Predictive Modeling of Heart Failure: Integrating Clinical Data With Machine Learning

BY

Tanvir Ahmed
ID: 0242310005103020

This Report Presented in Partial Fulfilment of the Requirements for the Degree of Masters of Science in Computer Science and Engineering

Supervised By

Mr. Md. Sazzadur Ahamed
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Abdus Sattar
Associate Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2024

APPROVAL

This Project titled "**Predictive Modeling of Heart Failure: Integrating Clinical Data with Machine Learning**", submitted by **Tanvir Ahmed** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of **M.Sc. in Computer Science and Engineering (MSc.)** and approved as to its style and contents. The presentation has been held on **06/07/2024.**

BOARD OF EXAMINERS

Dr, Sheak Rashed Haider Noori,PhD
Professor and Head — Chairman
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr.Md. Zahid Hasan, PhD
Associate Professor — Internal Examiner
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr,Arif Mahmud,PhD
Associate Professor — Internal Examiner
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud — External
Professor
Department of CSE
Jahangirnagar University

DECLARATION

I hereby declare that, this thesis has been done by us under the supervision of **Mr.Sazzadur Ahamed, Assistant Professor, Department of CSE** Daffodil International University. I also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Mr. Sazzadur Ahamed
Assistant professor
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Tanvir Ahmed
ID: 0242310005103020
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First of all, I would like to express our gratefulness to the almighty Allah for blessing usto complete the final year thesis successfully.

I am really grateful and wish our profound our indebtedness to **Mr. Md. Sazzadur Ahamed**, Assistant Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Data Mining, Big Data, Image Processing, Cloud Computing” to carry out this thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this thesis.

I would like to express our heartiest gratitude to **Mr. Abdus Sattar,Assistant Professor and Coordinator, Department of CSE**, for his kind help to finish our thesis and also to other faculty members and the staff of CSE department of Daffodil International University.

I would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and passion of our parents.

.

ABSTRACT

Heart failure is a critical health condition that poses significant challenges to the medical community due to its high mortality rate and substantial healthcare costs. Early prediction of heart failure can greatly improve patient outcomes by enabling timely interventions and personalized treatment plans. This study focuses on leveraging machine learning algorithms to predict heart failure at an early stage, utilizing a comprehensive set of clinical data and patient metrics.

In this research, I employ various machine learning algorithms, with a particular emphasis on the Random Forest algorithm, renowned for its robustness and accuracy in classification tasks. Our model is trained and validated using multiple datasets, including those sourced from established medical repositories, to ensure its reliability and applicability across different patient populations. Key performance metrics such as accuracy, precision, recall, and F-measure are utilized to evaluate the model's effectiveness.

The findings of our study indicate that the Random Forest algorithm excels in predicting heart failure, delivering significant improvements in prediction accuracy and reliability. The model not only identifies high-risk patients but also provides actionable insights for healthcare professionals to implement preventative measures and optimize treatment strategies early in the care continuum.

The proposed machine learning-based approach to heart failure prediction offers a powerful tool for enhancing clinical decision-making processes. By integrating this predictive model into healthcare systems, medical practitioners can reduce hospital readmission, lower healthcare costs, and improve overall patient care quality.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

High-quality healthcare delivery impacts patient satisfaction, clinical outcomes, and overall healthcare success. Medical errors, including missed diagnoses of conditions like heart failure, can have serious financial and operational consequences if not caught early. This research utilizes machine learning techniques to predict heart failure early, encouraging proactive healthcare interventions.

Healthcare data analysis is complicated, comprising patient history, diagnostic tests, treatments, and follow-ups. Each phase can have data points that indicate potential problems. Traditional heart failure prediction methods involve manual analysis of patient records and diagnostic tests, which are effective but time-consuming and expensive. These methods often identify conditions late, resulting in costly treatments and adverse health outcomes.

Automating and enhancing heart failure prediction with machine learning is a potential solution. Machine learning algorithms can identify patterns that predict heart failure by examining historical patient data and medical metrics. This predictive feature lets healthcare providers focus on the most at-risk patients, optimizing resource allocation and improving.

I analyze HF prediction using machine learning algorithms. I evaluate algorithms on multiple datasets to find the most effective ones. The study trains models using medical measures like blood pressure, cholesterol levels, patient history, and other relevant health metrics.

Effective heart failure prediction has huge potential. Early detection prevents costly and severe health complications, saving lives and money. It improves patient dependability and performance, enhancing healthcare quality and patient satisfaction. Automating condition identification allows healthcare professionals to focus on patient care and preventive measures rather than just diagnostic procedures.

Our research has the potential to transform healthcare. Incorporating machine learning-based heart failure prediction into the healthcare lifecycle can improve healthcare efficiency, cost-effectiveness, and reliability. This proactive strategy addresses immediate healthcare needs and helps healthcare projects succeed in the long run.

In conclusion, our machine learning-based heart failure prediction research advances medical science. I use data and predictive analytics to offer machine learning techniques that improve early detection, optimize healthcare resources, and create high-quality healthcare solutions.

1.2 Motivation

Heart failure is a major concern in healthcare, leading to increased costs and adverse patient outcomes. Early detection of heart failure can save significant resources and improve healthcare quality. Machine learning offers a promising solution by automating the prediction process. Create a model that can predict heart failure accurately predict heart failure, enabling healthcare providers to focus on preventive measures rather than reactive ones. By improving prediction, I aim to enhance the overall efficiency and reliability of healthcare processes.

1.3 Research Objective

Develop a ML model that can predict Heart failure with high accuracy which is out primary objectives. Specifically, we aim to:

1. Classify the maximum effective machine ML algorithm for HF(Heart Failure) prediction.
2. Accuracy, precision, recall, and F-measure asses the performance of this algorithm.
3. Provide a tool that helps healthcare providers detect heart failure early, thereby reducing treatment and maintenance costs.

1.4 Research Questions

To guide our research, I focus on the following questions:

1. Which machine learning algorithms are most effective for predicting heart failure?

2. What data prepossessing methods are required to enhance the accuracy of heart failure prediction?
3. How can I achieve the highest possible precision and recall in heart failure detection?
4. What are the challenges in generalizing the model across different patient populations?

1.5 Report Layout

Into six chapters the article is designed, each detailing different features of our research:Chapter 1, The research topic, outlines the importance of heart failure prediction, and presents the research questions, motivation, and objectives the article introduces. Chapter 2,This chapter reviews existing work on heart failure prediction, discussing various approaches, limitations, results, and methods used by other researchers. It also addresses the scope and challenges of our research. Chapter 3, Research Methodology, This section details the methodology used in our research, including data collection procedures, data statistics, classifiers used, and implementation requirements. Chapter 4, Results, This section presents the results of our research, showing the performance of different classifiers and the effectiveness of our model. Chapter 5, Impact and Implications, This chapter discourses the potential impact of our research on the healthcare industry, emphasizing its importance and sustainability. Finally, Chapter 6, Discussion and Conclusion, This chapter discusses on the research findings, conclusions and suggestions for future work. It also addresses the limitations of the current study.

CHAPTER 2
BACKGROUND

2.1 Introduction

In healthcare field, timely and accurate diagnosis is crucial for patient outcomes and operational efficiency. Predicting heart failure early remains a significant challenge. I utilize machine learning algorithms to enhance heart failure prediction by applying various models to historical patient data, identifying patterns to predict future cases. This method helps healthcare providers detect potential issues early, reducing diagnostic and treatment costs. Traditional approaches consuming time and prone-error, often resulting in late detection. Machine learning offers a solution by automating the prediction process, analyzing medical data to focus on high-risk patients. Our research aims to revolutionize healthcare by improving early detection, optimizing resources, and enhancing patient care. This proactive approach can transform heart failure management and overall healthcare quality.

