A Comparative Study of Machine Learning Algorithms for Heart Failure Survival Prediction

Submitted by

Mithun Kumar Das

ID: 213-35-796

Department of Software Engineering

Daffodil International University

Supervised by

Dr. Md. Fazla Elahe

Assistant Professor and Associate Head

Department of Software Engineering

Daffodil International University

This Thesis paper has been submitted in fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering.

Summer 2025

APPROVAL

This thesis titled on "**A Comparative Study of Machine Learning Algorithms for Heart Failure Survival Prediction**", submitted by **Mithun Kumar Das (ID: 213-35-796)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Fazla Elahe

--- **Chairman**

Dr. Md. Fazla Elahe
Assistant Professor & Associate Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Marzia

--- **Internal Examiner 1**

Dr. Marzia Ahmed
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

13.09.2025

--- **Internal Examiner 2**

Dr. Shabnom Mustary
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

13.09.25

--- **External Examiner**

Mohammod Abul Kashem
Professor
Department of Computer Science and Engineering
Dhaka University of Engineering & Technology, Gazipur.

A Comparative Study of Machine Learning Algorithms for Heart Failure Survival Prediction

MITHUN KUMAR DAS
ID: 213-35-796

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

SUPERVISOR's DECLARATION

It is hereby my declaration I have read this thesis entitled " A Comparative Study of Machine Learning Algorithms for Heart Failure Survival Prediction " and, to my judgment, it meets the degree requirements to honor the degree of Bachelor of Science in Software Engineering.

Fazla Elahe

(Supervisor's Signature)

Full Name : Dr. Md. Fazla Elahe

Position : Assistant Professor and Associate Head

Date : 13 September 2025

STUDENT’S DECLARATION

I would now verify the fact that the paper in this thesis is founded on my own work unless the quotations and quotations are illustrated by referencing that mentioned accordingly. I also confirm that it has neither been filed before or simultaneously with any other degree in Daffodil international university or any other university.

Mithun

____

(Student’s Signature)

Full Name : Mithun Kumar Das

ID Number : 213-35-796

Date : 13 September 2025

A Comparative Study of Machine Learning Algorithms for Heart Failure Survival Prediction

MITHUN KUMAR DAS
ID:213-35-796

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

September 2025

ACKNOWLEDGEMENTS

To start with, I would like to thank and thank almighty God, who has given me the heavenly gift, which allows finishing the final year thesis successfully. So, I am truly indebted as per my research supervisor Dr. Md. Fazla Elahe, who helped me in the entire research activities.

I would like to thank especially Dr. Imran Mahmud, the Head of the Faculty, who has ensured that all required facilities were provided to the purpose of the research. I am also gratified to the entire lecturers in, Department of Software Engineering who had heartedly plotted me through my misfortunes. I owe my friend the debt of gratitude because he was there with me in this business. Lastly, I would have to give due credit to my parents who were constantly in my support.

DEDICATION

This thesis is owed to my loving parents, whose unwavering love, encouragement, and support have been my pillars ever since. To my parents, for hardships and guidance that shaped my path. Finally, thanks to the almighty God.

ABSTRACT

Heart failure (HF) is one of the most common causes of death and morbidity in the world and poses to be a serious problem in early diagnosis and survival prognosis. In this study for predicting heart disease survival using a dataset of 5000 patients. Precise and early prognostication potential in automating and improving survival analysis. This paper documents the comparison of different ML techniques applicable to predict survival among HF patients: Random Forest, Decision Tree, Gradient Boosting, K-Nearest Neighbours, Support Vector Machine, Ad Boost, Logistic Regression, and Naive Bayes. This study will be based on the data that we will use to include some of the clinical parameters that were read by the patients who had heart failure. Before training the models, data pre-processing, balancing with ADASYN and feature scaling have been used. The assessment was done based on standard metrics of performance, including accuracy, precision, recall, F1-score, and ROC AUC. Model performance was analysed using visualization tools such as a confusion matrix, ROC, and importance of features plots. In this study, using analogical algorithms depends on accuracy, precision, recall, F1-score, and Random Forest (RF) shows the highest accuracy of of survival events among patients with HF also continues to be an imminent obstacle because of the heterogeneous and complex characteristics of the disease. Nonetheless, the current developments in machine learning (ML) have demonstrated 99.5%.

Keywords: Heart Failure, Machine Learning, Survival Prediction, Classification, Imbalanced Data, ROC-AUC, ADASYN, Clinical Decision Support.

TABLE OF CONTENT

Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
SVM	Support Vector Machine
KNN	K-Nearest Neighbours
RF	Random Forest
DT	Decision Tree
GB	Gradient Boosting
NB	Naive Bayes
LR	Logistic Regression
AB	AdaBoost
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
MSE	Mean Squared Error
R^2	Coefficient of Determination
MCC	Matthews Correlation Coefficient
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
F1-Score	Harmonic Mean of Precision and Recall
CV	Cross Validation
TPU	Tensor Processing Unit
CPU	Central Processing Unit
EF	Ejection Fraction
CHF	Congestive Heart Failure
CAD	Coronary Artery Disease

ECG	Electrocardiogram
BP	Blood Pressure
HR	Heart Rate
HF	Heart Failure
CSV	Comma Separated Values
API	Application Programming Interface
IoT	Internet of Things
KPI	Key Performance Indicator

CHAPTER 1

INTRODUCTION

1.1 Introduction

Heart failure (HF) is usually a highly developed, and even life threatening, illness that impairs the pumping ability of the heart, and, eventually, can cause a person to die. It is among the deadliest murderers in the world that is fast gaining momentum through ageing of the population and other risk causes [1]. According to the WHO estimates, majority of approximately over 64 million people have been living with HF at this time [2]. Forecasting survival of heart failure patients is an essential strategy of enhancing clinical and alleviating healthcare systems, especially when in time [3]. The potential of machine learning in many clinical applications such as diagnosis, prognosis, and patient stratification is also high. The algorithms presented by ML help provide an effective method of data driven survival prediction in patients using multifaceted and complex health data [4]. We analyzed and evaluated the predictive accuracy of 8 widely used ML models RF, DT, GB, KNN, Support Vector Machine (SVM), AdaBoost, Logistic Regression (LR) and NB) in heart failure survival prediction. The information utilized to achieve the aim of the current work was taken out of clinical records of patients with heart failure, and the chosen attributes were patient age, pressure level, ejection fraction, serum creatinine, and other essential factors. Several pre-processing steps, i.e., normalization, class balancing with ADASYN and feature selection are carried on to enhance the prediction performance. A model's performance is assessed on a set of visualization and evaluation techniques and metrics, such as accuracy, precision, recall, F1-score, and ROC-AUC. [5]. The machine learning models may take the nonlinear, hidden relationships between the clinical covariates which may not appear to be apparent by the traditional statistical models. Estimating risk of mortality is an integral part of HF management, providing physicians the means to guide care, modulate treatment, and convey prognostic information. Comparing different ML models in this field will help doctors to choose better tools for predictions according to their practice. The present study also has the potential to add to a burgeoning research field of AI-based clinical decision support systems by determining the best-suited algorithm for the prediction of heart failure survival. Based on the detailed analysis and visualization, the findings are indicative for robust deployment of ML models in real-world healthcare settings.

