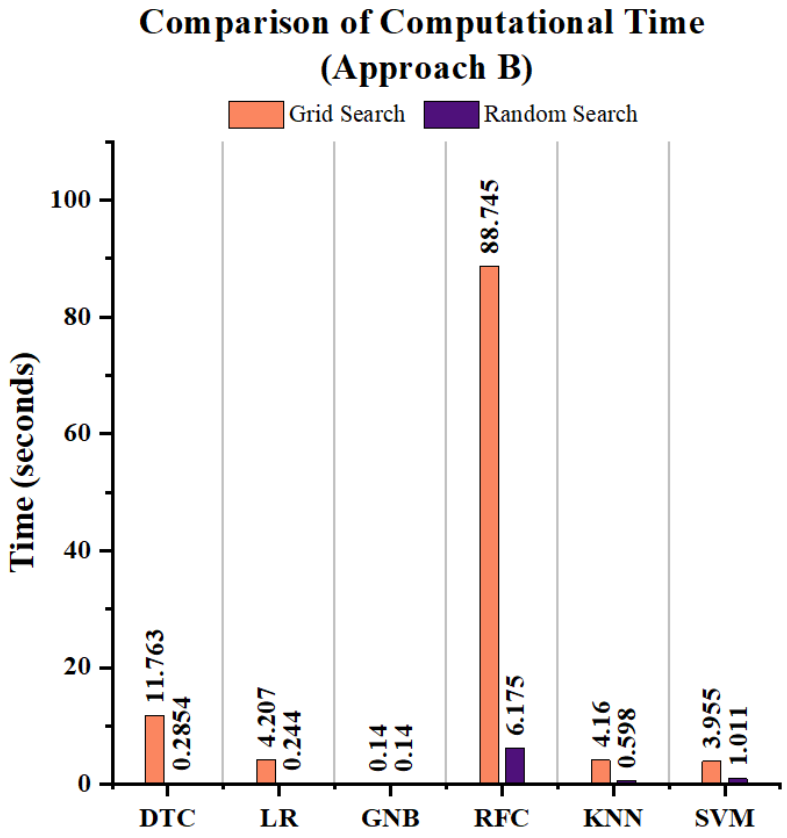


Figure 5.3: Comparison of computational time by Approach B

5.2.4 Performance Metrics

Table 5.16: Performance metrics of ML algorithms by Approach B

Algorithms	Accuracy	Precision	F1	Recall	ROC_AUC
DTC	0.767	0.902	0.841	0.787	0.688
LR	0.833	0.927	0.884	0.844	0.779
GNB	0.683	1.000	0.812	0.683	0.500
RFC	0.833	0.902	0.881	0.860	0.793
KNN	0.700	0.805	0.786	0.767	0.639
SVM	0.800	0.878	0.857	0.837	0.755