2.2 Related Works

A number of research studies have helped us figure out how to predict software bugs. Here, I look at the different ways that experts have looked into to predict software bugs.

Authors of [1] have made Incident HF Prediction in the Aging. This study evolves and authenticates a prediction model for incident heart failure in elderly individuals. The HF Score is a compact model that uses 9 common clinical variables to accurately predict the 5-year risk of heart failure. The model had good discrimination and calibration and performed better than the Framingham Heart Failure Risk Score when applied to the current study population. HF Score enables the identification of high-risk elderly individuals and can guide preventive strategies, with the goal of alleviating the burden of heart failure in this population by preventing its occurrence.

Jing Wang et al. [2] have made Heart Failure Prediction using machine learning algorithms. Various methods were employed, including classifiers, deep learning, boosted decision trees, and deep neural networks. The research compares 18 models using F1 score and accuracy metrics, highlighting the effectiveness of the SMOTE method for data imbalance. Results show z-score normalization outperforming min-max normalization. The study addresses the need for comprehensive model comparisons and introduces innovative approaches like Deep Forest.Significant implications for enhancing early detection and management of heart failure is these findings.

The study paper [3] presents a machine learning-based approach for predicting HF using common clinical variables. It develops and validates a risk prediction model, called the Health ABC HF Score, which demonstrates strong discrimination and calibration for 5-year incident heart failure risk in an elderly population. The model outperforms the existing Framingham Heart Failure Risk Score and can detect high-risk individuals who may benefit from tailored prevention strategies. The study highlights the potential of leveraging readily available patient data and advanced analytics to enhance heart failure risk assessment, a critical public health challenge given the aging population

The authors of [4] risk prediction in patients with heart disaster. The review on this paper on risk prediction of heart failure identified key predictors across various models. Demographics, commodities, and physiological measures were commonly used for risk assessment. Social and quality-of-life factors were less frequently considered. The study reviewed 43 main models for death prediction, 10 for hospitalization, and 11 for combined outcomes. Most studies focused on hospitalized patients, with varying data collection timings. Left ventricular ejection fraction was a significant predictor in many models. The review highlighted the importance of accurate risk prediction for guiding clinical decision-making in HF management.

Another Malaysian Journal of Computing study [5] by Huang et al. compared four ML techniques to predict early HF. RF has the highest recall and sensitivity instead of Naïve Bayes, SVM, LR. The dataset was organized and splits into training and testing sets. The best results were derived from SVM with a linear kernel function. This study underlines the significance of precise heart failure prediction to ensure proper treatment. This research provides valuable information about the paper's topic, as it examines how machine learning can help detect heart failure prematurely.

The research paper [6] shows that the Seattle HF model prediction of survival in heart failure. The Seattle HF Model is a predictive tool for estimating 1-, 2-, and 3-year survival rates in heart failure patients. It uses easily obtainable clinical, pharmacological, device, and laboratory characteristics to make predictions. The model was derived and validated using data from multiple cohorts, showing excellent accuracy in predicting survival rates. It aims to assist physicians in counseling patients about prognosis and making informed decisions regarding treatment options.

Douglas S. Lee et al. [7] study that developed a model for predicting acute heart failure mortality in emergency care settings. The model utilized clinical data from over 12,000 patients , creatinine concentration, BP, oxygen saturation, and serum troponin levels to stratify mortality risk with high discrimination. The study found that lower initial systolic blood pressure, oxygen saturation, and hemoglobin concentration, as well as higher leukocyte count and potassium, creatinine. A c-statistic of 80.5% in the derivation set and 82.6% in the validation set multivariate model showed high discrimination. The study also developed the EHMRG score, comprising additive and multiplicative variables, to calculate 7-day mortality risk.

The authors of [8] discuss the evolution of risk prediction models in heart failure, highlighting the shift towards ML and AI and techniques for improved predictive accuracy. It evaluates a study on heart failure with preserved ejection fraction using various predictive models, noting the limitations and potential future directions in the field.

Improving risk prediction in HF using ML shows that the research paper [9]. This paper presents the development and validation of a ML based on risk prediction model, named MARKER-HF, for mortality in HF patients. The model was derived using a boosted decision tree algorithm on a large datasets from the University of California, San Diego, and demonstrated excellent discriminatory power with an area under the curve of 0.88. External validation in two independent cohorts also showed superior performance compared to other established risk scores. The paper highlights the advantages of using machine learning techniques to capture complex multi-dimensional interactions in patient data for improving risk prediction in heart failure.

Wayne C. Levy et al. [10] developed heart failure risk prediction models. The model utilizes the challenges in validating and developing effective risk prediction models for heart failure patients. It highlights the importance of both discrimination and calibration for clinical utility, using examples from coronary artery disease models. The authors review a comprehensive analysis of heart failure risk models and recommend starting with a baseline age-sex model to better quantify the added value of new variables. Key recommendations include reporting changes in C-statistic and model chi-square, as well as conducting rigorous external validation in diverse populations before clinical implementation of these risk models.

The authors of [11] Suggested multiple biomarkers for risk prediction in chorionic heart HF. The study paper investigated whether a panel of multiple biomarkers could improve risk prediction in chronic heart failure patients beyond the commonly used Seattle HF Model . The researchers measured 8 biomarkers related to different biological pathways in 1,513 chronic heart failure patients and derived a parsimonious multimarker score. They found that this multimarker score was a strong liberated analyst of adverse outcomes like death, transplant, or ventricular assist device placement, and significantly improved risk prediction compared to the SHFM alone. The Extra multimarker score to the SHFM improved individual risk discrimination, reclassifying 25% of patients into higher risk categories. These findings back up the use of a multimarker approach to enhance prognosis in chronic HF patients.

The research paper [12] has made a risk model to predict outcomes in chronic heart failure patients with elevated heart rate. The researchers identified the top 10 predictors of adverse outcomes, which included increased resting heart rate ,levated creatinine, advanced symptoms, longer disease duration, older age, low BP, and history of left bundle branch block. The SHIFT Risk Model demonstrated good discriminatory ability, with c-statistics ranging from 67.6% to 69.5% for the various outcomes examined. This simple, clinically applicable risk model may help guide management decisions and target higher-risk patients for more intensive therapy in chronic heart failure.

This review paper [13] showed machine learning techniques applied to various facts of heart failure management. It covers methods for detecting HF classifying its subtypes, estimating disease severity, and predicting adverse events like destabilizations, rehospitalizations, and mortality. The authors highlight the pros and cons of the different approaches, noting that while progress has been made. Overall, the paper demonstrates the potential of data-driven approaches to improve risk stratification,diagnosis and treatment of this complex chronic condition.

The Seattle HF Model (SHFM) developed by Dariush Mozaffarian et al. [14]. TO predicts (sudden vs. pump failure) among 10,538 HF patients. The SHFM strongly predicted both the relative risk and the proportions of deaths due to sudden versus pump failure across different risk scores. Patients with higher SHFM scores had significantly greater risks of pump failure death compared to sudden death. These findings suggest the SHFM could provide prognostic information to guide the use of specific HF medications or devices targeted at preventing different modes of morality. Further research is warranted to determine the clinical utility of these predictions.

The research paper [15] shows that ML forecast of morality and hospitalization in HF with preserved ejection fraction. This study developed machine learning models to predict mortality and HF hospitalization in HFpEF patients. The random forest model was most accurate, with blood tests, body mass index, and patient-reported health status as key predictors. These models can help assess prognosis and guide care for this complex syndrome lacking predictive tools. The findings highlight the value of collecting patient-reported outcomes to improve risk assessment in HFpEF.