1.2 Motivation

Heart failure continues to be among the most pressing global health problems as it results in millions of deaths every year and imposes enormous burdens on healthcare systems across the globe [12]. This study is essentially aimed at better predicting the survival of people suffering from heart failure, which is an area of considerable significance, in terms of clinical outcomes and public health. Relatively little work has been done to investigate heart failure survival with a wide range of machine learning models in previous studies. But with recent progress in machine learning, we face a critical need for more accurate and data-driven survival prediction models.

The research is underpinned by practical implications. In practice, quickly and effectively predicting controller survival probability could be of great support to clinicians in Approach, resource distribution and Individual treatment program. Such an approach can also serve to decrease the time taken for manual evaluations, and help physicians to concentrate more on critical interventions. Finally, getting this work out there is our ultimate goal to provide better, AI-driven, tools for decision making, which may contribute to greater accuracy in predicting heart failure and in-patient care.

1.3 Rationale of the Study

Heart failure is a significant public health problem with high global morbidity and mortality. Early prediction of the survival results is very important for the treatment planning and the management of patients. Despite a plethora of machine learning (ML) methods, there is a requirement for evaluating and comparing how well individual ML methods perform on congruent data to identify which models are the most trustworthy. Most of the existing works concentrate on algorithms individually without performing thorough comparisons. The goal of this study is to fill this gap by assessing and comparing different machine learning (ML) algorithms, such as Random Forest, Logistic Regression, Support Vector Machine and others from a systematic review perspective, on the structured heart failure dataset. In this way, the present study is also not only oriented to find the best accurate model, but also to give a solid basis to prepare such models for a real clinical decision support system. The comparative study provides insights into the model behavior and assists in the choice of the most relevant technique for the future deployment in healthcare applications.

1.4 Objectives

This work addresses the motivation to develop a robust and high-performing prognostic ML model for predicting survival in HF patients, which is a major clinical need in HF medicine for enhancing prognostic tools. One practical issue is how to reduce the extremely high risk due to disproportionate negative outcome cases in outcome prediction - that is, non-survival cases are few. To reduce model instability and the extent to which the models can be generalised, we will be very cautious with how we apply and evaluate balancing in the study. This besides, the work features systematic development, training and optimization of a number of most successful machine-learning models. The models will be compared detail by detail for prediction with respect to accuracy, sensitivity, specificity and clinical net benefit in order to determine the model for survival estimation of heart failure.

1.5 Scope of the Problem

The burden of HF remains a major public health issue in Bangladesh and elsewhere in many resource-poor countries and this is because it is a major determinant, which influences morbidity and coverage of death. The majority of heart failure patients remain un-diagnosed and unless diagnosed, a significant proportion would not be properly treated owing to the lack of knowledge or access to healthcare. Early stage symptoms are ignored or confused and then lead to very serious complications including death. Further, a large proportion of predictive models that are available are out-of-date or pre-train on data from population groups that do not represent the other population groups, making them not generally applicable in real life settings.

This gap is bridged in our work by leveraging a validated dataset available at Kaggle, which contains a well-structured data consisting of a set of clinically relevant features that are useful in the prediction of survival of heart failure. We then train various machine learning models and check their prediction accuracy by putting them in comparisons against each other, to find out the best and the most reliable models. The work has a strong potential for real-life applications, such as development of intelligent decision-support tools for clinicians, mobile apps for early risk identification and patient education. The approach we chose is going to bring more personalized and earlier-crash-improvement outcomes for patients.

CHAPTER 2

BACKGROUND STUDIES

2.1 Overview

Heart failure (HF) is the most frequent and complex clinical syndrome with a high rate of hospitalization and mortality of patients worldwide. Estimation of the prognosis of patients with heart failure is essential for optimal clinical decision-making, management plans, and use of resources. However, traditional prediction models are only based on the statistical model or subjective expert experience of the physician traditions are hard to choose the extremely complex progress of the heart failure process due to a large number of overlapping clinical factors. In recent years, machine learning (ML) approaches are increasingly recognized as a promising tool in the medical prognosis resulting in solutions of multidimensional, large-scale data. They will be identified subtle patterns and none-linear association among patients' data that not be discoverable by means of conventional analysis. Application of ML in predicting heart failure survival reportedly has the potential to provide better overall performances and lock in high-performance prognostic models, thus driving more patient-specific and timely interventions. It will go running through different machine learning models like Naive Bayes, Logistic Regression, AdaBoost, Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Gradient Boosting, Decision Tree, and Random Forest for the purpose of how much better they predict for the people who suffer from heart failure diseases. The selection criterion of these models was to span a wide range of statistical and machine learning approaches, from simple probabilistic classifiers to ensemble learning methods, and thus to enable a full comparison of the predictive performance of such models. Also, in this chapter, an in-depth review of related works has been incorporated in the field of heart failure survival prediction and machine learning in health care application. The findings of the study identify the gap in the existing research methods and give pointers to work on the development of effective and implementable prognostic models that, eventually, can affect positive patient outcomes.

2.2 Literature Review

This research based on ML algorithms for early HF survival prediction, Nasiruddin at al. (2024) in his paper attain high accuracy with SVM and the result are (accuracy: 88.41%, precision: 89.76%, recall: 90.85%, F1-score: 90.30%, ROC-AUC: 94.97%). For this prediction, he is using a dataset of 918 patients [6]. Qadri et al. (2024) This study achieves the highest accuracy with Random Forest (RF) of 97.5%. To analyze the data using 299 hospitalized heart failure patients. Addressing the imbalance in data using the SMOTE method [7]. Ali et al. (2024) focus on predictive analysis in survival heart failure patients. A 299 dataset of patient records was used in this paper for prediction. Random Forest (RF) achieves the highest accuracy of 97.78%. phosphokinase and SS Additionally, a further breakdown of these factors reveals a large cluster of the features. It is discovered that increment of SC, age, and SS, and the decrement of EF, are the strongest risk factors of HF patients, in the survival analysis. [8]. Sahoo et al. (2020) The data set utilized in this paper was recorded in the UCI repository and contains 13 significant attributes. Most algorithms used in this work include SVM, NB, LR, Decision Tree and KNN. It is discovered that SVM offered the most desirable findings with a maximum accuracy of 85.2% [9]. Bhatt et al. (2023) The AUC values of the proposed models are as follows: DT (0.94), XGB (0.95), RF (0.95), and MLP (0.95). The findings of this study show that the MLP model, evaluated using cross-validation, outperformed the other algorithms in accuracy, achieving the highest score of 87.28% [10].