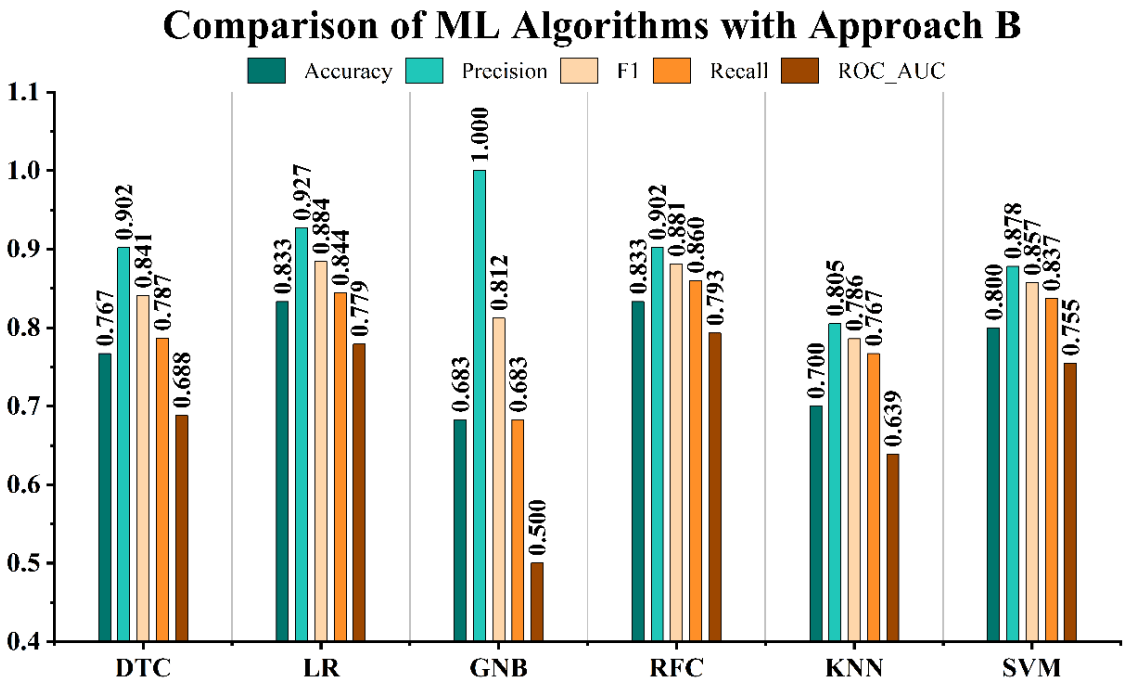


Figure 5.4: Comparative analysis of the performance metrics by Approach B

5.2.5 Receiver Operating Characteristic Curve (ROC)

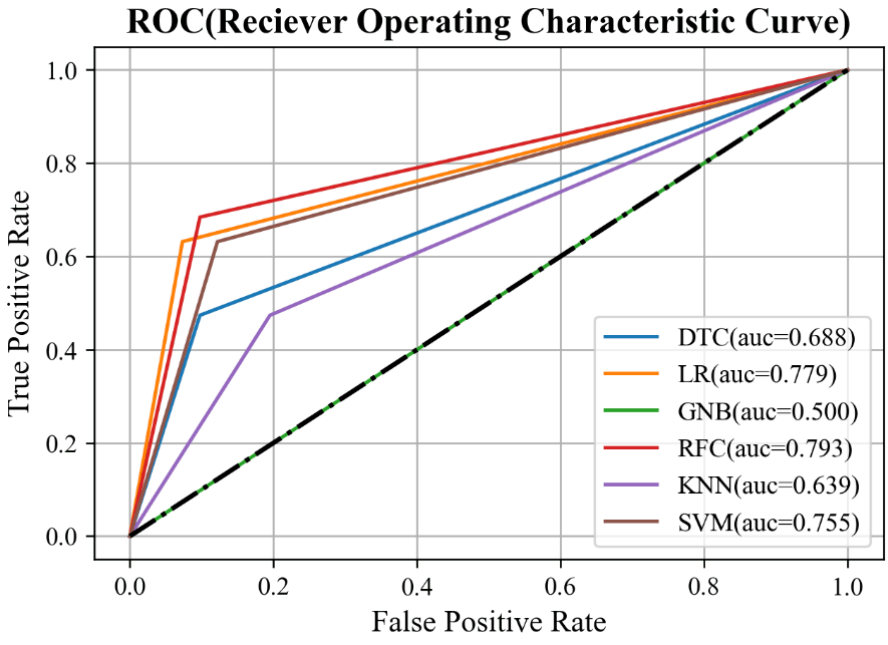


Figure 5.5: ROC curve of all the ML models for Approach B

5.3 Approach C

Because this dataset was significantly imbalanced, with a class ratio of 203:96, one class is nearly double that of the other, this strategy incorporates a class balancing technique called SMOTE-ENN. This kind of imbalance can cause machine learning algorithms to fail, and there's a propensity for the majority class in the prediction to win. Researchers employed sampling approaches like SMOTE and SMOTE-ENN to balance class in this sort of dataset to correct the imbalance and improve the outcomes. In this study, SMOTE-ENN is combined with scaling and hyperparameter optimization (HPO), resulting in more promising results. The calculation times for these approaches are compared in Table 5.3.3, and it is evident that GSCV takes longer than RSCV, but GSCV delivers superior accuracy, as shown visually in Fig. 5.3.3.