The authors of this paper [16] have implemented multiple machines learning models such as Naive Bayes, Decision Tree, Random Forest(RF), Logistic Regression(LR), and SVM, to predicts heart failure disease using a UCI heart disease datasets. The models were evaluated using 10-fold cross-validation, and the results showed significant improvement in accuracy compared to previous studies. The Decision Tree and SVM models achieved the highest accuracy of over 90%. The paper highlights the value of using ML techniques to enhance the prediction of heart failure, which can aid clinicians in earlier diagnosis and treatment.

Hong Yang at el. Have made [17] the strength of association between common clinical factors and the risk of developing heart failure (HF) in community-based populations. The strongest independent predictors were body mass index, male gender, hypertension, smoking, age, diabetes, and coronary artery disease.

This study paper [18] developed a heart failure (HF) risk prediction model .Multi-Ethnic Study of Atherosclerosis (MESA) cohort this is the source of using data. Key predictors included age, gender, smoking, body mass index, blood pressure, heart rate, diabetes, and N-terminal pro-B-type natriuretic peptide. The model had excellent discrimination and identify high-risk HF individuals without pre-existing cardiac disease. This risk assessment tool can help primary care providers target preventive strategies for individuals at high risk of developing HF.

This review paper [19] developed a heart failure (HF) risk prediction model using ML algorithms. As well find the HF patients accurately for specific assessments.This risk assessment system can help primary care providers target preventive strategies for individuals at high risk of developing HF.

GianLucaDiTanna at el. [20] evaluating risk prediction model for aged people with HF. This literature review identified 58 risk prediction models for HF, evaluating their methodological quality and performance. Most models had concerns for bias, primarily due to lack of validation and calibration. Common predictors included age, sex, diabetes, and biomarkers like natriuretic peptides. Discrimination was generally moderate to good, but inferior for predicting heart failure hospitalizations. Overall, the review highlights the need for more rigorously developed and validated risk prediction tools to guide clinical care and optimize outcomes in heart failure patients.

This thesis paper [21] provides a comparison of several ML methods for predicting heart disease.The study uses the Cleveland heart disease datasets from the UCI repository and considers 14 attributes to evaluate the performance of these algorithms which is Decision-Tree,K-Nearest Neighbor, Naive Bayes (NB), and Random Forest(RF). The results show that the K-Nearest Neighbor algorithm achieves the highest accuracy outperforming the other techniques. This study Provides useful information about how to use data mining and ML algorithms to diagnose heart disease early and accurately.

The authors of this paper [22] shows a diagnostic prediction heart failure(HF) model that combines clinical assessment and NT-proBNP levels to improve the diagnosis of acute HF in patients with undifferentiated dyspnea. The model was derived and externally validated, showing excellent discrimination, especially in patients with indeterminate clinical probability. The model appropriately reclassified a significant proportion of these intermediate-risk patients to either low or high probability of acute heart failure, with minimal inappropriate redirection. The use of NT-proBNP as a continuous variable, rather than relying on specific cut-off values, appears to enhance the diagnostic accuracy of the model.

This study paper [23] shows 20 risk prediction models for death in ambulatory HF patients. Only 5 models were externally validated, with the HF Survival Score and Seattle HF Model being the most extensively studied. These models showed modest discrimination and questionable calibration, with performance potentially impacted by changes in heart failure management over time. The clinical impact of using these models to guide decision-making remains unexplored.

The author of this paper [24] shows that clinical prediction rule to identify HF patients admitted from the emergency room who are low-risk. The authors analyzed data from over 33,000 heart failure patients in Pennsylvania and developed a rule using 21 prognostic factors. The rule classified 17.2% of patients as low risk, with a 0.3% inpatient mortality rate and 1.0% rate of serious complications. The 30-day mortality and readmission rates were also low in the low-risk group. The authors conclude that this prediction rule can identify a substantial proportion of heart failure patients who may be suitable for outpatient management, though further validation is needed.

The research paper [25] shows that risk prediction models for incident HF-heat failure. The review found that while the models had acceptable-to-good discrimination in derivation samples, most have not been externally validated. Only a few models were validated in multiple cohorts, with modest-to-acceptable discrimination. The incorporation of novel biomarkers modestly improved risk prediction in some studies. However, the review found no recommendations for using these models in clinical practice guidelines, and the effects of implementing the models on clinical outcomes remain unclear. The authors conclude that further research is needed to develop and validate robust incident heart failure risk prediction models that can guide effective prevention strategies.

This study paper [26] developed a Balanced-Random-Forest(BRF) classifier to predict the survival of HF patients using their clinical records. The BRF classifier incorporated under sampling into the model construction.Two feature selection techniques were used to identify the key predictors of survival which is Chi-square test and Recursive Feature Elimination. The proposed approach achieved a maximum G-mean score of 0.768 and a sensitivity of 0.821, significantly outperforming other machine learning models. The study identified ejection fraction, serum creatinine, and patient age as the most important factors in predicting heart failure mortality.

This paper [27] proposes a novel hybrid ML algorithm called Hybrid Random-Forest (HRF) with Linear Model (HRFLM) to effectively predict heart disease. The method uses all 13 features of the UCI heart disease dataset without any restrictions on feature selection.

Experiments show that the proposed HRFLM outperforms existing methods in terms of accuracy, sensitivity, specificity, precision, and F-measure. The authors highlight the importance of using a combination of techniques to increase the performance of HF prediction.

The authors of this paper [28] show that the clinical applications of ML in the diagnosis, classification, and prediction of HF. It introduces various ML techniques and their uses in heart failure research. The paper discusses how machine learning can improve the detection and prediction of heart failure, classify patients into distinct phenotypes, and identify responders to advanced therapies. The authors highlight the potential of machine learning to enhance personalized medicine in heart failure management, while also addressing the current limitations and obstacles in widespread clinical implementation.

Wouther Ouwerkerk at el. [29] have made systematic reviews and compares existing prediction models for death and HF hospitalization in patients with HF. It identifies the strongest predictor variables, such as blood urea nitrogen and sodium, and the model characteristics associated with higher predictive powerThe analysis indicates that mortality predictions are most accurate when using prospective registry-type studies that incorporate a wide range of clinical variables. However, the overall predictive capability of the models remains only moderate, highlighting the need for further improvement in risk prediction for heart failure patients.

The study paper [30] focuses on predicting heart disease using ML algorithms like (LR), KNN, and RF classifier.Dataset have 13 medical attributes of 304 patients. The proposed model achieves an accuracy of 87.5%, with KNN outperforming the other algorithms with 88.52% accuracy. The paper emphasizes the importance of using multiple ML techniques to enhance the prediction of heart disease. Showed result proposed model is cost-effective and can help doctors diagnose heart disease efficiently.

2.3 Scope of the problem

Our research aims to develop a model that predicts HF using ML algorithms, providing healthcare professionals with an early-warning tool. However, this task is challenging due to the complexity and imbalance of medical data, where different risk factors often exhibit overlapping characteristics. Additionally, the variability in patient demographics and health conditions makes it difficult to generalize the model. Despite these challenges, our initial experiments with a Random Forest classifier show promising results, indicating high accuracy and the potential for effective heart failure prediction.

2.4 Challenges

The primary challenge in developing our heart failure prediction model is achieving high precision and accuracy. I need a large, high-quality dataset with comprehensive health information to improve the model's performance. One challenge is the variance in patient demographics and health conditions, which influences model generalization. Furthermore, distinguishing between similar health conditions given overlapping symptoms is challenging. Our model needs to adapt to both common and unique patterns of heart failure to be effective. Nonetheless, our preliminary results are encouraging, suggesting that with further refinement, our model can significantly aid in early heart failure detection.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

A road map for our thesis and the method I applied the section of the article serves. It supplies us with the most comprehensive data and tactics. In our situation, this entails a focused delivery followed by realization. So, our data gathering efforts will have the same result. It will demonstrate how I use those to prediction heart failure using machine learning methods. This is part two of the research methodology.