CHAPTER 3

METHODOLOGY

3.1 Introduction

I describe the method of the study in this section. My first step in research was gathering correct data from Kaggle. Based on heart failure, I elected to use the MLAs and P for being highly predictive in clinical thought. The data pre-processing was necessary before model construction. This work is intended to increase the understanding of, and aid in predicting the risk of headache in health care, which in turn will assist in better planning treatment and patient care. The research phases consist of collecting data, pre-processing, selecting feature, implementing algorithm, evaluating model, and comparing performance. To get secondary data in analysis or ML models ready(focusing on cleaning and organizing) for use through various python libraries as per data handling or preprocessing.

1. Pandas: to organize the data and manage datasets
2. Numpy: used for numerical computing and matrix operations.
3. train _test_split_from sklearn_model_selection*:* used for dividing the dataset into training and testing subsets (75% training and 25% testing)
4. *S*tandardScaler from sklearn. pre-processing: to scale the features and normalize the dataset.

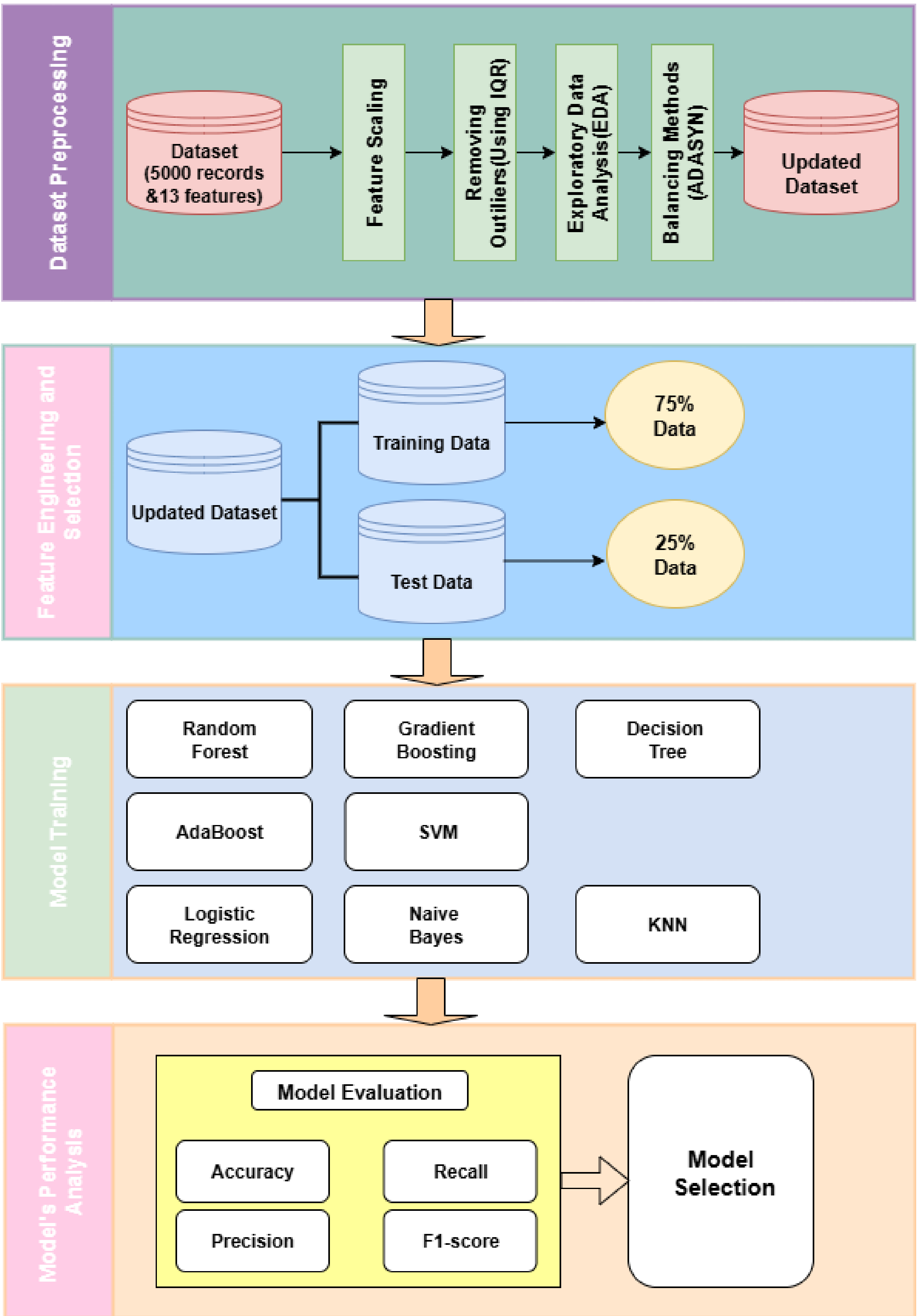

Figure 3.1 Methodology of a heart failure survival prediction model using machine learning.

3.1 Data Collection

Data is the most important part of any research that involves machine learning, as models cannot be trained and tested without the data. Data can play a role as the foundation of the entire machine learning approach. In this study, a secondary dataset sourced from Kaggle was utilized for the analysis of heart failure survival patients. The data set consists of 5,000 records, and each record represents a case by itself as well as a binary target for survival (0 = survived, 1 = death). Among these, 1,568 are cases of death and 3,432 are survivors, hence the creation of an unbalanced data set where the cases of death account for about 31.36% of the overall set. This imbalance is an important consideration for predictive modeling, as it may affect classification performance and necessitate appropriate handling techniques to ensure accurate and unbiased results.

