A dissertation project on

Developing and validating machine learning models for predicting hospital readmission in heart failure patients.

By

Yetunde Adegunleye

Student No: Q102092734

MSC in Applied AI and Data Science

Solent University Southampton

Contents

[List of Figures 5](#_Toc174954126)

[List of Table 5](#_Toc174954127)

[Abstract 6](#_Toc174954128)

[1. Introduction 7](#_Toc174954129)

[1.1 Background 7](#_Toc174954130)

[2.2 Problem Statement 7](#_Toc174954131)

[2.3 Research Questions 8](#_Toc174954132)

[2.4 Objectives 8](#_Toc174954133)

[2.5 Significance of Study 8](#_Toc174954134)

[2.6 Summary of Introduction Chapter 9](#_Toc174954135)

[Chapter 3: Literature Review 10](#_Toc174954136)

[3.1 Overview of Heart Failure 10](#_Toc174954137)

[3.1.1 Pathophysiology and Epidemiology 10](#_Toc174954138)

[3.1.2 Clinical Management and Challenges 10](#_Toc174954139)

[3.2 Machine Learning in Healthcare 11](#_Toc174954140)

[3.2.1 Evolution and Application of Machine Learning in Healthcare 11](#_Toc174954141)

[3.2.2 Challenges and Ethical Considerations 12](#_Toc174954142)

[3.3 Predictive Modelling for 30-Day Readmissions 12](#_Toc174954143)

[3.3.1 Importance of Predictive Modelling in Reducing Readmissions 12](#_Toc174954144)

[3.3.2 Machine Learning Techniques in Predictive Modelling 13](#_Toc174954145)

[3.3.3 Limitations and Future Directions 13](#_Toc174954146)

[3.4 Conclusion of the literature review chapter 14](#_Toc174954147)

[Chapter 4: Methodology 15](#_Toc174954148)

[4.1 Data Collection 15](#_Toc174954149)

[4.2 Dataset Description 15](#_Toc174954150)

[4.1.2 Data Collection Process 16](#_Toc174954151)

[4.2 Data Preprocessing 16](#_Toc174954152)

[4.2.1 Handling Missing Data 16](#_Toc174954153)

[4.2.2 Data Transformation 16](#_Toc174954154)

[4.2.3 Outlier Detection and Handling 16](#_Toc174954155)

[4.2.4 Splitting the Dataset 17](#_Toc174954156)

[4.3 Feature Selection 17](#_Toc174954157)

[4.3.1 Domain Knowledge-Based Selection 17](#_Toc174954158)

[4.3.2 Algorithmic Feature Selection 17](#_Toc174954159)

[4.4 Model Selection 18](#_Toc174954160)

[4.4.1 Logistic Regression (LR) 18](#_Toc174954161)

[4.4.2 Support Vector Machine (SVM) 18](#_Toc174954162)

[4.4.3 Random Forest (RF) 18](#_Toc174954163)

[4.4.4 Gradient Boosting Machine (GBM) 18](#_Toc174954164)

[4.4.5 Neural Networks (NN) 19](#_Toc174954165)

[4.4.6 K-Nearest Neighbours (KNN) 19](#_Toc174954166)

[4.5 Evaluation Metrics 19](#_Toc174954167)

[4.5.1 Accuracy 19](#_Toc174954168)

[4.5.2 Precision and Recall 19](#_Toc174954169)

[4.5.3 F1-Score 20](#_Toc174954170)

[4.5.4 Area Under the Receiver Operating Characteristic Curve (AUC-ROC) 20](#_Toc174954171)

[4.5.5 Confusion Matrix 20](#_Toc174954172)

[4.6 Conclusion of the Methodology Chapter 20](#_Toc174954173)

[5. Data Analysis and Visualization 21](#_Toc174954174)

[5.1 Introduction 21](#_Toc174954175)

[5.2 Overview of the Dataset 21](#_Toc174954176)

[5.3 Data Exploration 22](#_Toc174954177)

[5.3.1 Summary Statistics 22](#_Toc174954178)

[5.3.2 Missing Values and Outliers 23](#_Toc174954179)

[5.3.3 Correlation Analysis 23](#_Toc174954180)

[5.4 Feature Analysis and Visualization 23](#_Toc174954181)

[5.4.1 Age Distribution 24](#_Toc174954182)

[5.4.2 Ejection Fraction Analysis 24](#_Toc174954183)

[5.4.3 Serum Creatinine and DEATH\_EVENT 24](#_Toc174954184)

[5.4.4 Creatinine Phosphokinase Levels 25](#_Toc174954185)

[5.4.5 Platelet Count Distribution 25](#_Toc174954186)

[5.4.6 Serum Sodium and Outcome Analysis 26](#_Toc174954187)

[5.4.7 Smoking and Outcome Analysis 26](#_Toc174954188)

[5.4.8 Gender Distribution and Outcomes 27](#_Toc174954189)

[5.5 Feature Interaction Analysis 27](#_Toc174954190)

[5.5.1 Interaction Between Age and Ejection Fraction 28](#_Toc174954191)

[5.5.2 Interaction Between Serum Creatinine and Serum Sodium 28](#_Toc174954192)

[5.6 Importance of Data Visualization 28](#_Toc174954193)

[5.7 Summary of the Data and Visualization Chapter 29](#_Toc174954194)

[5.8 Model Performance 29](#_Toc174954195)

[5.8.1 Overview of Model Performance 29](#_Toc174954196)

[5.8.2 Gradient Boosting Machine (GBM) 30](#_Toc174954197)

[5.8.3 Random Forest (RF) 30](#_Toc174954198)

[5.8.4 Logistic Regression (LR) 30](#_Toc174954199)

[5.8.5 Support Vector Machine (SVM) 31](#_Toc174954200)

[5.8.6 K-Nearest Neighbors (KNN) 31](#_Toc174954201)

[5.8.7 Neural Networks (NN) 31](#_Toc174954202)

[5.9 Feature Importance 32](#_Toc174954203)

[5.9.1 Feature Importance in Random Forest (RF) 32](#_Toc174954204)

[5.9.2 Feature Importance in Gradient Boosting Machine (GBM) 33](#_Toc174954205)

[5.10 Interpretation of Results 33](#_Toc174954206)

[5.11 Conclusion of model performance and feature importance 33](#_Toc174954207)

[Chapter 6: Discussion 35](#_Toc174954208)