3.2 Research Subject and Instrumentation

Develop a more reliable approach to predicting heart failure the main focus of this article .To achieve this, I analyzed various patient health datasets to identify patterns that could indicate the onset of heart failure. The healthcare industry faces significant challenges in managing heart disease, particularly due to the subtle and non-specific nature of early HF symptoms. Clinicians and healthcare providers often struggle to identify potential heart failure early in the disease progression, which can lead to increased morbidity, mortality, and healthcare costs. Accurately predicting heart failure can help in early intervention, improving patient outcomes, and reducing hospital admissions.

I employed machine learning algorithms, specifically concentrating on classification methods that use patient data from the past to estimate the risk of heart failure. These models can filter out irrelevant features and identify significant correlations between clinical attributes and heart failure occurrences. Our study investigates the effectiveness of various ML techniques in heart failure prediction and compares them to traditional methods to identify the most successful approach.

3.3 Data Collection Procedure

Collecting relevant and high-quality data is essential for our study. I utilized publicly available medical datasets, including details like demographic information, medical history, clinical measurements, and lifestyle factors. I gathered approximately 1,200 patient records from several studies and healthcare institutions, categorized into instances of heart failure and non-heart failure. All collected data were formatted consistently to ensure accuracy and ease of analysis.

3.3.1 Datasets

TABLE 1: ATTRIBUTE DEFINATION

Metric	Definition
Age	The patient age (years)
Sex	Patient gender (0-female, 1-male)
CP	Chest pain types (0-Typical angina, 1-Atypical angina, 2-non-anginal pain, 3-Asymptomatic)
chol	mg/dl of Serum Cholesterol
restecg	Resting electrocardiogram (0-Normal, 1-Abnormal ST-T wave, 2-Indicating likely or certain left ventricular hypertrophy)
fbs	Fasting blood sugar level(fbs) set as above 120 mg/dl (0-false, 1-true)
thalachh	Maximum heart rate achieved during a stress test exng: Exercise induced angina (0-No & 1-Yes)
oldpeak	ST depression induced by exercise relative to rest (unit -> depression)
slp	Slope of the peak exercise ST segment (0-Upsloping, 1-Flat, 2- Down sloping)
trtbps	mmHg for resting blood sugar
caa	The outcome of the Thallium stress test (0-Normal, 1-Fixed defect, 2-Reversible defect, 3-Not characterized) gives the count of major vessels (0–4) colored in the fluoroscopic.
output	Heart disease status (0-no disease, 1-presence of disease

TABLE 2: DATASET PROPERTIES

Dataset name	Number of Attribute	Number of Instances
Heart1	11	918
Heart2	13	304

3.4 Machine Learning Techniques

I have employed numerous ML techniques to predict HF. These approaches cover a broad spectrum of learner types, including Ensemble Learners, Bayesian Learners, Neural Networks, Support Vector Machines, Distance-Based methods, and Tree-Based Learners.

From these categories, I selected eight machine learning techniques to accurately predict heart failure. Figure 1 illustrates the algorithms used, along with their respective categories. Below is a list of each algorithm.

To evaluate the effectiveness of each learning method, I used a 3-fold Cross-Validation (CV) model.

Fig. 3.5.1. Sequential Model Diagram

3.4.1 Ensemble Learners

Bagging (Bootstrap Aggregation):

Introduced by Leo Breiman, this ensemble technique involves creating multiple subsamples from both the training dataset and the predictive model. These subsamples are selected through random replacement. The final model is a combination of several models, which helps reduce variance and enhance accuracy.

Random Forest:

This is a grouping of algorithms based on trees. Random Forests create many decision trees during the learning phase and either choose the option that is most common among them for classification or average the predictions for regression. By merging the forecasts from several decision trees, this method enhances the ability to generalize and lessens the risk of over fitting.

3.4.2 Bayesian Learner

Naive Bayes:

Naive Bayes is a simplistic probabilistic classifier that applies Bayes' theorem while assuming strong (naive) independence across the features. Due to its simplicity and efficiency, it is widely used in classification tasks.

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \qquad (1)$$

3.4.3 Neural Networks

MLP (Multilayer Perceptron):

The Multilayer Perceptron (MLP) is a form of artificial neural network that consists of multiple layers of interconnected nodes. Each node in a layer is directly connected to the nodes in the subsequent layer. This method is highly effective for supervised learning tasks such as classification and regression, and it excels at identifying complex patterns within data.

Radial Basis Function (RBF):

A Radial Basis Function (RBF) network represents a type of artificial neural network that employs radial basis functions for activation purposes. RBF networks are commonly used for tasks such as function approximation and classification. They consist of a hidden layer where each neuron calculates the distance between the input and a center, and an output layer used for classification or regression.

3.4.3 Support Vector Machine

SVM (Support Vector Machine):

Support Vector Machine (SVM) is a type of supervised machine learning technique employed for classification and regression purposes. The SVM algorithm operates by identifying the hyperplane that most effectively separates the classes in the feature space, while also maximizing the distance between them. SVMs are highly efficient in high-dimensional environments. Through the use of various kernel functions SVMs can manage non-linear decision boundaries.

3.4.4 Distance-Based Methods

KNN (K-Nearest Neighbors):

KNN, an abbreviation for K-Nearest Neighbors, is a basic ML approach used for both classification and regression. It works by finding the K closest data points to a given query point within the feature space to predict outcomes based on the labels or values of these neighboring points.

3.4.6 Tree-Based Learner

Decision tree algorithm:

The decision tree is a commonly used supervised learning method for tasks involving classification and regression. It divides the feature space into subsets based on input feature values through recursive partitioning. This process forms a tree-like structure where internal nodes represent decisions using features, and leaf nodes signify class labels or numerical values. Decision trees are valued for their interpretability and flexibility in dealing with both numerical and categorical data, making them useful in various applications.

3.5 Data Preprocessing

Effective data preparation is crucial for creating reliable and precise machine learning models. This section outlines the techniques necessary to prepare datasets for model training and evaluation.

3.5.1 Handling Missing values

Missing values in the datasets were addressed using the following strategies:

Imputation: Imputation is a statistical method for filling in missing values in a datasets. The purpose of imputation is to create a complete datasets by guessing missing values using the existing data. Missing numerical values were imputed using the mean of the respective feature, while missing categorical values were imputed using the mode.

3.5.2 Standardization

In order to maintain equal contribution of each feature to the model's performance, numerical features were standardized to have an average of 0 and a standard deviation of 1.

3.5.3 Handling Imbalanced Data

The datasets frequently showed an uneven distribution of classes. To address this issue, oversampling methods such as SMOTE (Synthetic Minority Over-Sampling Technique) were used to balance the class distribution.

3.5.4 Hyper-parameter Tuning

Hyper-parameter tuning was performed to optimize the model's performance. Techniques such as Grid Search and Random Search were used to find the best combination of hyper-parameters.

3.6 Statistical Analysis

In this section, I present a comprehensive statistical analysis of the machine learning models used for predicting heart failure. The primary goal is to assess the performance, reliability, and robustness of these models in predicting heart failure in patients.

3.6.1 Statistical Analysis

First, I will provide a concise overview of the important descriptive statistics of the datasets that were utilized for both training and testing the models. This encompasses calculations of both central tendency and dispersion for each individual characteristic within the datasets.

Mean and Median: The average and middle values of the features.

Standard Deviation (SD): The measure of the amount of variation or dispersion of the features **Range:** The difference between the maximum and minimum values.

3.6.2 Implementation Requirements

To evaluate how well learning algorithms perform, commonly used metrics include accuracy, precision, recall, and the F-measure. The confusion matrix (sometimes called an error matrix) is used to summarize the prediction results for a classification problem. Model evaluation is crucial for classification issues that involve multiple output categories.