The SVM delivers the most accuracy with a value of 0.9889, which is the best result of the three approaches in terms of accuracy. The most accurate method is now 10-fold GSCV and standard scalar. A summary of the experiments is shown in Table 5.3.2. The performance metrics for this method were then assessed using a test dataset and parameters from a conventional scalar 10-fold GSCV, as shown in Table 5.3.4. The RFC is the most accurate, with a score of 0.9, followed by DTC at 0.8667. In terms of F1 score, recall, and ROC AUC, the KNN is the clear winner. The DTC has the best precision. This method produces significantly better outcomes than the other two. A comparison of algorithms utilizing technique C is shown in Figure 5.7.

5.3.1 Confusion Matrices

Table 5.17: Confusion matrix of Decision Tree Classifier

Decision Tree Classifier (DTC)		Predicted	
		Positive	Negative
Actual	True	38	14
	False	3	5

Table 5.18: Confusion matrix of Logistic Regression

Logistic Regression (LR)		Predicted	
		Positive	Negative
Actual	True	29	17
	False	12	2

Table 5.19: Confusion matrix of Gaussian Naive Bayes

Gaussian Naive Bayes (GNB)		Predicted	
		Positive	Negative
Actual	True	33	13
	False	8	6

Table 5.20: Confusion matrix of Random Forest Classifier

Random Forest Classifier (RFC)		Predicted	
		Positive	Negative
Actual	True	36	18
	False	5	1

Table 5.21: Confusion matrix of K-Nearest Neighbors

Random Forest Classifier (KNN)		Predicted	
		Positive	Negative
Actual	True	32	15
	False	9	4

Table 5.22: Confusion matrix of Support Vector Machine

Support Vector Machine (SVM)		Predicted	
		Positive	Negative
Actual	True	21	16
	False	10	3

5.3.2 Hyperparameter Optimization and Data Preprocessing

Table 5.23: Highest accuracies of classifiers in conducted eight experiments

Experiment no.	Preprocessing method	K fold CV	HPO method	Highest accuracy classifier	Highest accuracy
1	Standard scalar	5-fold	Random search	SVM	0.970
2	Standard scalar	10-fold	Random search	SVM	0.970
3	Min–max scalar	5-fold	Random search	RFC	0.987
4	Min–max scalar	10-fold	Random search	RFC	0.980
5	Standard scalar	5-fold	Grid search	KNN	0.986
6	Standard scalar	10-fold	Grid search	SVM	0.989
7	Min–max scalar	5-fold	Grid search	RFC	0.981
8	Min–max scalar	10-fold	Grid search	KNN	0.983

5.3.3 Computational Time

Table 5.24: Comparison of computational time

Algorithms	Computation time Grid search CV (sec)	Computation time Random search CV (sec)
DTC	11.428	0.310
LR	4.078	0.315
GNB	0.355	0.148
RFC	65.935	3.558
KNN	4.225	0.578
SVM	2.233	0.880