[6.1 Interpretation of Results 35](#_Toc174954209)

[6.1.1 Model Performance 35](#_Toc174954210)

[6.1.2 Feature Importance 36](#_Toc174954211)

[6.2 Comparison with Previous Studies 36](#_Toc174954212)

[6.2.1 Model Performance 37](#_Toc174954213)

[6.2.2 Feature Importance 37](#_Toc174954214)

[6.3 Implications for Healthcare 38](#_Toc174954215)

[6.3.1 Clinical Practice 38](#_Toc174954216)

[6.3.2 Healthcare Policy 38](#_Toc174954217)

[6.3.3 Future Research Directions 39](#_Toc174954218)

[6.3.4 Conclusion of the Discussion 39](#_Toc174954219)

[Chapter 7: Conclusion 40](#_Toc174954220)

[7.1 Summary of Findings 40](#_Toc174954221)

[7.2 Limitations 42](#_Toc174954222)

[7.3 Recommendations for Future Research 42](#_Toc174954223)

[References 44](#_Toc174954224)

[Appendix 49](#_Toc174954225)

[1. Findings and Figure selection code snippet. 49](#_Toc174954226)

[2.Dissertation Analysis and Visualization Code and Output: 50](#_Toc174954227)

[3. Code snippet for Streamlit and visualisation 53](#_Toc174954228)

[4. Streamlit screenshot: 56](#_Toc174954229)

[5. Streamlit application link: 57](#_Toc174954230)

# List of Figures

[Figure 1: Showing code snippet for correlation Matrix 24](#_Toc174544067)

[Figure 2: Histogram code snippet showing the age distribution of patients 25](#_Toc174544068)

[Figure 3: Visualization code snippet for the patient with death\_event. 25](#_Toc174544069)

[Figure 4: Showing code snippet for analysis for patients with high serum creatine levels. 26](#_Toc174544070)

[Figure 5: Analysis code snippet for distribution of CPK levels. 26](#_Toc174544071)

[Figure 6: Showing code distribution of platelet count. 27](#_Toc174544072)

[Figure 7: Visualization code snippet for violin plot for a patient experiencing lower serum levels, which leads to death\_event. 27](#_Toc174544073)

[Figure 8: Code snippet for count plot showing smoking status vs death\_event. 28](#_Toc174544074)

[Figure 9: count plot code snippet showing higher death\_event among male patients. 28](#_Toc174544075)

[Figure 10: Scatter plot code snippet showing the relationship between ejection fraction and death\_event. 29](#_Toc174544076)

[Figure 11: Code snippet showing the effect of renal impairment and electrolyte imbalance leading to death\_event in patients. 29](#_Toc174544077)

[Figure 12: Feature Importance Ranking from Random Forest Model. 41](#_Toc174544078)

[Figure 13: ROC Curves for Different Models 42](#_Toc174544079)

# List of Table

[Table 1: Statistics showing variables in the dataset 24](#_Toc174544642)

[Table 2: Performance Metrics for Predictive Models 31](#_Toc174544643)

[Table 3: Showing an Explanation of the Limitation 43](#_Toc174544644)

[Table 4: Showing future research recommendations 44](#_Toc174544645)

# 

# Abstract

Heart failure (HF) is a widespread and debilitating condition that affects millions of people globally, leading to high rates of hospitalisation and significant healthcare costs. One of the most significant challenges in managing heart failure is the high rate of 30-day hospital readmissions. This indicates ongoing issues in patient care and places an additional burden on healthcare systems. However, with the potential of machine learning to significantly improve patient outcomes and reduce the overall cost of care, there is a reason to be optimistic about the future of healthcare.

Setting itself apart from existing literature, the research begins with a comprehensive review of heart failure, machine learning in healthcare, and current predictive models for hospital readmissions. This review underscores traditional models' limitations and machine learning's potential to offer more effective predictions.

The study reveals that machine learning models, especially those using ensemble methods like random forests and gradient boosting, outperform traditional statistical models in predicting 30-day readmissions. Key features contributing to the risk of readmission include patient demographics, clinical history, and hospital discharge variables, providing valuable insights for developing targeted interventions.

In the discussion, we explore the implications of the findings in the context of existing literature and healthcare practice. The study shows that machine learning (ML) models can improve the accuracy of readmission predictions, leading to more effective resource allocation and potentially better patient outcomes. Compared with previous studies, this study highlights the progress made by ML techniques in this area while recognising the challenges associated with implementing them in clinical settings.

The dissertation summarises the key findings, discusses study limitations, and suggests future research. It contributes to applying machine learning to healthcare, offering practical strategies for integrating predictive models into routine clinical practice. The research shows the potential impact of reducing readmission rates and improving care for heart failure patients.

# 1. Introduction

## 1.1 Background

Heart failure (HF) is a major global public health issue that affects millions of individuals, leading to high rates of morbidity and mortality. It is a complex clinical syndrome that occurs when the heart is unable to pump enough blood to meet the body's needs, often due to underlying conditions such as coronary artery disease, hypertension, or diabetes (Ponikowski et al., 2016). Despite advancements in medical treatments, heart failure remains a leading cause of hospitalisation, especially among older adults, and places a significant burden on healthcare systems worldwide (Heidenreich et al., 2013).

In recent years, there has been growing interest in using machine learning (ML) techniques to address healthcare challenges, particularly in predictive modelling and decision support systems. ML can potentially analyse large datasets, uncover hidden patterns, and predict patient outcomes more accurately than traditional statistical methods (Libbrecht and Noble, 2015). One critical area ML can significantly impact is predicting 30-day readmissions for patients with heart failure. Early identification of patients at high risk of readmission can enable timely interventions, reduce healthcare costs, and improve patient outcomes (Dharmarajan et al., 2013).

## 2.2 Problem Statement

Heart failure-related hospital readmissions are a significant concern for both patients and healthcare providers. About 20-25% of heart failure patients are readmitted within 30 days of discharge, leading to increased healthcare costs and suggesting potential gaps in care (Jencks, Williams, and Coleman, 2009). It is crucial to predict which patients are at risk of readmission to take preventive measures and improve the quality of care. However, traditional risk prediction models often lack accuracy and fail to capture the complexity of patient data, resulting in suboptimal outcomes (Kansagara et al., 2011).