The confusion matrix includes True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values.

- Positive (P): The observation is identified as positive.
- Negative (N): The observation is identified as Negative.
- True Positive (TP): Predicted correctly positive cases.
- False Negative (FN): Predicted incorrectly negative cases.
- True Negative (TN): Predicted correctly negative cases.
- False Positive (FP): Predicted incorrectly positive cases.

Accuracy: The ratio of correctly predicted instances to the total instances. The following equation provides the accuracy, often known as the classification rate:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Precision: The proportion of accurately predicted positive observations to total anticipated positives.

$$\text{Precision}= \frac{TP}{TP+FN} \quad (3)$$

Recall (Sensitivity): The ratio of correctly predicted positive observations to all observations in the actual class.

$$\text{Recall}= \frac{TP}{TP+FN} \quad (4)$$

F1 Score: The harmonic mean of precision and recall, providing a single metric that balances both concerns.

$$\text{Precision}= \frac{2*precision*Recall}{Precision+Recall} \quad (5)$$

Area Under the Receiver Operating Characteristic Curve (AUC-ROC): A measurement of the model's the capacity to differentiate between classes.
Confusion Matrix: A table that illustrate the performance of a classification model, showing the true positives, true negatives, false positives, and false negatives.

3.6.3 Cross validation

To ensure the robustness of the models, I perform k-fold cross-validation:
k-Fold Cross-Validation: The datasets are divided into k subsets, and the model is trained and validated k times, each time using a different subset as the validation set and the remainingk-1subsets as the training set. This helps in understanding the model's performance and stability across different data splits.

3.6.2 Model Comparison

Using Various machine learning models to find the one that works best for predicting software defects:
Visualization: To see how well a model is doing, use box plots, ROC curves, and a training vs. validation accuracy and loss graph.
Models for Ranking: I put the models in order of performance metrics and statistical tests to find the best one for this prediction job.

3.7 Proposed Methodology

In this research, I propose a model to predict heart failure using datasets collected from the Kaggle public repository. To design an effective predictive system for heart failure, I conducted extensive experimentation with several machine learning algorithms. I employed K-fold Cross Validation (CV) for each learning algorithm to ensure robust model verification. Initially, the datasets underwent preprocessing to handle missing data, encode categorical variables, and balance the datasets using SMOTE (Synthetic Minority Over-Sampling Technique). Subsequently, the data were standardized and split into training and test sets. After determining the features and target attributes, I trained and tested each model. To enhance performance, I utilized hyper-parameter tuning. I evaluated the models using various metrics such as accuracy, precision, recall, F1 score, and also examined the confusion matrix, ROC curve, validation loss, and validation accuracy to comprehensively understand the models' performance. The model demonstrating the best evaluation metrics was finalized for predicting heart failure.

3.7.1 Proposed Model Workflow

In order to effectively predict the software defects, A range of actions must be undertaken. The flowchart that represents the proposed model process of this work is shown in Figure 3.7.1.

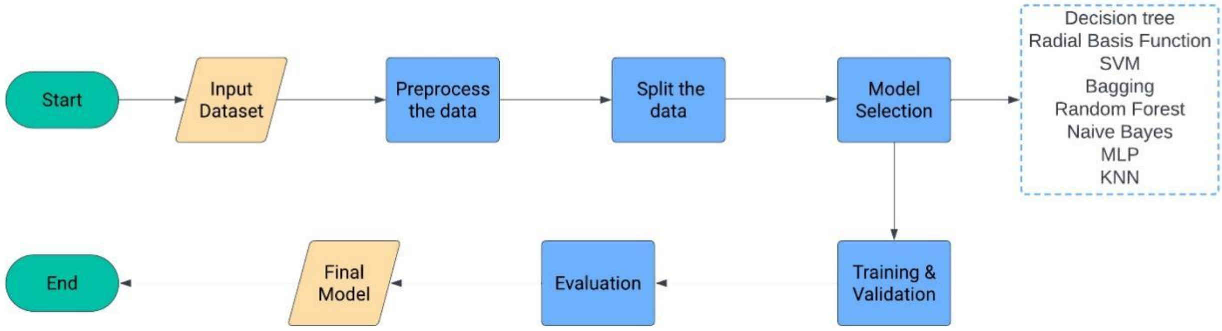

Fig. 3.7.1. Proposed Model Workflow

3.8 Implementation Requirements

To implement the project for software defect prediction using machine learning, specific hardware and software requirements are essential:

Hardware Requirements:

- Windows 7/10 or above operating system.
- Trans-Portable or in-built Hard Disk with a capacity of at least 300 GB
- Minimum 2 GB RAM to handle data processing efficiently
- Google Chrome or Mozilla Firefox.

Software Requirements:

- Mozilla Firefox web browsers for installation

Developing Tools

- Python with Tensor flow library for ML Implementation
- Jupiter (Google Colab) for interactive development and documentation
- Notepad++ for code editing

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

In this paper, I explore the vivid analysis of the data I have used in our study, and I demonstrate the experimental results obtained from our project on heart failure prediction using machine learning. I acknowledge the necessity for validating with different methodological and theoretical approaches to guarantee generalization findings. I also assess our vision in a quantitative manner by employing our proposed strategy as a dynamic framework. This chapter clarifies the value proposition of the framework (internally and externally) and displays its construction, and its boundaries in terms of performance.

4.2 Experimental Setup

The experimental setup involved several steps to ensure comprehensive evaluation and comparison of various machine learning algorithms:

1. **Data Preprocessing:** Handling lost data, encrypting categorical variables, oversampling with SMOTE, and standardizing features.
2. **Train-Test Split:** The datasets are divided into training and testing sets using the conventional 70-30 split.
3. **Model Selection:** Evaluating several machines learning algorithms, including Random Forest, Bagging, and other methods.
4. **K-fold Cross-Validation:** Using 3-fold cross-validation for robust model validation.
5. **Hyper-parameter Tuning:** Fine-tuning model parameters to optimize performance.6.
6. **Evaluation Metrics:** Assessing models using accuracy, precision, recall, F1score, confusion matrix, and ROC curve.

4.3 Experimental Results & Analysis

In this Study, I developed and analyzed eight various models using four separate datasets to forecast software problems. The assessment of these models employed a blend of metrics, however certain studies depend on a single metric. The main criteria I used for evaluation were accuracy, graphical analysis, and confusion matrices.

TABLE 3: RESULTS FOR EACH DATASETS

Dataset	Metrics	Decision Tree	Radial Basis Function	Support Vector Machine	Bagging	Random Forest	Naive Bayes	MLP (Multilayer Perceptron)	KNN
Heart1	Accuracy	81.37%	87.75%	87.75%	85.29%	**88.24%**	87.25%	82.35%	87.75 %
	Precision	0.842	0.882	0.888	**0.898**	0.877	0.881	0.857	0.895 7
	Recall	0.827	0.905	0.896	0.836	**0.922**	0.896	0.827	0.887
	F1 Score	0.834	0.893	0.892	0.866	**0.899**	0.888	0.842	0.891
Heart2	Accuracy	78.79%	84.85%	77.27%	**86.36%**	81.82%	77.27%	81.82%	74.24 %
	Precision	0.794	0.833	0.771	**0.878**	0.789	0.743	0.805	0.729
	Recall	0.794	0.882	0.794	0.852	**0.882**	0.852	0.852	0.794
	F1 Score	0.794	0.857	0.782	**0.865**	0.833	0.794	0.828	0.760

For any dataset and machine learning technique, the optimal classification performance outcome is displayed in bold fonts. Across four different datasets, various machine learning algorithms were employed to predict outcomes, with each dataset exhibiting different characteristics. In Dataset 1 (Heart1), random Forest demonstrated the maximum accuracy, precision, recall, and F1 score, showcasing its overall effectiveness. Meanwhile, in Dataset 2 (Heart2), Random Bagging emerged as the top performer inaccuracy and F1 score, while Bagging achieved the highest precision. These results highlight the significance of selecting the suitable algorithm based on the specific characteristics and requirements of each dataset to achieve optimal predictive performance.