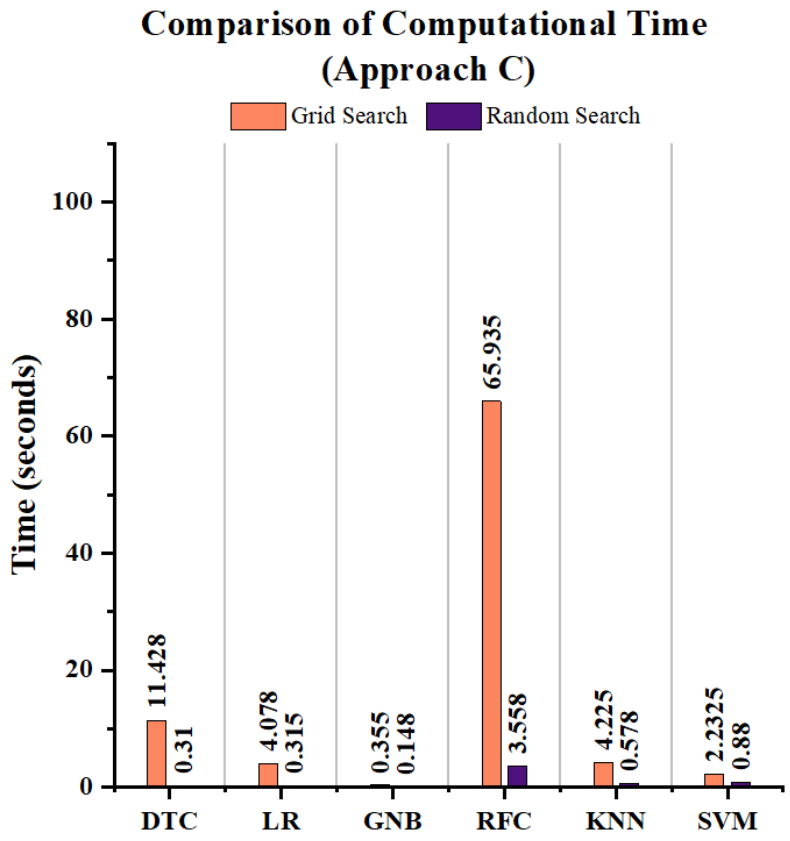


Figure 5.6: Comparison of computational time by Approach C

5.3.4 Performance Metrics

Table 5.25: Performance metrics of ML algorithms by Approach C

Algorithms	Accuracy	Precision	F1	Recall	ROC_AUC
DTC	0.867	0.927	0.905	0.884	0.832
LR	0.767	0.707	0.806	0.936	0.801
GNB	0.767	0.805	0.825	0.846	0.745
RFC	0.900	0.878	0.923	0.973	0.913
KNN	0.783	0.781	0.831	0.889	0.785
SVM	0.783	0.756	0.827	0.912	0.799