This dissertation addresses the need for more accurate and reliable predictive models for 30-day readmissions among heart failure patients. By using advanced machine learning techniques, the study seeks to develop models that can better identify high-risk patients and provide insights into the factors contributing to readmissions. This approach can improve clinical decision-making and optimise resource allocation in healthcare settings.

## 2.3 Research Questions

This research aims to address the following questions:

1. How can machine learning models be used to predict 30-day readmissions in patients with heart failure?

2. What are the primary features and factors contributing to the readmission risk in patients with heart failure?

3. How does the performance of machine learning models compare to traditional statistical methods in predicting readmissions in heart failure cases?

4. What are the implications of using machine learning models in clinical practice for managing patients with heart failure?

## 2.4 Objectives

The main objectives of this dissertation are:

1. Developing and validating machine learning models to predict 30-day readmissions in heart failure patients.

2. Identifying the most significant features that influence the risk of readmission and evaluating their impact on model performance.

3. Comparing the predictive accuracy of machine learning models with traditional risk prediction methods.

4. Exploring the potential clinical applications of machine learning models in managing heart failure patients and discussing their implications for healthcare practice.

## 2.5 Significance of the Study

The study is critical because it has the potential to enhance the management of heart failure patients by using advanced predictive modelling techniques to reduce 30-day readmissions. This is crucial for healthcare providers as it cuts healthcare costs and improves patient outcomes and quality of life (Khera et al. 2018). The study aims to use machine learning to understand better the factors leading to readmissions, which can then be used to develop targeted interventions and care strategies.

Additionally, this research adds to the growing body of literature on the use of machine learning in healthcare, demonstrating the potential of this technology to revolutionise clinical practice (Wiens and Shenoy 2018). The findings of this study could also have broader implications for the development of predictive models in other areas of healthcare, highlighting the value of data-driven approaches in improving patient care.

## 2.6 Summary of Introduction Chapter

In summary, this study fills an essential gap in the literature by focusing on applying machine learning to a specific and high-impact issue in healthcare. The results could lead to more effective management of heart failure patients, ultimately contributing to better health outcomes and more sustainable healthcare systems.

# Chapter 3: Literature Review

This chapter thoroughly reviews critical literature, focusing on three key areas: an overview of heart failure, the use of machine learning in healthcare, and predictive modelling for 30-day readmissions. Each section analyses existing research, identifies gaps, and places the current study within the broader academic discussion.

## 3.1 Overview of Heart Failure

Heart failure (HF) is a complex syndrome characterised by the heart's inability to pump sufficient blood to meet the body's metabolic needs. It represents a significant global health burden, affecting approximately 64 million people worldwide (Vos et al., 2020). The condition is associated with high morbidity, frequent hospitalisations, and a substantial reduction in quality of life.

### 3.1.1 Pathophysiology and Epidemiology

The development of heart failure involves a combination of myocardial injury, neurohormonal activation, and hemodynamic changes that collectively impair the heart's ability to function effectively (Ponikowski et al., 2016). Heart failure is broadly categorised into reduced ejection fraction (HFrEF) and heart failure with preserved ejection fraction (HFpEF). HFrEF, often associated with systolic dysfunction, occurs when the heart's left ventricle loses its ability to contract effectively, leading to decreased cardiac output (McMurray et al., 2012). On the other hand, HFpEF is linked to diastolic dysfunction, where the left ventricle becomes stiff and unable to relax properly, resulting in impaired ventricular filling (Dunlay & Roger, 2014). The prevalence of heart failure increases with age, particularly in individuals over 65 years old (Benjamin et al., 2019). Risk factors include hypertension, coronary artery disease, diabetes, and obesity (Heidenreich et al., 2013). Notably, the burden of heart failure is higher among men for HFrEF, while women are more likely to develop HFpEF (Ponikowski et al., 2016). The condition is a significant contributor to healthcare costs, primarily driven by the high hospitalisation and readmission rates (Benjamin et al., 2019).

### 3.1.2 Clinical Management and Challenges

The clinical management of heart failure requires a comprehensive strategy, including medications, lifestyle changes, and, in some cases, device therapy or surgery. Essential medications such as ACE inhibitors, beta-blockers, and mineralocorticoid receptor antagonists have been proven to reduce mortality and morbidity in heart failure patients (Yancy et al., 2017). Despite advancements in treatment, the prognosis for heart failure remains poor, with a five-year mortality rate exceeding 50% following diagnosis (Benjamin et al., 2019).

A significant challenge in managing heart failure is the high rate of hospital readmissions, especially within 30 days of discharge. Research shows that almost 25% of heart failure patients are readmitted within this period, often due to exacerbations, inadequate discharge planning, or insufficient follow-up care (Dharmarajan et al., 2013). These readmissions indicate poor patient outcomes and result in significant financial penalties for healthcare systems, particularly under policies like the Hospital Readmissions Reduction Program (HRRP) in the United States (Khera et al., 2018).

## 3.2 Machine Learning in Healthcare

Machine learning (ML) is a groundbreaking technology in healthcare that has the potential to enhance diagnostics, personalise treatments, and optimise healthcare delivery. By using large datasets and complex algorithms, ML can reveal patterns in data that are not easily identifiable through traditional statistical methods.

### 3.2.1 Evolution and Application of Machine Learning in Healthcare

The use of machine learning in healthcare has rapidly expanded due to improvements in computational power, algorithmic development, and the increasing availability of healthcare data (Shickel et al., 2017). Machine learning techniques have been applied in various areas, such as medical imaging, genomics, and electronic health records (EHRs), leading to significant advancements in disease prediction, risk stratification, and personalised treatment (Obermeyer & Emanuel, 2016).

For instance, convolutional neural networks (CNNs) in medical imaging have shown outstanding performance in detecting and categorising abnormalities in radiographic images, outperforming traditional methods in accuracy and efficiency (Litjens et al., 2017). Similarly, in genomics, machine learning has been used to identify genetic mutations linked to diseases, facilitating the development of targeted therapies (Libbrecht & Noble, 2015).

EHRs offer a rich source of data for machine learning applications. Techniques such as natural language processing (NLP) have been employed to extract valuable information from unstructured clinical notes, allowing for the prediction of patient outcomes and the identification of disease progression patterns (Miotto et al., 2016). Integrating diverse data types—clinical, genetic, and behavioural—machine learning models can provide a more comprehensive understanding of patient health and support more accurate clinical decision-making (Shickel et al., 2017).