3.2 Dataset Sample

```
df.head(10)
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
0	55.0	0	748	0	45	0	263358.03	1.3	137	1	1	88	0
1	65.0	0	56	0	25	0	305000.00	5.0	130	1	0	207	0
2	45.0	0	582	1	38	0	319000.00	0.9	140	0	0	244	0
3	60.0	1	754	1	40	1	328000.00	1.2	126	1	0	90	0
4	95.0	1	582	0	30	0	461000.00	2.0	132	1	0	50	1
5	70.0	0	232	1	30	0	302000.00	1.2	132	1	0	210	0
6	63.0	1	122	1	60	0	172000.00	1.2	145	0	0	147	0
7	70.0	1	171	0	50	1	358000.00	0.9	141	0	0	196	0
8	50.0	0	482	1	30	0	300000.00	0.9	132	1	0	109	0
9	53.0	1	446	0	45	1	274000.00	1.0	133	1	0	215	0

Figure 3.2 Dataset sample overview.

3.3 Data Pre-processing

Cleaned up the heart failure dataset and standardized it using StandardScaler, and divided it into a 75% training and a 25% test dataset. To deal with class imbalance, the ADASYN approach was adopted to oversample the minority class.

```
[13]  1  df.describe()
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	60.288736	0.474400	586.760600	0.439400	37.734600	0.364800	265075.404370	1.369106	136.808200	0.645600	0.311800	130.678800	0.313600
std	11.697243	0.499394	976.733979	0.496364	11.514855	0.481422	97999.758622	1.009750	4.464236	0.478379	0.463275	77.325928	0.464002
min	40.000000	0.000000	23.000000	0.000000	14.000000	0.000000	25100.000000	0.500000	113.000000	0.000000	0.000000	4.000000	0.000000
25%	50.000000	0.000000	121.000000	0.000000	30.000000	0.000000	215000.000000	0.900000	134.000000	0.000000	0.000000	74.000000	0.000000
50%	60.000000	0.000000	248.000000	0.000000	38.000000	0.000000	263358.030000	1.100000	137.000000	1.000000	0.000000	113.000000	0.000000
75%	68.000000	1.000000	582.000000	1.000000	45.000000	1.000000	310000.000000	1.400000	140.000000	1.000000	1.000000	201.000000	1.000000
max	95.000000	1.000000	7861.000000	1.000000	80.000000	1.000000	850000.000000	9.400000	148.000000	1.000000	1.000000	285.000000	1.000000

Figure 3.3 Descriptive Statistics of Heart Failure Patient Dataset.

The dataset comprises 5,000 samples and 13 patient survival of HF patients. The median age of the cases is about 60.29 years, and the range is from 40 to 95 years with a standard deviation of 11.69 years. The prevalence of diabetes, anemia and hypertension are seen in about 47%, 43% and 36.48% of the patients. It ranges widely from 23-7,861 with a mean of 586.76 (U/L). The ejection fraction mean = 37.73%; (range, from 14%, to 80%). The platelet count is normal with a mean of 260,575/mL and a wide variation of 25,100 to 850,000. The overall mean of serum creatinine is 1.37 mg/dL and sodium is 136.86 mEq/L, 64.57% of them are males and 31.18% of patients are smokers. The time range ("time") is 4-285 day in a mean value as 130.68 day. Nearly 31.36% of the patients died according to DEATH_EVENT.

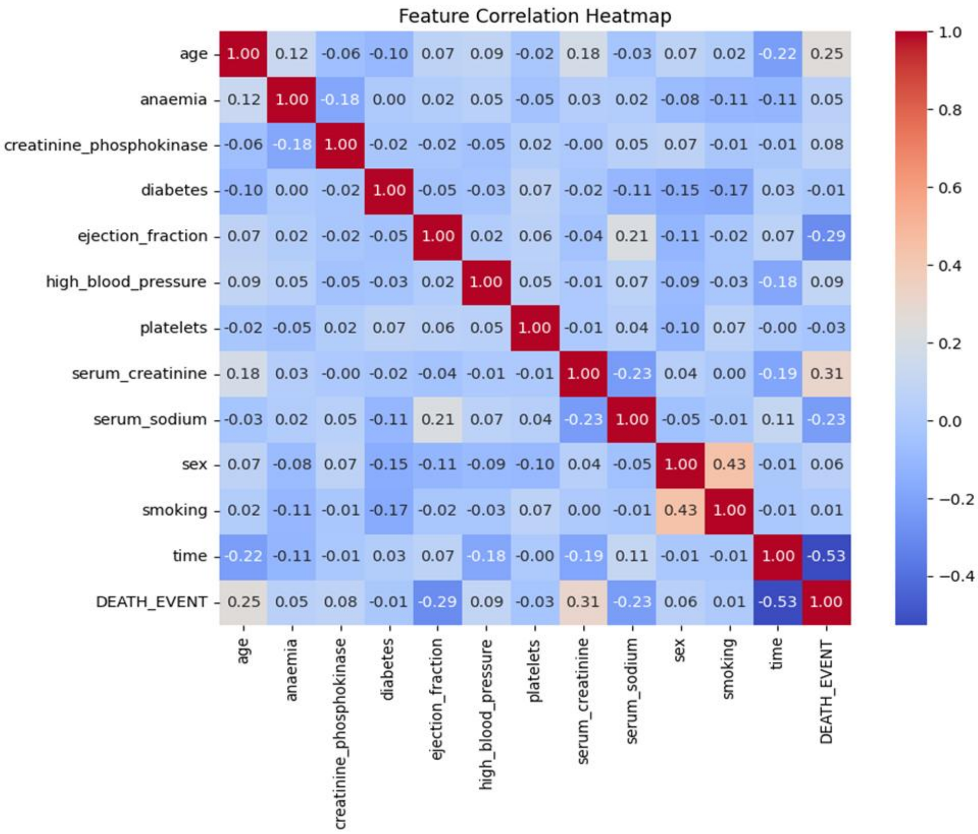

Figure 3.4 Survival - feature correlation heat map outcomes

Figure 3.4 visualizes a feature correlation heatmap with pairwise Pearson correlation coefficients among variables. Some of the findings are: DEATH_EVENT is positively correlated with age (0.25) and serum creatinine (0.31), and negatively correlated with ejection fraction (−0.29) and time (−0.53). This suggests that rising age and serum creatinine are associated with increased mortality, while rising ejection fraction and longer follow-up times are associated with better survival. All but most of the clinical features have low correlations with one another, indicating low multicollinearity within the data. The strongest positive non-target correlation is between sex and smoking (0.43), indicating a higher prevalence of smoking in one gender group.