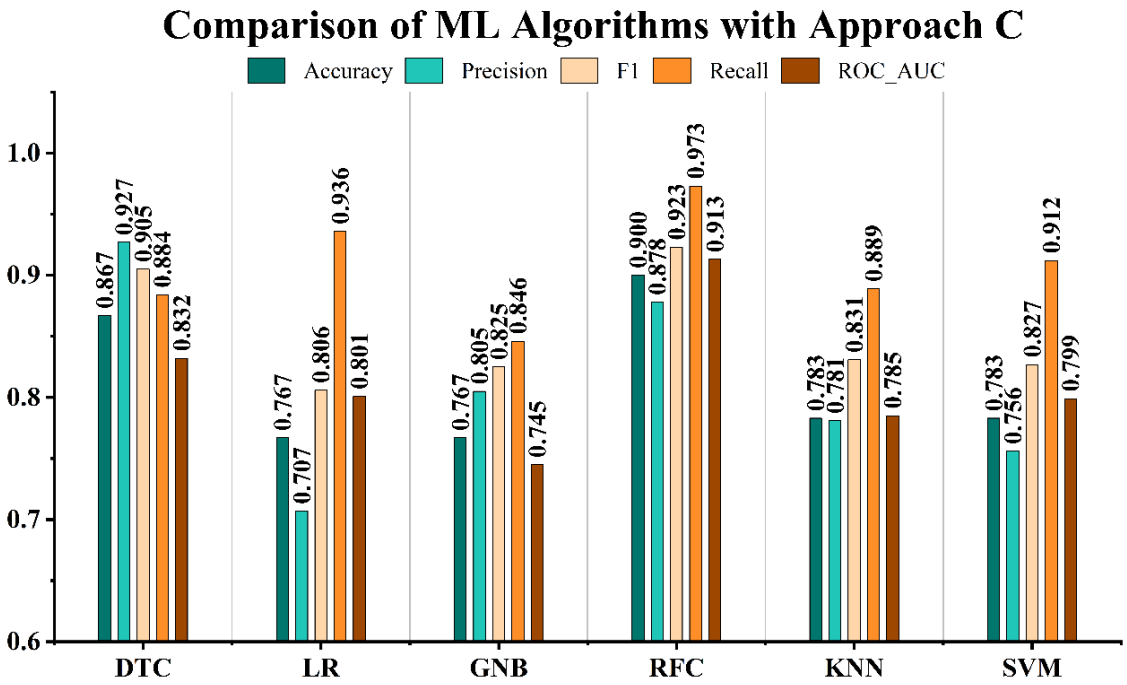


Figure 5.7: Comparative analysis of the performance metrics by Approach C

5.3.5 Receiver Operating Characteristic Curve (ROC)

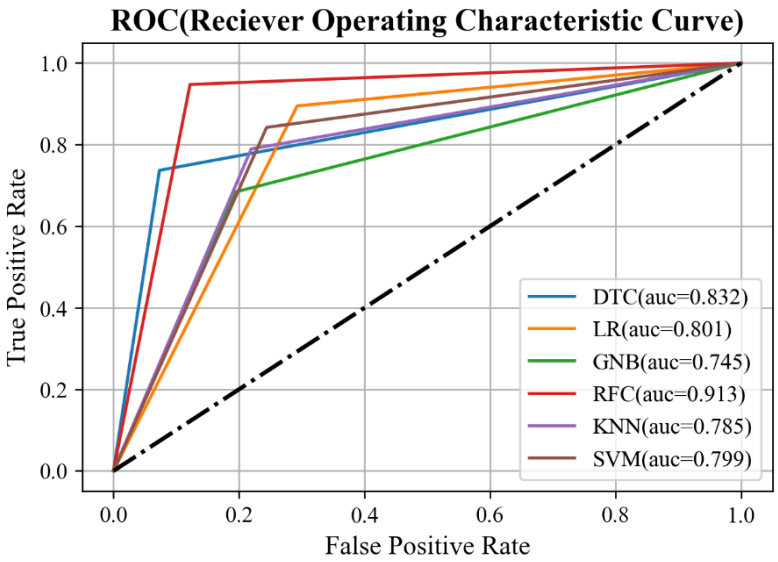


Figure 5.8: ROC curve of all the ML models for Approach C

5.4 Comparison Between Approaches

Approach C outperforms all other experimental approaches on all performance parameters. The highest accuracy was 0.8 in Approach A, 0.8333 in Approach B, and 0.9 in Approach C, suggesting that the models were improving over time. The final Approach C did an outstanding job of forecasting the survival of heart failure patients. The comparison of three techniques based on accuracy, precision, F1 score, recall, and ROC AUC is shown in Figure 5.9, 5.10, 5.11, 5.12, 5.13.