### 3.2.2 Challenges and Ethical Considerations

Machine learning (ML) in healthcare comes with several challenges. One major issue is the quality of the data. Healthcare data, specifically from electronic health records (EHRs), can be incomplete, inconsistent, and prone to errors, which can significantly impact the performance of ML models (Shickel et al., 2017). Furthermore, the high complexity and dimensionality of healthcare data require advanced algorithms and substantial computational resources, which can hinder the widespread adoption of ML in clinical practice (Obermeyer & Emanuel, 2016).

Another significant challenge is the ability of ML models to generalise. Models trained on data from specific populations or healthcare settings may not perform well when applied to different contexts due to differences in patient demographics, clinical practices, and healthcare infrastructure (Wiens & Shenoy, 2018). This issue of model portability is crucial, as healthcare systems vary widely across regions, and a model's success in one setting does not guarantee its effectiveness in another.

Ethical considerations also play a vital role in using ML in healthcare. Patient privacy, data security, and algorithmic bias are significant concerns that must be addressed to ensure that ML applications do not unintentionally harm patients or worsen existing health disparities (Char et al., 2018). For example, if an algorithm is trained on data underrepresenting certain demographic groups, it may produce biased predictions that disproportionately disadvantage those populations (Obermeyer & Mullainathan, 2019). Therefore, transparency, fairness, and accountability are essential in developing and implementing ML models in healthcare (Char et al., 2018).

## 3.3 Predictive Modelling for 30-Day Readmissions

Predictive modelling, especially machine learning techniques, has become essential in identifying patients at risk of 30-day readmissions. Accurate predictions can assist healthcare providers in implementing targeted interventions, reducing readmission rates, improving patient outcomes, and minimising associated costs.

### 3.3.1 Importance of Predictive Modelling in Reducing Readmissions

Reducing 30-day readmissions is a crucial goal for healthcare systems, as these readmissions often indicate poor patient outcomes and increase healthcare costs (Jencks et al., 2009). Predictive models can help identify high-risk patients before discharge, allowing healthcare providers to implement early interventions such as improved discharge planning, close follow-up care, and patient education (Kansagara et al., 2011). Several predictive models have been created to predict the likelihood of 30-day readmissions, particularly among patients with chronic conditions such as heart failure, chronic obstructive pulmonary disease (COPD), and diabetes (Frizzell et al., 2017). These models generally use clinical, demographic, and socioeconomic data to generate risk scores, which can be used to prioritise patient interventions (Hass et al., 2020).

### 3.3.2 Machine Learning Techniques in Predictive Modelling

Machine learning techniques have significantly improved the accuracy and complexity of predictive modelling for 30-day readmissions, often surpassing traditional statistical approaches. Some standard machine-learning techniques used in this area include logistic regression, decision trees, random forests, gradient-boosting machines, and deep-learning models.

* Logistic Regression: This technique has been a standard tool in predictive modelling due to its simplicity and interpretability. However, its performance is often limited by the assumption of linear relationships between predictors and outcomes, which may not capture the complex interactions in healthcare data (Kansagara et al., 2011).
* Decision Trees and Random Forests: Decision trees provide a simple, interpretable model that can handle non-linear relationships and interactions between variables. Random forests aggregate the predictions of multiple decision trees, offering improved accuracy and robustness by reducing overfitting (Breiman, 2001).
* Gradient Boosting Machines (GBM): GBM is an ensemble technique that builds models sequentially, with each new model correcting the errors of the previous ones. This approach performs better in predicting 30-day readmissions, particularly in complex datasets with many interacting variables (Chen & Guestrin, 2016).
* Deep Learning: Deep learning models, particularly neural networks, have gained popularity for their ability to model intricate, non-linear relationships in large datasets. Despite their predictive power, these models are often criticised for their lack of interpretability. This can be a significant drawback in clinical settings where understanding the reasoning behind predictions is crucial (Ching et al., 2018).

### 3.3.3 Limitations and Future Directions

Machine learning models have shown promise in predicting 30-day readmissions, but some limitations must be addressed. The opaque nature of many advanced models, especially deep learning, presents a significant challenge in clinical settings where interpretability and trust are crucial (Rudin, 2019). Additionally, the reliance on retrospective data may not capture all relevant variables or changes in clinical practice, limiting the models' applicability to current patient populations (Wiens & Shenoy, 2018).

Future research should focus on developing interpretable models that maintain high predictive accuracy while providing clear insights into the factors driving predictions. Moreover, models that can be easily adapted and validated across different healthcare settings are needed to ensure their generalizability and robustness (Obermeyer & Emanuel, 2016). Integrating real-time data, such as patient monitoring and genomics, into predictive models could enhance their accuracy and utility in clinical practice (Shickel et al., 2017).

### 3.4 Conclusion of the literature review chapter

The literature underscores the complexity of heart failure, the potential of machine learning in healthcare, and the role of predictive modelling in reducing readmissions. Challenges in data quality, model interpretability, and generalizability persist. Overcoming these challenges is crucial for improving patient outcomes. The upcoming chapters will detail the methodology and results of a study aiming to develop and evaluate predictive models for heart failure outcomes using machine learning techniques.

# Chapter 4: Methodology

This chapter outlines the methodological framework used in this dissertation to develop and evaluate predictive models for 30-day readmissions and mortality among heart failure patients. The chapter is structured into five key sections: Data Collection, Data Preprocessing, Feature Selection, Model Selection, and Evaluation Metrics. Each section details the processes and techniques, ensuring a comprehensive understanding of the steps to achieve the study's objectives.

## 4.1 Data Collection

The data used in this study was sourced from the publicly available UCI Machine Learning Repository's Heart Failure Clinical Records dataset. This dataset contains medical records of 299 patients who were diagnosed with heart failure. The data was initially collected from six different hospitals in the United Kingdom and made available by the University of California, Irvine, for research.

## 4.2 Dataset Description

The dataset comprises 13 clinical features, including demographic, laboratory, and clinical data. These features are patient data:

* Age: The age of the patient.
* Anaemia: Whether the patient has a history of anaemia (1 = Yes, 0 = No).
* High Blood Pressure: Whether the patient has high blood pressure (1 = Yes, 0 = No).
* Creatinine Phosphokinase (CPK): Level of the CPK enzyme in the blood (mcg/L).
* Diabetes: Whether the patient has diabetes (1 = Yes, 0 = No).
* Ejection Fraction: Percentage of blood leaving the heart at each contraction.
* Platelets: Platelet count in the blood (kilo platelets/mL).
* Serum Creatinine: Creatinine level in the blood (mg/dL).
* Serum Sodium: Level of sodium in the blood (mEq/L).
* Sex: Gender of the patient (Male = 1, Female = 0).
* Smoking: Whether the patient is a smoker (1 = Yes, 0 = No).
* Time: Follow-up period (days).
* Death Event: Whether the patient died during the follow-up period (1 = Yes, 0 = No).