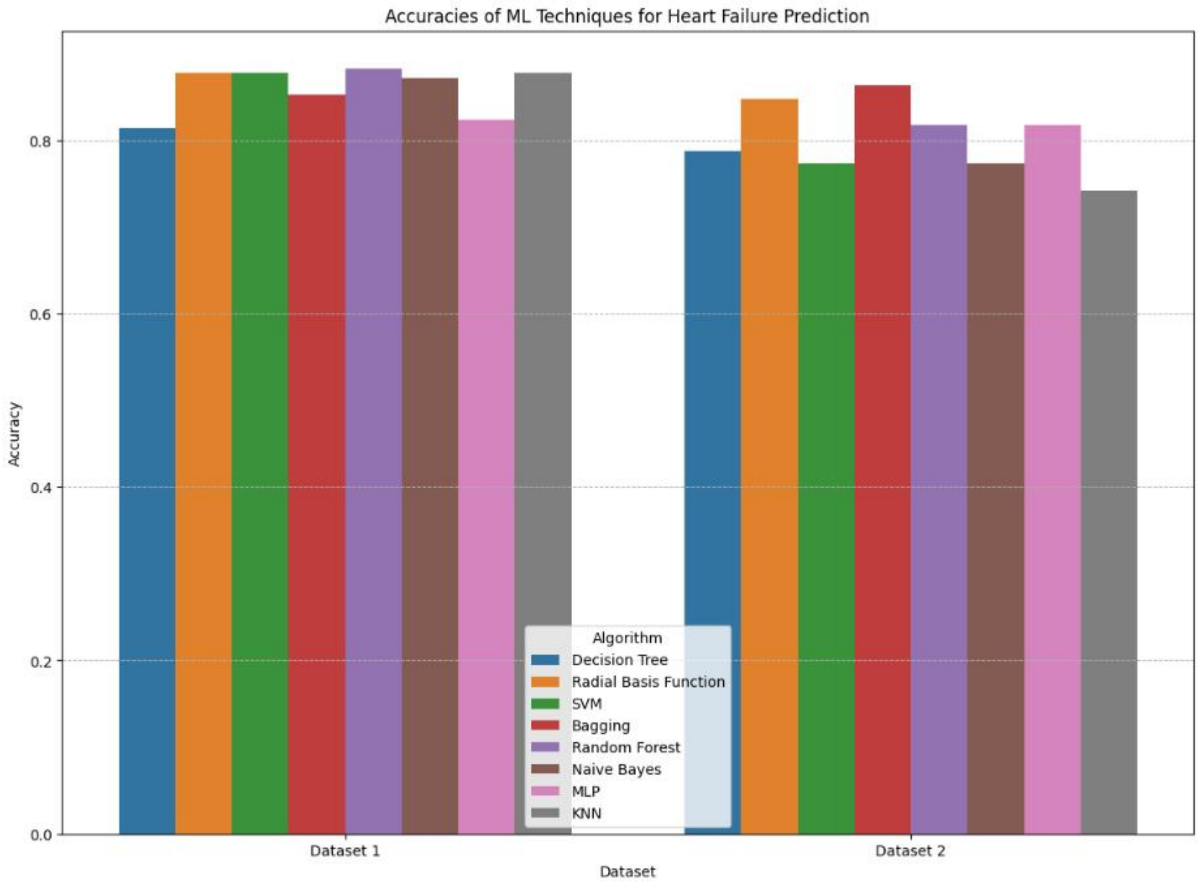

Fig. 4.3.1. Accuracies of ML Techniques for Heart failure Prediction

TABLE 4: AUC & PR AUC FOR EACH DATASETS

Dataset	**Metrics**	**Decision Tree**	**Radial Basis Function**	**Support Vector Machine**	**Bagging**	**Random Forest**	**Naive Bayes**	**MLP (Multilayer Perceptron)**	**KNN**
Heart1	AUC	0.811	0.930	0.924	0.914	**0.941**	0.925	0.910	0.912
	PR PUC	0.794	0.940	0.933	0.888	**0.950**	0.937	0.923	0.896
Heart2	AUC	0.787	0.856	0.859	0.871	**0.891**	0.831	0.865	0.832
	PR AUC	0.736	0.852	0.771	**0.825**	0.890	0.851	0.890	0.808

In this work, 20 epochs have been worked and the following Figures shows the graph between training accuracy vs validation accuracy and training loss vs validation loss for every dataset.

Fig. 4.3.2. Training vs Validation Accuracy & Loss (Random Forest in Heart1)

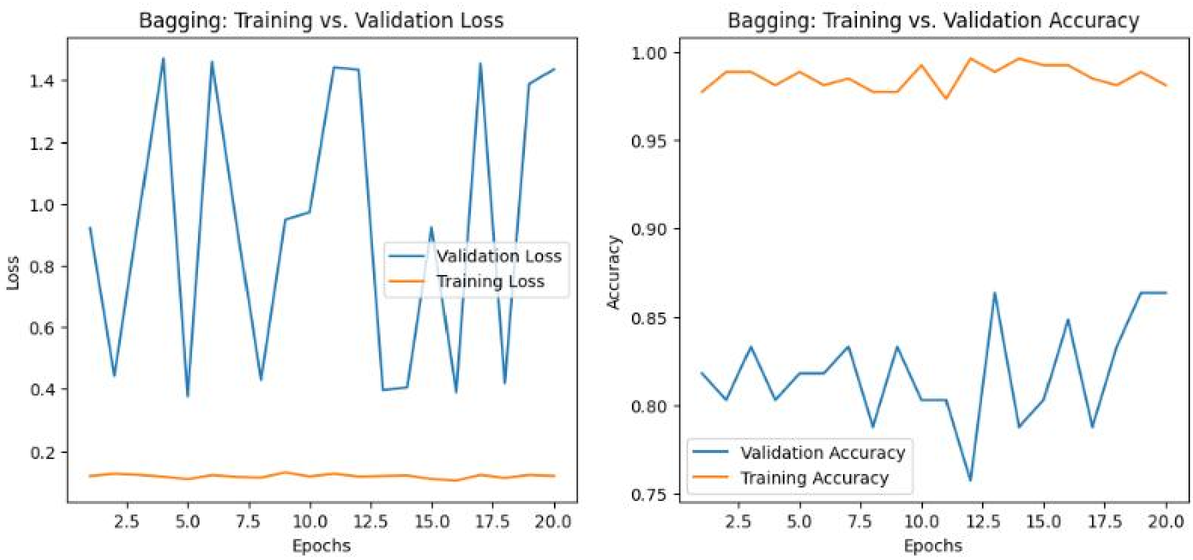

Fig. 4.3.2. Training vs Validation Accuracy & Loss (Bagging in Heart2)

4.4 Discussion

The results show that ensemble learners, especially Random Forest and Bagging, were very good at predicting bugs in software across a variety of datasets. Random Forest did better than all the other algorithms in terms of accuracy and other evaluation measures after hyper-parameter tuning was applied. In particular:

Random Forest: Demonstrated superior performance with accuracies of 88% Dataset 1. It also showed high precision and recall, indicating its robustness in defect prediction.
Bagging: Achieved notable results with accuracies of 86% on Dataset 1.It performed well in terms of precision and recall, making it a reliable choice for defect prediction.