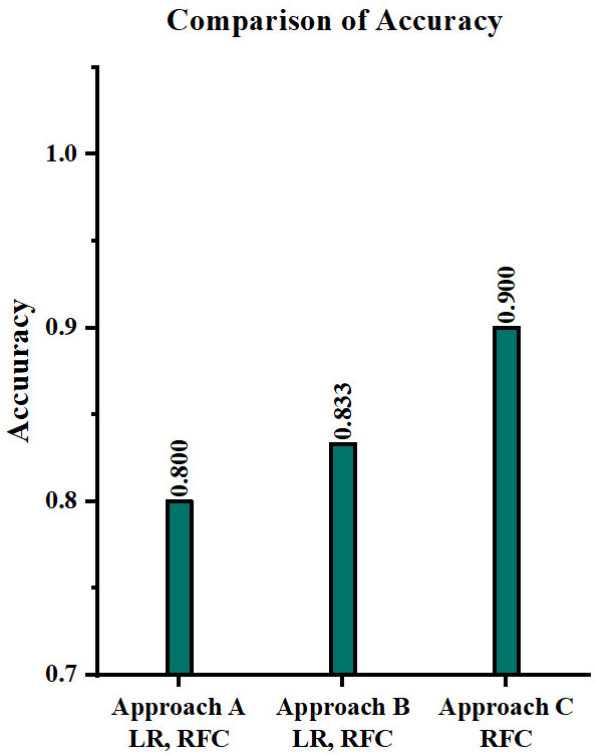


Figure 5.9: Comparison of accuracy

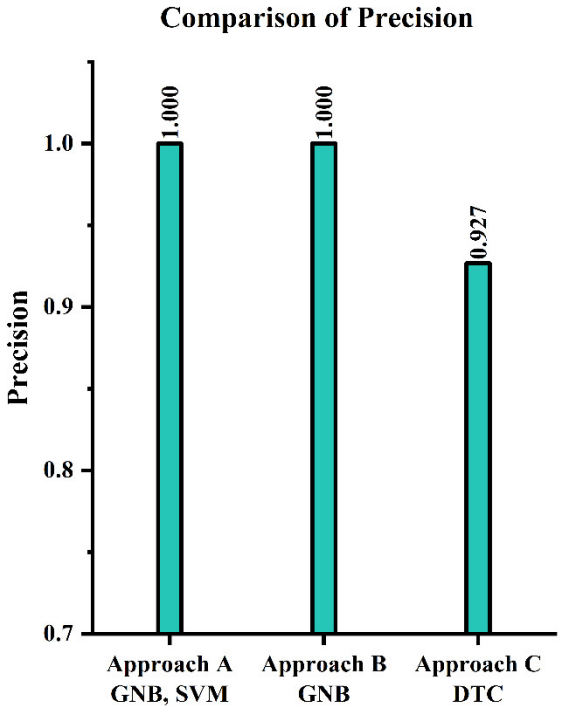


Figure 5.10: Comparison of precision

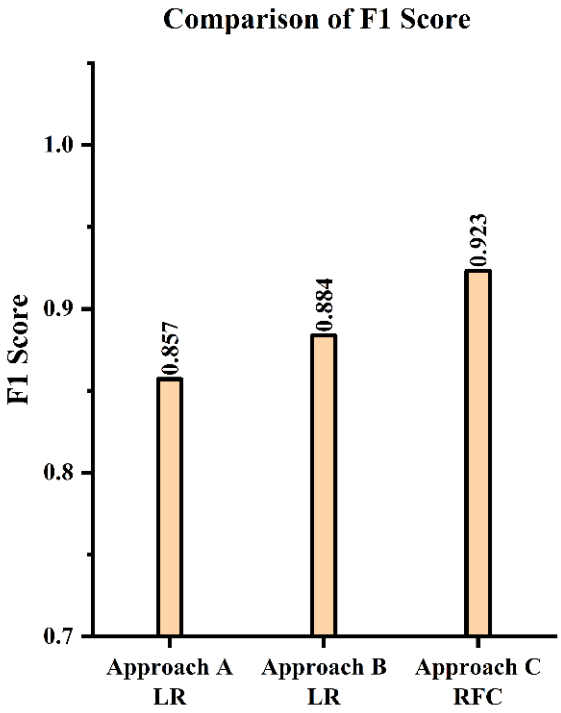


Figure 5.11: Comparison of F1

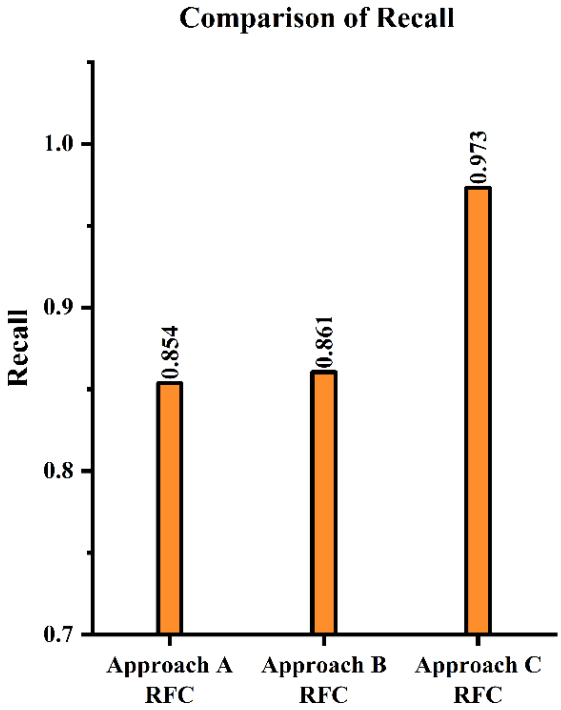


Figure 5.12: Comparison of recall

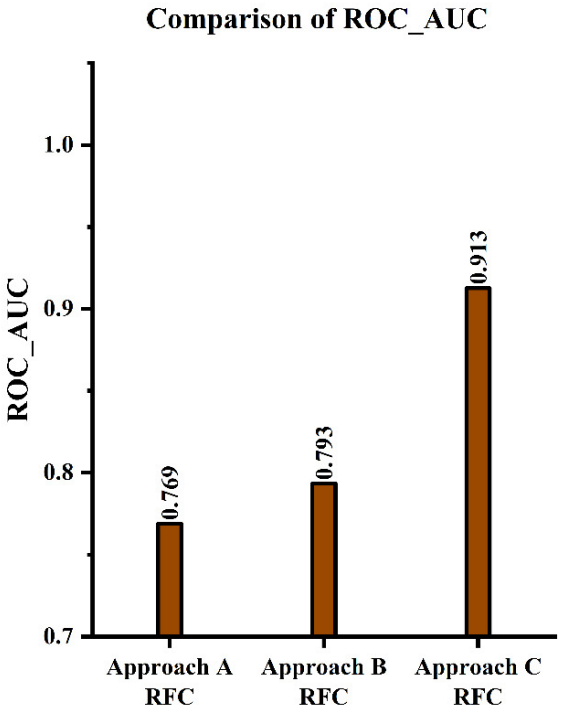


Figure 5.13: Comparison of ROC-AUC

Table 5.26: Comparison of performance with other works

References	Authors	Algorithms	Best accuracy (validation set)	Best accuracy (test set)
1	O. O. Oladimeji and O.Oladimeji	Random Forest Classifier (RFC)	83.1788	—
2	D. Chicco and G. Jurman	Logistic Regression (LR)	83.80	—
3	P. E. Rubini and C. A.Subasini	Random Forest Classifier (RFC)	84.81	—
4	R. G □urfidan and M. Ersoy	Support Vector Machine (SVM)	90.00	—
5	S. Elyassami and A. A.Kaddour	Stochastic gradient descent with chi-squaretest	91.43	—
6	A. Ishaq et al.	Extra Tree Classifier (ETC) with SMOTE	92.62	—
7	L. Ali et al.	Gaussian Naive Bayes (GNB) with $\chi 2$ statistical model	93.33	—
8	S. Rahayu and J. JayaPurnama	Random Forest Classifier (RFC) with SMOTE	94.31	—
This work		KNN SVM with SMOTE-ENN	98.9(SVM)	90(RFC)

Chapter 6

Conclusion and Future Works

Patients must seek medical guidance on a frequent basis since heart failure is such a fatal ailment, and because prevention is so important in the treatment of the condition. A range of events and features should be taken into consideration by healthcare professionals while counseling or treating their patients. The lack of diagnostic equipment and medical technologists makes it difficult to establish an accurate and timely diagnosis of a patient's condition in many parts of the world. Massive amounts of data are collected and processed for real-world patient scenarios in order to provide support to healthcare professionals. For detecting hidden patterns in large clinical datasets, machine learning and data mining offer immense potential. It is possible to use these patterns to aid doctors in the diagnosis of their patients. When compared to statistics, it is a more advanced and efficient technique for evaluating big volumes of data since it allows for prediction based on previous cases and helps healthcare practitioners to make informed judgments based on that information. There are three different methodologies being used in this study to examine the usefulness of machine learning models in predicting the survival of heart failure patients for the goals of this research.