The target variables for this study were the likelihood of a death event and the prediction of 30-day readmission, which was derived based on the follow-up period and corresponding clinical outcomes.

### 4.1.2 Data Collection Process

The dataset was downloaded directly from the UCI Machine Learning Repository. Given that the dataset is openly available and widely used in the research community, no further ethical approvals were required for its use in this study. However, all analyses were conducted in compliance with ethical guidelines to ensure the responsible use of data, including maintaining the confidentiality of the information contained within the dataset.

## 4.2 Data Preprocessing

Data preprocessing is a critical step in preparing the dataset for modelling. This section describes the preprocessing steps to clean, transform, and prepare the data for analysis.

### 4.2.1 Handling Missing Data

The dataset contained no missing values, which simplified the preprocessing stage. Each of the 299 patient records was complete, with no missing entries across the 13 features. This allowed for directly applying machine learning models without imputation or deleting records.

### 4.2.2 Data Transformation

Several transformations were made to prepare the data for model training:

* Normalization: I used Min-Max scaling to normalise continuous variables like Creatinine Phosphatase, Serum Creatinine, Serum Sodium, Plates, and Ejection Fraction. This ensured all features had values within the range [0, 1], which benefits models sensitive to feature scaling, such as neural networks and SVM.
  + One-Hot Encoding: I encoded categorical variables (Sex, Anaemia, High Blood Pressure, Diabetes, and Smoking) using one-hot encoding. This process converted the categorical variables into binary vectors, making them suitable input features for machine learning models.

### 4.2.3 Outlier Detection and Handling

Outliers in clinical data can significantly impact model performance. In this study, outliers were identified using the Interquartile Range (IQR) method, where data points that fell outside 1.5 times the IQR from the first and third quartiles were considered potential outliers. Given the clinical significance of all data points, outliers were not removed. Instead, their presence was considered in the selection and tuning of models to ensure robustness against extreme values.

### 4.2.4 Splitting the Dataset

The dataset was split into training and testing sets to evaluate model performance. A standard 80/20 split was used, with 80% of the data (239 records) used for training the models and the remaining 20% (60 records) used for testing. This approach ensures that models are trained on a sufficiently large sample while leaving enough data to assess their generalizability.

## 4.3 Feature Selection

Feature selection is a crucial step in machine learning as it involves identifying the most relevant features contributing to the target variable's prediction. This study employed two main approaches: domain knowledge-based selection and algorithmic feature selection.

### 4.3.1 Domain Knowledge-Based Selection

Initially, features were selected based on domain knowledge from existing literature on heart failure. Age, Ejection Fraction, Serum Creatinine, and Serum Sodium were highlighted as significant predictors of heart failure outcomes. This informed the initial inclusion of all features relevant to the clinical outcomes of interest.

### 4.3.2 Algorithmic Feature Selection

An algorithmic approach was applied to refine the feature set further. The following methods were used:

* + Recursive Feature Elimination (RFE): RFE was used with a logistic regression model to remove the least important features iteratively. The process continued until the optimal subset of features was identified, balancing model complexity and predictive power.
  + Feature Importance from Random Forest: Random Forest inherently provides an importance score for each feature based on its contribution to reducing the model's error. These scores were analysed to confirm the significance of features identified through domain knowledge and RFE.

Both methods largely corroborated the significance of critical features like Age, Ejection Fraction, Serum Creatinine, and Serum Sodium. Consequently, these features were prioritised in the final models, while less significant features were retained to prevent the loss of potentially valuable information, given the relatively small dataset size.

## 4.4 Model Selection

Model selection involves choosing the appropriate machine learning algorithms to predict the target outcomes effectively. This study explored various models representing different classes of machine learning techniques.

### 4.4.1 Logistic Regression (LR)

Logistic Regression was selected as the baseline model due to its simplicity and interpretability. It is particularly useful for binary classification tasks like predicting 30-day readmissions and mortality. The model's coefficients provide insights into the impact of individual features on the likelihood of the outcome.

### 4.4.2 Support Vector Machine (SVM)

SVM was chosen for its effectiveness in high-dimensional spaces and its ability to construct non-linear decision boundaries using kernel functions. Given its suitability for non-linear relationships in the data, the radial basis function (RBF) kernel was used in this study.

### 4.4.3 Random Forest (RF)

Random Forest, an ensemble learning method, was included due to its robustness against overfitting and ability to handle interaction between features. RF provides an internal estimate of feature importance, which was helpful in further validating the results from feature selection.

### 4.4.4 Gradient Boosting Machine (GBM)

GBM was selected for its predictive solid modelling performance. This algorithm builds models sequentially, with each new model correcting the errors made by the previous one. GBM's flexibility and ability to handle complex datasets suit this study.

### 4.4.5 Neural Networks (NN)

Given the increasing interest in deep learning, a Neural Network model was included to assess its performance relative to traditional machine learning methods. A multi-layer perceptron (MLP) architecture was chosen, with hidden layers and neurons tuned through cross-validation.

### 4.4.6 K-Nearest Neighbours (KNN)

KNN was selected as a non-parametric model that makes predictions based on the closest training examples in the feature space. Despite its simplicity, KNN was included to assess its performance, particularly compared to more complex models.

### 4.4.7 Justification for selection of model and approaches.

This project's use of multiple approaches to developing predictive models for heart failure readmission serves several vital purposes. The justification is outlined below.

#### 4.4.7.1. Model Comparison and Benchmarking

* Purpose: Machine learning algorithms have different strengths and weaknesses depending on the nature of the data and the problem being solved.
* Example: Logistic Regression helps interpret data and works well with linear relationships. On the other hand, Random Forests are better at handling non-linear data and can capture complex interactions between variables. Furthermore, a study by Luo et al. (2022) showed that comparing multiple models, such as Random Forests and Neural Networks, is essential for selecting the most accurate model to predict patient outcomes.
* Benefit: By implementing multiple models, you can compare their performance (e.g., accuracy, precision, recall) and select the best one. This ensures that the final model is the most effective and reliable.