The strong performance of these ensemble methods underscores their effectiveness in handling the complexities of heart failure prediction. Random Forest, in particular, emerged as the best-performing algorithm after hyper-parameter tuning, making it a highly reliable model for practical applications in the healthcare industry.

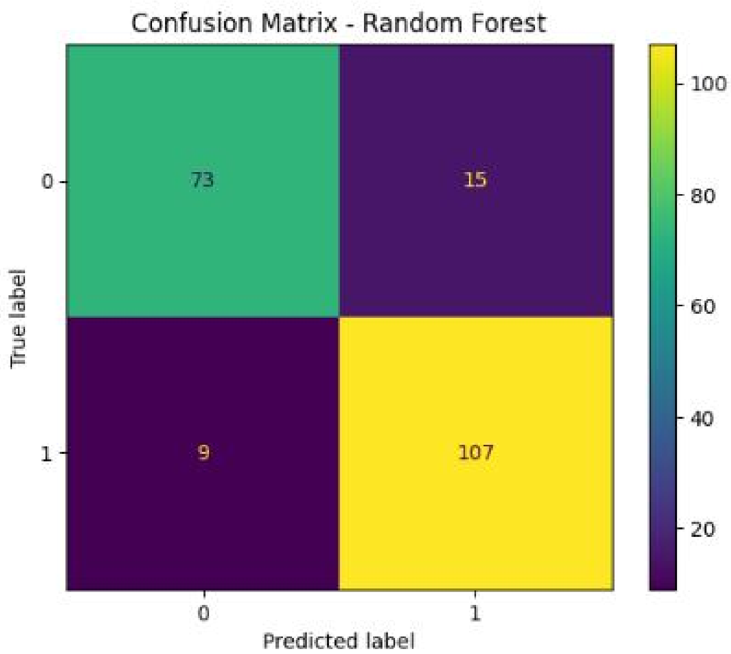

Fig. 4.4.1. Confusion Matrix (Random Forest in Heart1)

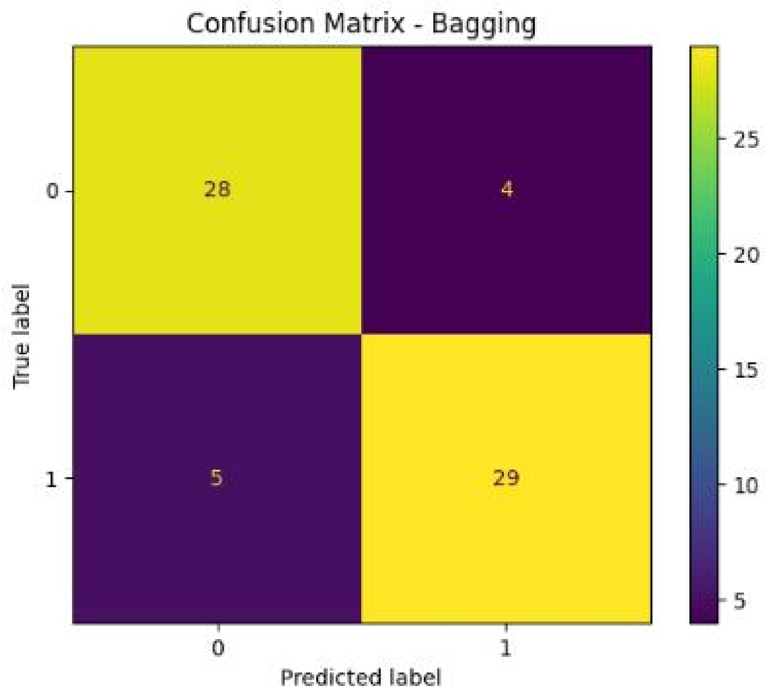

Fig. 4.4.1. Confusion Matrix (Random Forest in Heart2)

The Random Forest algorithm consistently outperformed others, exhibiting superior accuracy rates on Dataset 1, in the evaluation of two datasets. The confusion matrices corroborated these findings, as Random Forest exhibited high precision and recall in defect prediction tasks, thereby demonstrating its reliability. In contrast, Bagging exhibited significantly improved performance on Dataset 2, with higher accuracies. Random forest's efficacy in defect prediction across diverse datasets was further reinforced by the confusion matrices that highlighted its strong precision and recall values.

.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Society

Heart failure worldwide is a significant concern, leading to high morbidity, mortality, and healthcare costs. Predicting HF using ML algorithms is crucial for addressing this issue. By analyzing extensive datasets to identify patterns and risk factors, machine learning enables early detection and timely intervention, significantly improving patient outcomes and reducing heart failure incidence. Personalized risk assessments allow for tailored treatment plans, enhancing the management and overall quality of patient care. Accurate predictions help healthcare systems optimize resource allocation, prioritize high-risk patients, and reduce unnecessary tests and treatments, ultimately lowering healthcare costs. Early and precise predictions lead to timely interventions that slow disease progression, reduce hospitalizations, and improve patients' quality of life. This technology is particularly beneficial for low-income and underserved communities, providing cost-effective predictive tools that help bridge healthcare disparities and ensure high-quality care for more people. By improving heart failure prediction, this research contributes to better health outcomes, advances medical knowledge, resulting in a healthier society generally.

5.2 Impact on Environment

The application of machine learning in predicting heart failure, while primarily focused on healthcare, also has notable environmental impacts. Efficient predictive algorithms can optimize healthcare operations by reducing the need for extensive and repeated medical tests and procedures, thereby lowering energy consumption in medical facilities. Additionally, improved management and early intervention can decrease the reliance on frequent and resource-intensive treatments, contributing to a reduction in medical waste. By enhancing the precision of heart failure prediction, this research supports more sustainable healthcare practices, leading to reduced energy usage and waste generation. Environmentally, this research promotes a greener healthcare system and aligns with broader goals of minimizing the carbon footprint of the medical industry.

5.3 Ethical Aspects

This This research on predicting heart failure using machine learning adheres to strict ethical guidelines, ensuring the security and privacy of patient data. The results presented are accurate and verifiable, and the research has been conducted with integrity and transparency. No unethical practices were involved, and the study was carried out with full collaboration among team members under the guidance of our supervisor. The ethical implications of deploying heart failure prediction models, such as ensuring fairness and avoiding bias in medical assessments, have also been carefully considered. Additionally, patient consent and confidentiality were prioritized throughout the research process, upholding the highest standards of ethical conduct in medical research.

5.4 Sustainability Plan

The prediction of heart failure using machine learning is a continuous and evolving process that is critical for the sustainability of the healthcare industry. This research aims to provide long-term benefits by improving the accuracy and efficiency of heart failure detection, thereby enhancing patient care and outcomes. By assisting healthcare providers in identifying potential heart failure risks early, this research can significantly reduce the time and resources required for treatment and management. Expanding the range of medical data and employing additional algorithms to further improve prediction accuracy future work will focus. The goal is to develop a robust, scalable solution that can adapt to the evolving needs of the healthcare industry and contribute to its sustainable development. This includes ensuring the technology remains accessible and effective in diverse clinical settings, ultimately supporting a more resilient and sustainable healthcare system.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Summary of the study

Following the end of the research, I concluded that predicting and identifying heart failure is of utmost significance for ensuring the reliability and efficacy of healthcare systems. High-quality predictive models are essential for maintaining patient health and safeguarding critical medical data across various sectors. The main concern of this study was to construct a viable ML algorithm for predicting heart failure utilizing datasets from the Kaggle public repository. By employing various machine learning algorithms and applying K-fold Cross-Validation for model validation, I were able to preprocess the data effectively, handle missing values, and balance the datasets. The ultimate goal was to enhance patient outcomes and prevent potential failures that could have significant impacts on society.