3.6 Experimental Result

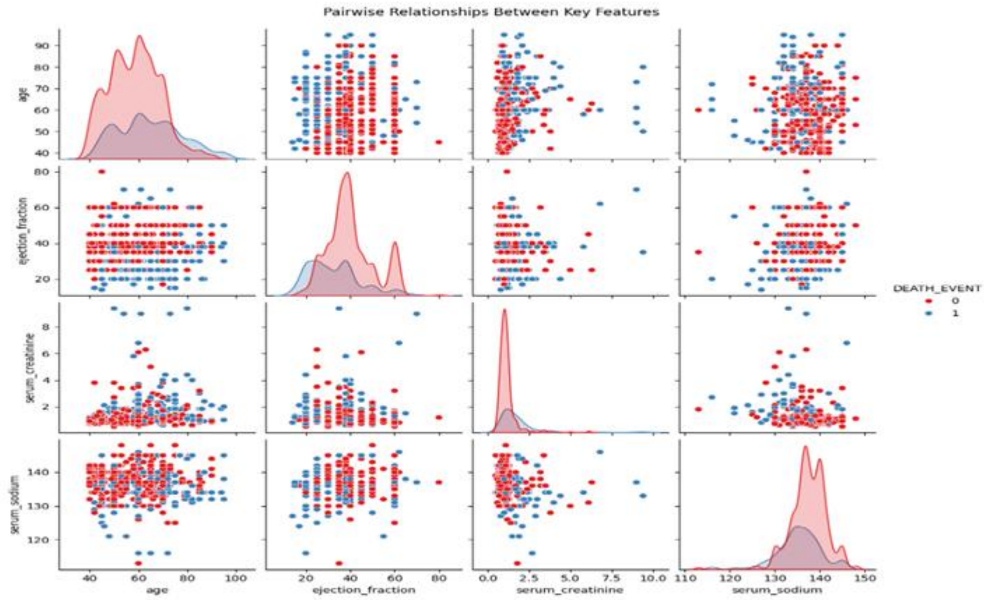

Figure 3.5 Pairwise Relationships Between Key Features.

In Figure 3.5 the pairwise relationships between age, Ejection fraction, Serum creatinine, and serum sodium are mapped against the target variable DEATH_EVENT. 1 indicates the event, and 0 indicates the absence of the event. Patients who suffered the event are marked in red, while the remaining patients are marked in blue. As shown in the diagonal plots, the kernel density estimations (KDE) are shown for each feature, which separates the groups based on their distributions. Patients with lower ejection fraction and higher serum creatinine seem to be more frequent, and ejection fraction and serum creatinine can be considered significant predictive features.

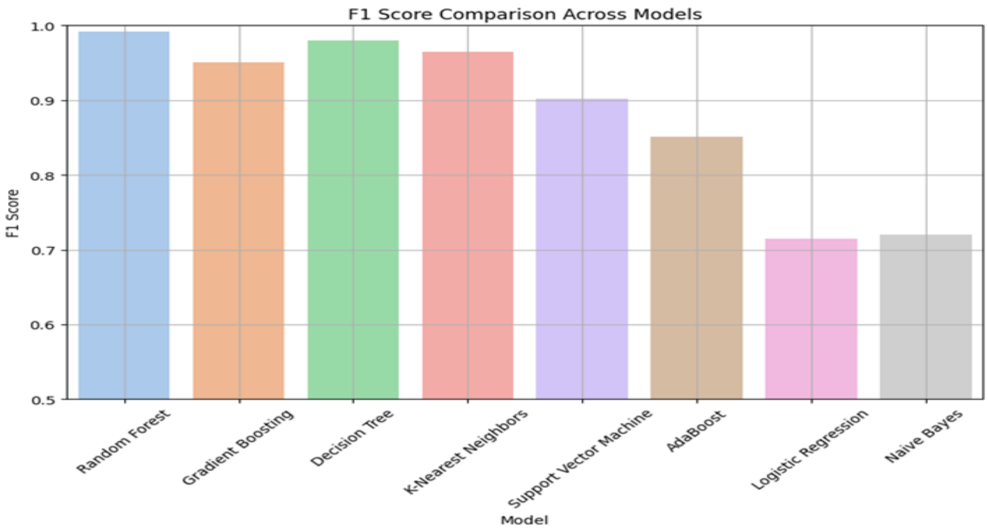

Figure 3.6 F1-Score Comparison Across Models

As shown in Figure 3.6 , the machine learning models applied for the prediction of heart failure have also been evaluated on the F1-score. As shown before, the Random Forest classifier proved to have the highest F1 score, with close competition from Decision Tree and K-Nearest Neighbors, which also scored high in their balance of precision and recall. This prediction does seem concerning for Logistic Regression and Naïve Bayes, as they appear to have weaker performance on the FC score caused by the confusion matrix.

3.7 Machine Learning Models Evaluation

3.7.1 Logistic Regression

```
📊 Logistic Regression Evaluation
----------------------------------------
Accuracy      : 0.807
Precision     : 0.6631
Recall        : 0.7834
F1 Score      : 0.7182
MCC           : 0.5776
Cohen's Kappa : 0.573
ROC AUC       : 0.8965
Confusion Matrix:
 [[561 125]
 [ 68 246]]
```

Figure 3.7 Logistic Regression Evaluation

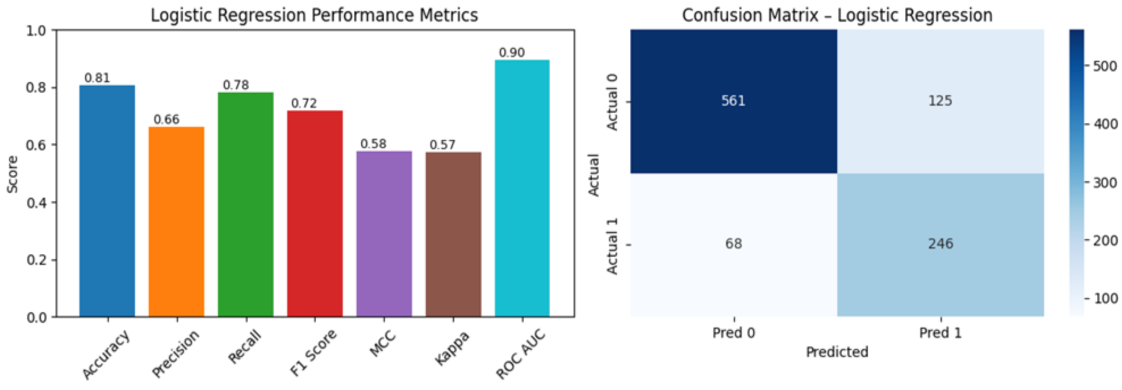

Figure 3.8 Logistic Regression Performance Metrics and Confusion Matrix

3.7.2 Support Vector Machine

```
📊 Support Vector Machine Evaluation
----------------------------------------
Accuracy      : 0.942
Precision     : 0.8743
Recall        : 0.9522
F1 Score      : 0.9116
MCC           : 0.8703
Cohen's Kappa : 0.8685
ROC AUC       : 0.9766
Confusion Matrix:
 [[643  43]
 [ 15 299]]
```