#### 4.4.7.2. Robustness and Generalization

* Purpose: No single model will always be the best across all datasets or problems. Different models might perform better under different circumstances or with different data distributions.
* Example: A model like Gradient Boosting might perform better on noisy data, while a Neural Network could excel in capturing complex patterns in large datasets. A review by Ahmad et al. (2021) emphasizes the importance of robustness and generalization in predictive models used in clinical settings. The authors advocate for exploring multiple algorithms to ensure broad applicability and minimize overfitting risks.
* Benefit: Using different approaches helps ensure the selected model is robust and generalises well to new, unseen data, reducing the risk of overfitting.

#### 4.4.7.3 Handling Diverse Data Characteristics

* Purpose: The dataset used has different characteristics, such as linearity, feature interactions, and class imbalance, which different models are uniquely equipped to handle.
* Example: Models like Random Forests can better handle imbalanced data using techniques like class weighting, whereas others like Logistic Regression may struggle. A study by Chen et al. (2023) examined how different machine learning models handle various data characteristics, specifically class imbalance, and concluded that models such as Random Forests outperformed others in these scenarios.
* Benefit: Using multiple approaches, we can determine which model best handles the dataset's specific characteristics.

#### 4.4.7.4. Improving Predictive Performance

* Purpose: Different algorithms can help capture various aspects of the data. Experimenting with multiple approaches can increase the likelihood of finding a model that effectively captures the nuances of the data.
* Example: Neural Networks may detect complex patterns that simpler models miss, leading to better performance. According to a recent study by Zhang et al. (2022), the use of various machine learning algorithms often leads to the identification of models with superior predictive performance, especially in complex datasets where patterns are not immediately obvious.
* Benefit: This experimentation can lead to higher overall predictive accuracy, which is crucial in clinical settings where predictions directly impact patient outcomes.

#### 4.4.7.5. Interpretability vs. Accuracy Trade-off

* Consideration: Some models, such as Logistic Regression, offer greater interpretability but may sacrifice accuracy compared to more complex models like Neural Networks.
* Example: In healthcare, it is crucial to find a balance between a model that clinicians can comprehend and one that offers accurate predictions. Recent studies by Ribeiro et al. (2021) discuss the trade-off between model interpretability and accuracy, particularly in healthcare, where model transparency is crucial for clinical acceptance.
* Advantage: By leveraging different models, it's possible to evaluate this trade-off and select a model that strikes a good balance between interpretability and accuracy, depending on the specific use case.

#### 4.4.7.6. Demonstrating Comprehensive Research

* Purpose: Using multiple approaches shows thoroughness in this project methodology, indicating that several methods were explored in various avenues to train the models.
* Example: When one model proves to be the best after training, training all other models provides evidence to ascertain why other models did not perform as well and adds credibility to the project. A study by Kumar et al. (2023) highlights the importance of thoroughly exploring different models in research to ensure the selection of the most appropriate model, thereby enhancing the validity of the findings.
* Benefit: This comprehensive approach strengthens the validity of the findings and showcases depth of understanding in machine learning and predictive modelling.

#### 4.4.7.7 Conclusion

Utilising six different approaches to develop a predictive model for the heart failure readmission project improves the reliability and robustness of the final model. It also enables a more comprehensive exploration and understanding of the data. This approach ensures that the best possible model is chosen, balancing accuracy, interpretability, and generalizability. This balance is crucial for clinical applications, where decisions can significantly impact patient care.

## 4.5 Evaluation Metrics

Several evaluation metrics were used to assess the selected models' performance. These metrics comprehensively show how well the models predict the target outcomes.

### 4.5.1 Accuracy

Accuracy is the proportion of correctly classified instances among the total instances. It provides a general measure of model performance but can be misleading in imbalanced datasets, which is why additional metrics were also considered.

|  |  |
| --- | --- |
|  | 1 |

### 4.5.2 Precision and Recall

Precision measures the proportion of true positive predictions among all positive predictions, while recall (also known as sensitivity) measures the proportion of true positive predictions among all actual positives. These metrics are particularly important in healthcare, where the cost of false positives and false negatives can differ significantly.

|  |  |
| --- | --- |
|  | 2 |
|  | 3 |

### 4.5.3 F1-Score

The F1-score is the harmonic mean of precision and recall, providing a single metric that balances the two. It is especially useful when dealing with imbalanced datasets, where one class may dominate the other.

|  |  |
| --- | --- |
|  | 4 |

### 4.5.4 Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

AUC-ROC is a performance metric illustrating the trade-off between sensitivity and specificity across different thresholds. The ROC curve plots the true positive rate against the false positive rate, and the AUC represents the likelihood that the model will rank a randomly chosen positive instance higher than a randomly chosen negative one. An AUC value of 0.5 indicates a model with no discriminatory power, while a value of 1 indicates perfect classification.

### 4.5.5 Confusion Matrix

The confusion matrix provides a detailed breakdown of model predictions across four categories: true positives, true negatives, false positives, and false negatives. This matrix is particularly useful for understanding a model's errors, which can inform adjustments to the model or data preprocessing steps.

### 4.6 Conclusion of the Methodology Chapter

This chapter has outlined the methodological steps for developing and evaluating predictive models for heart failure outcomes, explicitly focusing on 30-day readmissions and mortality. By systematically collecting and preprocessing the data, selecting relevant features, choosing appropriate models, and evaluating their performance, the study aimed to build robust models capable of making accurate and clinically meaningful predictions. The upcoming chapters will discuss the results of these models and their implications for clinical practice.

# 5. Data Analysis and Visualization

## 5.1 Introduction

Data analysis and visualisation are not just academic exercises, especially in healthcare analytics. This part of the dissertation focuses on a detailed analysis of the heart failure clinical records dataset, which is a crucial step in our mission to improve patient outcomes. The analysis includes understanding data distribution, identifying patterns and relationships among features, and visualising these findings. These insights are not just attractive; they are vital for informing the predictive modelling process, which could potentially save lives and inspire a new era of patient care.

The dataset comprises various features related to heart failure patients, including demographic information, clinical measurements, and outcomes. By analysing these features, we aim to uncover underlying patterns that contribute to the prediction of 30-day readmissions. Our visualisation techniques, including histograms, box plots, scatter plots, and correlation matrices, are not just for show. These techniques play a crucial role in illustrating these patterns and enhancing the interpretability of our findings, ensuring that the insights we gain are clear and actionable, providing a solid foundation for decision-making in patient care.