6.2 Conclusions

The major purpose of this work was to create a machine learning-based method for forecasting Heart Failure. The findings show that machine learning models, particularly those fine-tuned via hyperparameter optimization, are extremely effective in this domain. Specifically, the Random Forest and Bagging algorithms accurately identified faults.

Random Forest achieved an accuracy of 88% on Dataset 1 while Bagging reached 86% on Dataset 1. Ensemble learners, in general, performed admirably. After applying hyperparameter tuning, Random Forest outperformed all other algorithms. In terms of precision and other evaluation metrics, both Random Forest and Bagging showed excellent performance.

I were able to predict heart failure with excellent accuracy and reliability through considerable experimentation, as demonstrated by multiple evaluation measures including precision, recall, and F1 score. Because of the possible savings in expenses and resources related to patient care and management, this research holds great promise for the healthcare sector. But the study also found several shortcomings, mostly with regard to the accuracy of the models and the quality of the data. Predictive power can be increased with larger datasets and more algorithmic optimization. The development of a comprehensive tool or application that incorporates these predictive models and gives medical practitioners real-time information into potential heart failure concerns will be the main focus of future effort.

6.3 Implication for further study

For future research, I recommend several enhancements to improve the accuracy and applicability of heart failure prediction models. Firstly, increasing the size and diversity of the dataset can lead to better generalization and model performance. Secondly, exploring and integrating additional machine learning algorithms may provide more robust predictions. Thirdly, incorporating contextual factors such as patient demographics and lifestyle attributes can further refine predictions. In future work,I aim to:

- Improve model accuracy by expanding and refining the dataset.
- Apply a broader range of machine learning models to identify the most effective approaches.
- Develop a real-time prediction tool that can assist healthcare professionals in monitoring and managing heart failure risks.
- Incorporate environmental and contextual factors into the prediction model to enhance its reliability and applicability.

By focusing on these specific areas, our aim is to make significant contributions to the advancement of healthcare systems, resulting in improved reliability and efficiency. Ultimately, this will have a positive impact on the overall medical and societal environment.

REFERENCE

[1] Butler, Javed, et al. "Incident heart failure prediction in the elderly: the health ABC heart failure score." *Circulation: Heart Failure* 1.2 (2008): 125-133.

[2] Wang, Jing. "Heart failure prediction with machine learning: a comparative study." *Journal of Physics: Conference Series*. Vol. 2031. No. 1. IOP Publishing, 2021.

[3] Sahoo, Prasanta Kumar, and Pravalika Jeripothula. "Heart failure prediction using machine learning techniques." *Available at SSRN 3759562* (2020).

[4] Rahimi, Kazem, et al. "Risk prediction in patients with heart failure: a systematic review and analysis." *JACC: Heart Failure* 2.5 (2014): 440-446.

[5] Mansur Huang, Nur Shahellin, Zaidah Ibrahim, and Norizan Mat Diah. "Machine learning techniques for early heart failure prediction." *Malaysian Journal of Computing (MJoC)* 6.2 (2021): 872-884.

[6] Levy, Wayne C., et al. "The Seattle Heart Failure Model: prediction of survival in heart failure." *Circulation* 113.11 (2006): 1424-1433.

[7] Lee, Douglas S., et al. "Prediction of heart failure mortality in emergent care: a cohort study." *Annals of internal medicine* 156.11 (2012): 767-775.

[8] Gottdiener, John S., and Alison E. Fohner. "Risk prediction in heart failure: new methods, old problems." *JACC: Heart Failure* 8.1 (2020): 22-24.

[9] Adler, Eric D., et al. "Improving risk prediction in heart failure using machine learning." *European journal of heart failure* 22.1 (2020): 139-147.

[10] Levy, Wayne C., and Inder S. Anand. "Heart failure risk prediction models: what have we learned?." *JACC: Heart Failure* 2.5 (2014): 437-439.

[11] Ky, Bonnie, et al. "Multiple biomarkers for risk prediction in chronic heart failure." *Circulation: Heart Failure* 5.2 (2012): 183-190.

[12] Ford, Ian, et al. "Top ten risk factors for morbidity and mortality in patients with chronic systolic heart failure and elevated heart rate: The SHIFT Risk Model." *International journal of cardiology* 184 (2015): 163-169.

[13] Tripoliti, Evanthia E., et al. "Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques." *Computational and structural biotechnology journal* 15 (2017): 26-47.

[14] Mozaffarian, Dariush, et al. "Prediction of mode of death in heart failure: the Seattle Heart Failure Model." *Circulation* 116.4 (2007): 392-398.

[15] Angraal, Suveen, et al. "Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction." *JACC: heart failure* 8.1 (2020): 12-21.

[16] Alotaibi, Fahd Saleh. "Implementation of machine learning model to predict heart failure disease." *International Journal of Advanced Computer Science and Applications* 10.6 (2019).

[17] Yang, Hong, et al. "Clinical prediction of incident heart failure risk: a systematic review and meta-analysis." *Open heart* 2.1 (2015): e000222.

[18] Chahal, Harjit, et al. "Heart failure risk prediction in the Multi-Ethnic Study of Atherosclerosis." *Heart* 101.1 (2015): 58-64.

[19] Guo, Aixia, et al. "Heart failure diagnosis, readmission, and mortality prediction using machine learning and artificial intelligence models." *Current Epidemiology Reports* 7 (2020): 212-219.

[20] Di Tanna, Gian Luca, et al. "Evaluating risk prediction models for adults with heart failure: A systematic literature review." *PLoS One* 15.1 (2020): e0224135.

[21] Shah, Devansh, Samir Patel, and Santosh Kumar Bharti. "Heart disease prediction using machine learning techniques." *SN Computer Science* 1.6 (2020): 345.

[22] Steinhart, Brian, et al. "Improving the diagnosis of acute heart failure using a validated prediction model." *Journal of the American College of Cardiology* 54.16 (2009): 1515-1521.

[23] Alba, Ana C., et al. "Risk prediction models for mortality in ambulatory patients with heart failure: a systematic review." *Circulation: Heart Failure* 6.5 (2013): 881-889.

[24] Auble, Thomas E., et al. "A prediction rule to identify low-risk patients with heart failure." *Academic emergency medicine* 12.6 (2005): 514-521.

[25] Echouffo-Tcheugui, Justin B., et al. "Population risk prediction models for incident heart failure: a systematic review." *Circulation: Heart Failure* 8.3 (2015): 438-447.

[26] Newaz, Asif, Nadim Ahmed, and Farhan Shahriyar Haq. "Survival prediction of heart failure patients using machine learning techniques." *Informatics in Medicine Unlocked* 26 (2021): 100772.

[27] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." *IEEE access* 7 (2019): 81542-81554.

[28] Olsen, Cameron R., et al. "Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure." *American Heart Journal* 229 (2020): 1-17.

[29] Ouwerkerk, Wouter, Adriaan A. Voors, and Aeilko H. Zwinderman. "Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure." *JACC: Heart Failure* 2.5 (2014): 429-436.

[30] Jindal, Harshit, et al. "Heart disease prediction using machine learning algorithms." *IOP conference series: materials science and engineering*. Vol. 1022. No. 1. IOP Publishing, 2021.

Predictive Modeling of Heart Failure Integrating Clinical Data with Machine Learning.docx

ORIGINALITY REPORT

20%	16%	11%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4%
2	Submitted to Coventry University Student Paper	1%
3	123dok.com Internet Source	1%
4	www.mdpi.com Internet Source	1%
5	dokumen.pub Internet Source	1%
6	www.mjbdsc.org Internet Source	1%
7	Submitted to Middlesex University Student Paper	<1%
8	Submitted to University of Hertfordshire Student Paper	<1%
9	Submitted to University of North Texas Student Paper	<1%