Figure 3.9 Support Vector Machine Evaluation

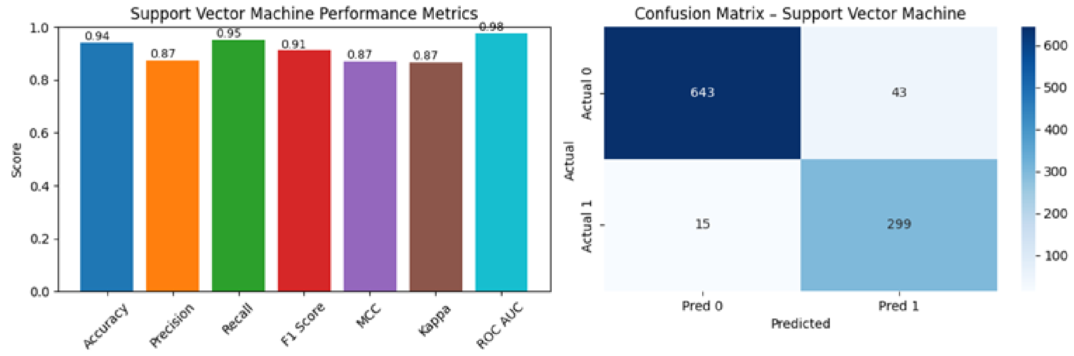

Figure 3.10 Support Vector Machine Performance Metrics and Confusion Matrix

3.7.3 Random Forest

```
📊 Random Forest Evaluation
----------------------------------------
Accuracy      : 0.995
Precision     : 0.9936
Recall        : 0.9904
F1 Score      : 0.992
MCC           : 0.9884
Cohen's Kappa : 0.9884
ROC AUC       : 0.9999
Confusion Matrix:
 [[684   2]
 [  3 311]]
```

Figure 3.11 Random Forest Evaluation

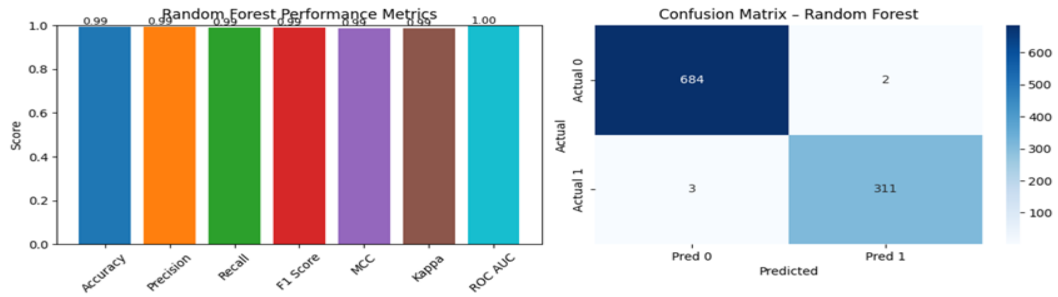

Figure 3.12   Random Forest Performance Metrics and Confusion Matrix

3.7.4 Gradient Boosting

```
Gradient Boosting Evaluation
-------------------------------------
Accuracy        : 0.977
Precision       : 0.9505
Recall          : 0.9777
F1 Score        : 0.9639
MCC             : 0.9472
Cohen's Kappa   : 0.947
ROC AUC         : 0.9954
Confusion Matrix:
 [[670  16]
 [  7 307]]
```

Figure 3.13   Gradient Boosting Evaluation

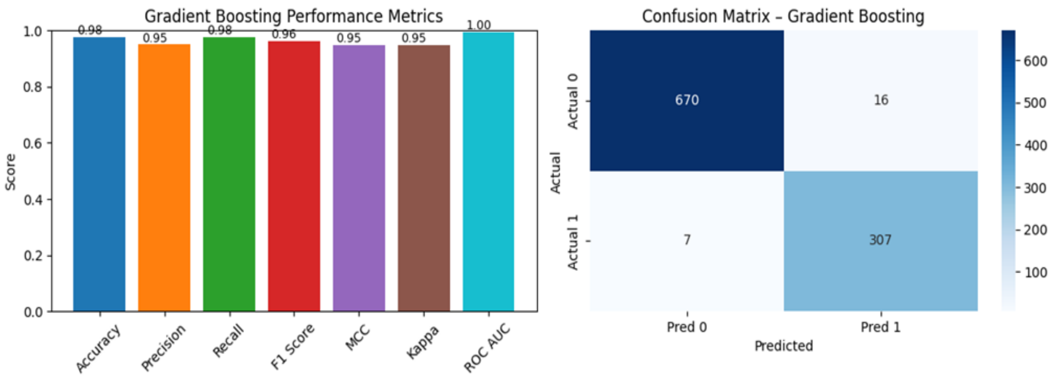

Figure 3.14 Gradient Boosting Performance Metrics and Confusion Matrix

3.7.5 AdaBoost

```
📊 AdaBoost Evaluation
----------------------------------------
Accuracy       : 0.933
Precision      : 0.8709
Recall         : 0.9236
F1 Score       : 0.8964
MCC            : 0.8478
Cohen's Kappa : 0.847
ROC AUC        : 0.9858
Confusion Matrix:
 [[643  43]
 [ 24 290]]
```

Figure 3.15 AdaBoost Evaluation

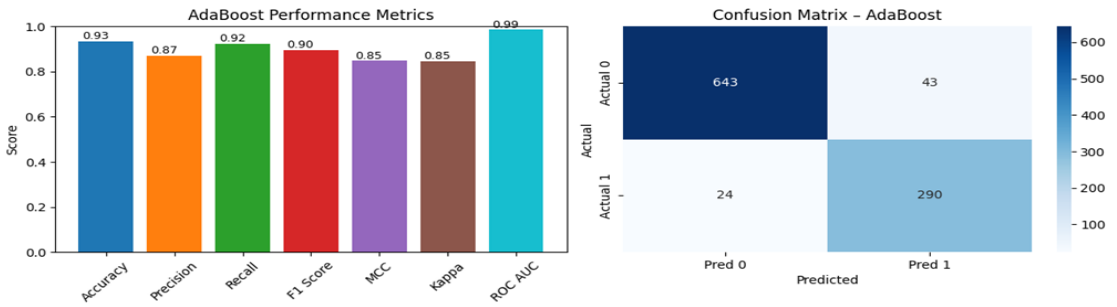

Figure 3.16 AdaBoost Performance Metrics and Confusion Matrix

3.7.6 K-Nearest Neighbours

```
📊 K-Nearest Neighbors Evaluation
------------------------------------
Accuracy       : 0.977
Precision      : 0.959
Recall         : 0.9682
F1 Score       : 0.9635
MCC            : 0.9468
Cohen's Kappa : 0.9467
ROC AUC        : 0.9813
Confusion Matrix:
 [[673  13]
 [ 10 304]]
```