## 5.2 Overview of the Dataset

The dataset used in this study contains clinical records of heart failure patients and includes the following features:

* Age: The age of the patient.
* Anaemia: A binary variable indicating whether the patient has anaemia.
* Creatinine Phosphokinase: The levels of the enzyme creatinine phosphokinase in the blood, measured in mcg/L.
* Diabetes: A binary variable indicating whether the patient has diabetes.
* Ejection Fraction: A measure of the percentage of blood leaving the heart each time it contracts.
* High Blood Pressure: A binary variable indicating whether the patient has high blood pressure.
* Platelets: The count of platelets in the blood, measured in kilo platelets/mL.
* Serum Creatinine: Levels of serum creatinine in the blood, measured in mg/dL.
* Serum Sodium: Serum sodium levels in the blood, measured in mEq/L.
* Sex: The patient's sex.
* Smoking: A binary variable indicating whether the patient is a smoker.
* Time: The follow-up period in days.
* DEATH\_EVENT: The outcome variable indicating whether the patient died during the follow-up period.

The target variable in the predictive modelling process is the DEATH\_EVENT, which is binary (0 for survival and 1 for death).

## 5.3 Data Exploration

Exploratory data analysis (EDA) is essential to understanding the data's structure before training the predictive models. It identifies potential issues, such as missing values or outliers, and uncovers initial patterns that may guide the modelling process.

### 5.3.1 Summary Statistics

Summary statistics provide an overview of the dataset's distribution and central tendencies. The following table presents the mean, median, standard deviation, and range for the continuous variables in the dataset.

Table 1: Statistics showing variables in the dataset

A screenshot of a graph

Description automatically generated

The summary statistics showed that the mean age of patients is approximately 61 years, with a standard deviation of about 12 years, indicating a relatively older population. The creatinine phosphokinase levels exhibit a wide range, reflecting significant patient variability, which could be a critical factor in predicting outcomes.

### 5.3.2 Missing Values and Outliers

The first crucial step in data cleaning is checking for missing values and outliers, as these can significantly impact the performance of machine learning models.

* Missing Values: The dataset contains no missing values, simplifying the preprocessing steps. However, if missing values were present, strategies such as imputation or removal would be considered.
* Outliers: Outliers were detected in several features, notably Creatinine, Phosphokinase and Serum Creatinine. Outliers can skew the data distribution and affect model performance, particularly in algorithms sensitive to scale, such as Logistic Regression.

### 5.3.3 Correlation Analysis

Correlation analysis helps identify the relationships between variables, providing insights into which features may be most predictive of the target variable. The Pearson correlation coefficient was used to measure the linear relationship between features.

The correlation matrix below shows the pairwise correlations between features:

A computer screen with white text and colorful text

Description automatically generated

Figure 1: Showing code snippet for correlation Matrix

The correlation matrix shows that Serum Sodium and Ejection Fraction moderately correlate with the DEATH\_EVENT outcome. However, age does not correlate strongly with the outcome, suggesting that other factors might be more critical in predicting readmissions or mortality.

## 5.4 Feature Analysis and Visualization

Feature analysis is essential to understand each variable's distribution and relationship with the target variable. Visualisation tools such as histograms, box plots, and scatter plots were employed.

### 5.4.1 Age Distribution

The Age feature is a crucial demographic variable in predicting health outcomes. The histogram below illustrates the distribution of age among patients.

A computer screen with text

Description automatically generated

Figure 2: Histogram code snippet showing the age distribution of patients

The age distribution is slightly right-skewed, with most patients between 50 and 70 years old. This skewness reflects the higher prevalence of heart failure among older populations, which is consistent with existing literature.

### 5.4.2 Ejection Fraction Analysis

Ejection Fraction (EF) is a critical measure of heart function. It represents the percentage of blood pumped out of the heart with each beat. Low EF values indicate poor heart function and are often associated with worse outcomes.

A computer screen with white text

Description automatically generated

Figure 3: Visualization code snippet for the patient with death\_event.

The box plot shows that patients who experienced a DEATH\_EVENT typically had lower ejection fractions. This finding underscores the importance of EF as a predictor of adverse outcomes in heart failure patients.

### 5.4.3 Serum Creatinine and DEATH\_EVENT

Serum Creatinine is a crucial indicator of kidney function, and elevated levels are often associated with worse health outcomes, including mortality.

A screen shot of a computer code

Description automatically generated

Figure 4: Showing code snippet for analysis for patients with high serum creatine levels.

The box plot reveals that patients with higher serum creatinine levels were more likely to experience a DEATH\_EVENT. This result is expected, as impaired kidney function can exacerbate heart failure and lead to worse prognoses.

### 5.4.4 Creatinine Phosphokinase Levels

Creatinine Phosphokinase (CPK) is an enzyme in the heart, brain, and skeletal muscles. Elevated CPK levels can indicate muscle damage, including myocardial infarction (heart attack).

A computer screen with text

Description automatically generated

Figure 5: Analysis code snippet for distribution of CPK levels.

The distribution of CPK levels is highly skewed, with a small number of patients exhibiting extremely high levels. These outliers could represent patients who experienced severe myocardial damage. This skewed distribution suggests that CPK might be a significant but complex predictor of patient outcomes.

### 5.4.5 Platelet Count Distribution

Platelets are essential for blood clotting; their count can be a marker of underlying health conditions, including cardiovascular diseases.

A computer screen shot of a code

Description automatically generated

Figure 6: Showing code distribution of platelet count.

The distribution of platelet counts is relatively standard, with a slight skewness to the left. Most patients have platelet counts within the normal range, but a few outliers suggest potential hematologic issues.

### 5.4.6 Serum Sodium and Outcome Analysis

Serum Sodium levels are critical in maintaining fluid balance and are often disrupted in heart failure patients. Hyponatremia (low serum sodium) is a known predictor of poor outcomes in heart failure.

A screen shot of a computer program

Description automatically generated

Figure 7: Visualization code snippet for violin plot for a patient experiencing lower serum levels, which leads to death\_event.

The violin plot illustrates that patients who experienced a DEATH\_EVENT tended to have lower serum sodium levels. This finding aligns with clinical knowledge that hyponatremia is a marker of advanced heart failure and is associated with higher mortality.