Compared to the other two ways, Approach C exceeds them in terms of accuracy, F1 score, recall, and area under the curve (ROC AUC). Therefore, SMOTE-ENN and hyperparameter tweaking resulted in a significant improvement in the overall performance of the classifiers when used together. Approach C has the highest accuracy rate of 90 percent among the methods, followed by Approaches A and B, which both have accuracy rates of 80 percent and 83.33 percent, respectively, among the approaches. When compared to the other approaches, Strategy C had the highest F1 score (0.9231), recall (0.973), and ROC AUC of the entire group (0.973). (0.9127). Approaches A and B produced F1 scores of 0.8571 and 0.8837, recall values of 0.8537 and 0.8605, and ROC AUC values of 0.7689 and 0.7933, respectively, whereas Approach C produced F1 scores of 0.8571 and 0.8837, recall values of 0.8537 and 0.8605, and ROC AUC values of 0.7689 and 0.7933, respectively. As a result, RFC outperforms all other strategies on the test dataset, with RFC (in combination with the SMOTE-ENN algorithm and hyperparameter optimization) achiev-

ing 90 percent accuracy on the dataset.

The results of this research may be used to develop an automated computer-aided diagnosis system for use in e-healthcare applications, as well as to calculate heart failure patient survival rates.Eventually, we hope to make our model more interactive by taking live input from a device rather than from a dataset, via an application or a graphical user interface, and producing a forecast based on the trained model in real time. We also intend to use this model in Bangladeshi hospitals so that we can anticipate the danger of heart failure based on patient data and take the required procedures to avoid it.

Bibliography

[1] S. Elyassami and A. A. Kaddour, "Implementation of an incremental deep learning model for survival prediction of cardiovascular patients," *IAES International Journal of Artificial Intelligence (IJ-AI*, vol. 10, no. 1, pp. 101–109, 2021.