Figure 3.17 K-Nearest Neighbours (KNN) Evaluation

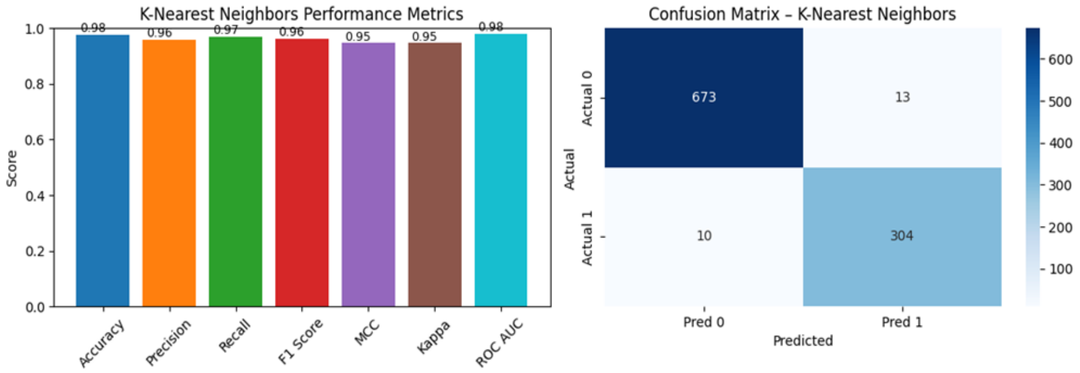

Figure 3.18 K-Nearest Neighbours (KNN) Performance Metrics and Confusion Matrix

3.7.7 Naïve Bayes

```
📊 Naive Bayes Evaluation
-----------------------------------
Accuracy       : 0.82
Precision      : 0.719
Recall         : 0.7006
F1 Score       : 0.7097
MCC            : 0.5794
Cohen's Kappa  : 0.5793
ROC AUC        : 0.883
Confusion Matrix:
 [[600  86]
 [ 94 220]]
```

Figure 3.19 Naive Bayes Naive Bayes Evaluation

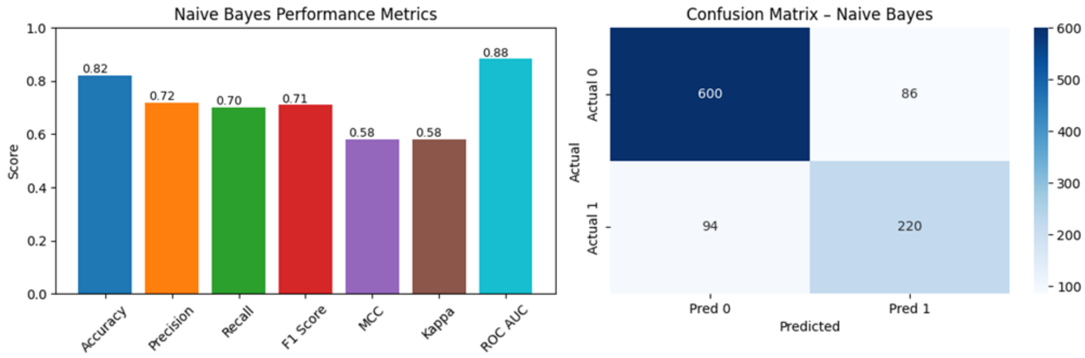

Figure 3.20 Naive Bayes Performance Metrics and Confusion Matrix

3.7.8 Decision Tree

```
📊 Decision Tree Evaluation
-----------------------------------
Accuracy        : 0.992
Precision       : 0.9873
Recall          : 0.9873
F1 Score        : 0.9873
MCC             : 0.9814
Cohen's Kappa   : 0.9814
ROC AUC         : 0.9907
Confusion Matrix:
 [[682   4]
 [  4 310]]
```

Figure 3.23 Decision Tree Evaluation

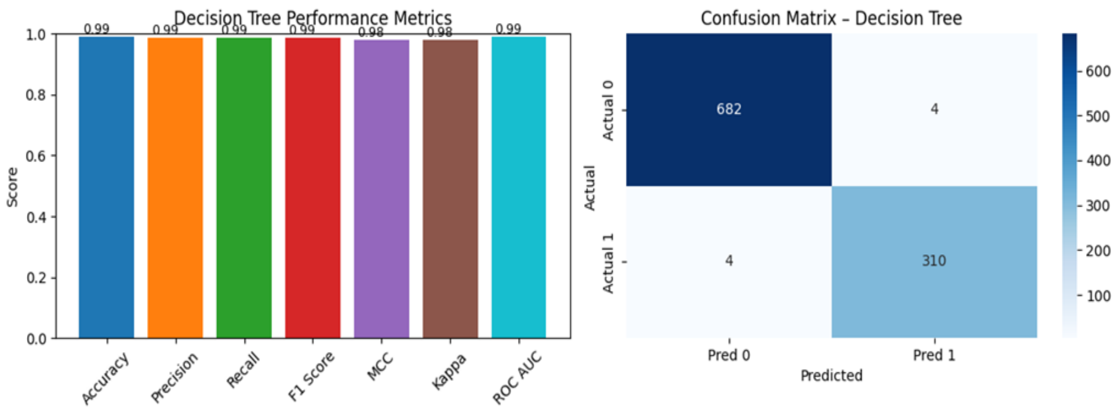

Figure 3.21 Decision Tree Performance Metrics and Confusion Matrix

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Result

Table 4.1 Independent Test Result Evaluation for Applied ML Algorithms.

Algorithm	Accuracy	Precision	Recall	F1-Score	MCC	Kappa	ROC AUC
Logistic Regression	0.807	0.6631	0.7834	0.7182	0.5776	0.573	0.8965
Support Vector Machine	0.942	0.8743	0.9522	0.9116	0.8703	0.8685	0.9766
Random Forest	0.995	0.9936	0.9904	0.9920	0.9884	0.9884	0.9999
Gradient Boosting	0.977	0.9505	0.9777	0.9639	0.9472	0.947	0.9954
AdaBoost	0.933	0.8709	0.9236	0.8964	0.8478	0.847	0.9858
K-Nearest Neighbors	0.977	0.9590	0.9682	0.9635	0.9468	0.9467	0.9813
Naive Bayes	0.820	0.7190	0.7006	0.7097	0.5794	0.5793	0.8830
Decision Tree	0.992	0.9873	0.9873	0.9873	0.9814	0.9814	0.9907

The evaluation results of various ML algorithms are outlined using key evaluation metrics such as Accuracy, Precision, Recall, F1-Score, MCC, Kappa and ROC AUC. RF achieved the topmost performance among all models with an accuracy of 99.5%, F1-score of 0.9920 and nearly perfect ROC AUC of 0.9999 demonstrating outstanding classification capability. High accuracy as well as strong F1-scores and ROC AUC was also noted with DT, GB and KNN over 97%. SVM was just behind with 94.2% accuracy and fairly stable performance across all metrics. On the other end, LR and NB were quite low performing with 80.7% and 82.0% accuracy, also lower Precision, Recall, MCC further showing limited capability in dealing with the dataset’s complexity. All in all, ensemble techniques such as those provided by RF and GB were shown to outperform conventional models, proving their resiliency and adeptness for classifier tasks.

CHAPTER 5

CONCLUSION

5.1 Introduction

The primary objective of this research was to develop an efficient and accurate ML-based system for predicting survival in HF patients. This work addresses the critical need for precise prognostic tools in clinical settings, aiming to assist healthcare professionals in early risk identification and personalized treatment planning. For this study, we utilized a publicly available Kaggle dataset containing clinical features such as age, serum creatinine, ejection fraction, and others. After comprehensive pre-processing—removing missing values and normalizing features we applied multiple ML algorithms to evaluate their predictive performance. The tested models included LR, SVM, RF, GB, AdaBoost, KNN, NB, and DT. The best results were achieved by RF (Acc 99.5%), DT (Acc 99.2%), and GB (Acc 97.7%). All models were evaluated using key metrics such as Acc, Pre, and Rec to ensure stable and balanced performance. The findings suggest that ensemble methods like RF and GB outperform conventional classifiers in predicting HF survival and are strong candidates for clinical adoption.

5.2 Summary

The given comparative study determines the optimal ML models to predict the survival of heart failure, which can be used in terms of developing clinical decision-support instruments. We have compared and tested various models, including LR, SVM,RF, GB, AdaBoost, KNN, NB, and DT to name a few. The best accuracies were obtained with Random Forest (99.5%), followed by Decision Tree (99.2%)

and GB (97.7%), as it could once again be seen that ensemble methods worked better in this case. The outcomes of the present study have revealed that the system of ML is useful in predictive modelling of heart failure as it has demonstrated well-timed and consistent estimates with the ability to develop interpretable results that can aid healthcare providers in early prediction and management planning. The results apply to general clinical settings and patient demographics. Although machine learning tools offer great potential, ethical considerations and the need for clinical validation remain important. Future efforts will focus on integrating DL such as LSTM or hybrid frameworks, incorporating semi-supervised learning, and deploying these models in real-time clinical applications and remote monitoring systems. These advancements aim to make predictive healthcare more accessible and effective, particularly in under-resourced or remote areas.

REFERENCES

[1] Ahmad T, Lund LH, Coles A, et al. Machine learning methods improve prognostication and identify clinically distinct phenotypes in heart failure. J Am Coll Cardiol. 2018;71(20):2060–2070. DOI: 10.1161/JAHA.117.008081

[2] Kwon JM, Kim KH, Jeon KH, et al. Deep learning-based algorithm for detecting heart failure using electrocardiography. J Am Heart Assoc. 2019;8(19):e012743.

[3] Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

[4] Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial intelligence in cardiology. J Am Coll Cardiol. 2018;71(23):2668–2679.

https://doi.org/10.1016/j.jacc.2018.03.521

DOI 10.7717/peerj-cs.1894.

[5] Ali, M. M., Al-Doori, V. S., Mirzah, N., Hemu, A. A., Mahmud, I., Azam, S., ... & Moni, M. A. (2023). A machine learning approach for risk factors analysis and survival prediction of Heart Failure patients. *Healthcare Analytics*, *3*, 100182. https://doi.org/10.1016/j.health.2023.100182

[6] Sahoo, P. K., & Jeripothula, P. (2020). Heart failure prediction using machine learning techniques. *Available at SSRN 3759562*.

[7] Zolbanin HM, Yadollahi F, Hussain S, Keshavarz H. Machine learning applications for predicting heart failure: A systematic review. IEEE Access. 2020;8:160003–160018.

[8] Ibrahim M, Kassem M, Yousri D, et al. Heart failure diagnosis and classification using multi-feature fusion and machine learning. Comput Methods Programs Biomed. 2021;205:106094.

[9] Huang K, Fu Y, Xing T, et al. Using machine learning methods to predict patient mortality in heart failure. Med Biol Eng Comput. 2020;58(7):1383–1394.

[10] Khera R, Haimovich J, Hurley NC, et al. Use of machine learning models to predict death after acute myocardial infarction. JAMA Cardiol. 2021;6(6):633–641.

[11] Sun X, Yan X, Xiao Y, et al. Machine learning for prognosis of heart failure with reduced ejection fraction. J Card Fail. 2019;25(3):198–205.

[12] DeRose SF, Stewart CR, Mielke JG, et al. Applying machine learning to predict hospitalization for heart failure. Comput Biol Med. 2020;126:104005.

[13] Park JJ, Choi D, Hong YJ, et al. Artificial intelligence-based prediction of heart failure with preserved ejection fraction outcomes. Eur Heart J. 2020;41(29):2921–2931.

[14] Kim M, Lim J, Seo S, et al. Machine learning for predicting mortality among heart failure patients. Sci Rep. 2020;10(1):20600.

[15] Li X, Yu X, Wang Y, et al. A machine learning approach to predict heart failure prognosis using clinical data. Int J Med Inform. 2021;146:104333.

[16] Zolfaghar K, Noorian M, Soltanian-Zadeh H, Rajabi M. Deep learning for early detection and prognosis of heart failure. IEEE Access. 2020;8:205248–205258.

[17] Shanthikumar S, Mitra A, Jha MK. Prognostic modeling of heart failure using supervised machine learning. Comput Methods Programs Biomed. 2019; 179:104982.

ACCOUNTS CLEARANCE