[2] S. Rahayu, J. Jaya Purnama, A. Baroqah Pohan, F. Septia Nugraha, S. Nurdiani, and S. Hadianti, "PREDICTION OF SURVIVAL OF HEART FAILURE PATIENTS USING RANDOM FOREST," Tech. Rep. 2, sep 2020.

[3] O. O. Oladimeji and O. Oladimeji, "Predicting Survival of Heart Failure Patients Using Classification Algorithms," *JITCE (Journal of Information Technology and Computer Engineering)*, vol. 4, pp. 90–94, sep 2020.

[4] R. Gürfidan and M. Ersoy, "Classification of Death Related to Heart Failure by Machine Learning Algorithms," Tech. Rep. 1, jan 2021.

[5] A. Ishaq, S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques," *IEEE Access*, vol. 9, pp. 39707–39716, 2021.

[6] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1–16, 2020.

[7] T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab, and M. A. Raza, "Survival analysis of heart failure patients: A case study," *PLoS ONE*, vol. 12, p. e0181001, jul 2017.

[8] P. E. Rubini, C. A. Subasini, A. Vanitha Katharine, V. Kumaresan, S. Gowdhamkumar, and T. M. Nithya, "A cardiovascular disease prediction using machine learning algorithms," *Annals of the Romanian Society for Cell Biology*, vol. 25, no. 2, pp. 904–912, 2021.

[9] L. Ali, A. Niamat, J. A. Khan, N. A. Golilarz, X. Xingzhong, A. Noor, R. Nour, and S. A. C. Bukhari, "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," *IEEE Access*, vol. 7, pp. 54007–54014, 2019.

[10] B. Dhomse Kanchan and M. Mahale Kishor, "Study of machine learning algorithms for special disease prediction using principal of component analysis," *Proceedings - International Conference on Global Trends in Signal Processing, Information Computing and Communication, ICGTSPICC 2016*, pp. 5–10, 2016.

[11] A. Rahman, A. W. Chowdhury, H. I. Khan, K. M. Sabah, M. G. Amin, and S. M. Azad, "Association of high sensitive C - Reactive protein with severity of the left ventricular systolic dysfunction in acute anterior ST elevation myocardial infarction," *Bangladesh Medical Research Council Bulletin*, vol. 44, pp. 71–76, nov 2018.

[12] M. Aljanabi, M. Qutqut, and M. Hijjawi, "Machine Learning Classification Techniques for Heart Disease Prediction: A Review," *International Journal of Engineering and Technology*, vol. 7, no. October, pp. 5373–5379, 2018.

[13] D. S. P. Jaymin Patel, Prof.TejalUpadhyay, "Heart Disease Prediction Using Machine learning and Data Mining Technique," *International Journal of Computer Science and Engineering*, vol. 7, no. 6, pp. 17–19, 2015.

[14] M. C. S. Dangare and D. M. S. S. Apte, "A DATA MINING APPROACH FOR PREDICTION OF HEART DISEASE USING NEURAL NETWORKS," *nternational Journal of Computer Engineering and Technology (IJCET), ISSN 0976 – 6367(Print), ISSN 0976 – 6375(Online) Volume 3, Issue*, vol. 3, no. Issue 3, October - December, pp. 30–40, 2012.

[15] I. A. Zriqat, A. M. Altamimi, and M. Azzeh, "A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods," vol. 14, no. 12, pp. 868–879, 2016.

[16] M. Shouman, T. Turner, and R. Stocker, "Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients," *International Journal of Information and Education Technology*, pp. 220–223, 2012.

[17] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1–16, 2019.

[18] A. Mustaqeem, S. M. Anwar, M. Majid, and A. R. Khan, "Wrapper method for feature selection to classify cardiac arrhythmia," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 3656–3659, 2017.

[19] N. Khateeb and M. Usman, "Efficient heart disease prediction system using K-nearest neighbor classification technique," *ACM International Conference Proceeding Series*, pp. 21–26, 2017.

[20] A. K. Sonam Nikhar, "Prediction of heart disease using machine learning algorithms," *International Journal of Advanced Engineering, Management and Science (IJAEMS)*, pp. 197–202, 2016.

[21] "UCI Machine Learning Repository: Heart failure clinical records Data Set."