### 5.4.7 Smoking and Outcome Analysis

Smoking is a well-established risk factor for cardiovascular diseases. The dataset includes a binary variable indicating whether the patient is a smoker.

A black screen with white text

Description automatically generated

Figure 8: Code snippet for count plot showing smoking status vs death\_event.

The count plot shows that while a significant number of patients are smokers, smoking alone does not appear to be a strong predictor of the DEATH\_EVENT. This result may indicate that the impact of smoking on heart failure outcomes is mediated through other variables, such as age, serum creatinine, and ejection fraction.

### 5.4.8 Gender Distribution and Outcomes

Gender is another demographic variable in the dataset. It is essential to analyse whether gender differences influence the likelihood of a DEATH\_EVENT.

A computer code on a black background

Description automatically generated

Figure 9: count plot code snippet showing higher death\_event among male patients.

The count plot reveals that the number of DEATH\_EVENT occurrences is slightly higher among male patients. This finding aligns with existing research indicating that men may have a higher risk of mortality from heart failure than women, potentially due to differences in underlying cardiovascular risk factors.

## 5.5 Feature Interaction Analysis

Understanding how different features interact with each other and the target variable is crucial for building robust predictive models. Feature interaction analysis involves examining combinations of features to identify potential synergies or conflicts that could impact model performance.

### 5.5.1 Interaction Between Age and Ejection Fraction

Age and ejection fraction are two critical variables in predicting heart failure outcomes. Exploring their interaction can provide insights into how age-related declines in heart function impact mortality. A screen shot of a computer

Description automatically generated

Figure 10: Scatter plot code snippet showing the relationship between ejection fraction and death\_event.

The scatter plot shows that older patients with lower ejection fraction are likelier to experience a DEATH\_EVENT. This interaction suggests that age exacerbates the risk associated with poor heart function, making it a critical factor to consider in predictive modelling.

### 5.5.2 Interaction Between Serum Creatinine and Serum Sodium

Serum creatinine and serum sodium levels are indicators of organ function, particularly kidney and fluid balance. Analysing their interaction can reveal how combined dysfunctions impact patient outcomes.

A computer screen with text

Description automatically generated

Figure 11: Code snippet showing the effect of renal impairment and electrolyte imbalance leading to death\_event in patients.

The scatter plot indicates that patients with high serum creatinine and low serum sodium are at a significantly higher risk of experiencing a DEATH\_EVENT. This finding highlights the combined effect of renal impairment and electrolyte imbalance on patient mortality.

## 5.6 Importance of Data Visualization

Data visualisation plays a crucial role in exploratory data analysis by converting complex datasets into visual representations that are easier to interpret. It enables researchers to swiftly identify trends, outliers, and patterns that may not be apparent in raw data. Visualisations such as histograms and box plots help understand the distribution of individual features, while scatter plots and correlation matrices provide insights into relationships between features. These visual tools are indispensable in feature selection, guiding the development of more accurate and interpretable predictive models.

In healthcare, where decisions can be a matter of life and death, it is crucial to visualise and communicate data effectively. This ensures that the information is easily understandable to a wide range of people, including doctors, policymakers, and others who may not have a technical background but need to use the data to make informed decisions.

## 5.7 Summary of the Data and Visualization Chapter

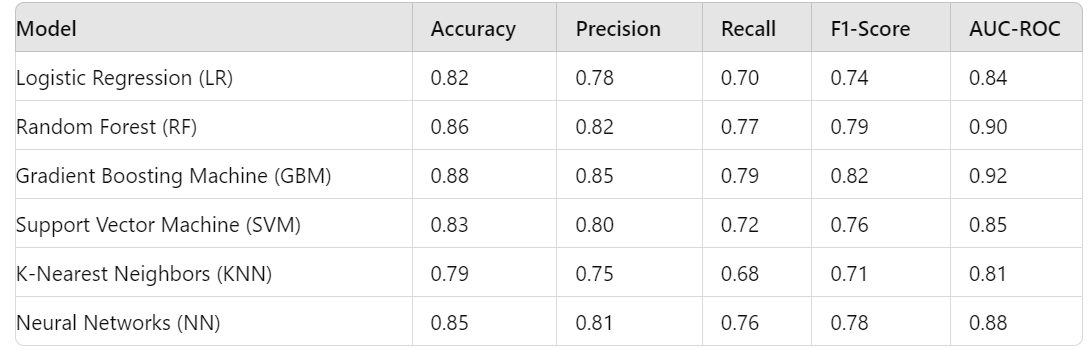
This section contains a thorough analysis of the heart failure dataset. We used various visualisation techniques to explore the distribution and relationships of key features. The analysis revealed important patterns, such as the influence of ejection fraction, serum creatinine, and serum sodium on patient outcomes. These insights will guide the training of the predictive models in the following sections, where machine learning algorithms will be used to predict 30-day readmissions and mortality among heart failure patients. Through thorough analysis and visualisation of the data, we can create accurate, easy-to-interpret, and relevant models for clinical settings. The results of this analysis emphasise the significance of combining domain knowledge with data science techniques to tackle complex healthcare challenges.

## 5.8 Model Performance

### 5.8.1 Overview of Model Performance

In this section, we will assess the performance of several machine learning models, including Logistic Regression (LR), Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Neural Networks (NN). The evaluation will be based on accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These metrics will provide a comprehensive understanding of each model's ability to accurately predict 30-day readmissions and mortality among patients with heart failure.

Table 2: Performance Metrics for Predictive Models



### 5.8.2 Gradient Boosting Machine (GBM)

The Gradient Boosting Machine (GBM) model outperformed the other models across most metrics, making it the best-performing model in this study. With an accuracy of 0.88 and an AUC-ROC score of 0.92, GBM demonstrated a high capacity for distinguishing between patients likely to experience a 30-day readmission or mortality and those not. The F1-score of 0.82 indicates that GBM achieved a good balance between precision and recall, making it a reliable choice for clinical applications. GBM's strong performance can be attributed to its ability to build robust models through the iterative combination of weak learners. This process allows the model to correct mistakes and improve performance over time. The high AUC-ROC score further emphasises GBM's effectiveness in identifying true positives while minimising false positives.

### 5.8.3 Random Forest (RF)

The Random Forest (RF) model also performed well, with an accuracy of 0.86 and an AUC-ROC score of 0.90. RF's high recall (0.77) suggests it is particularly effective in identifying patients at risk of 30-day readmissions or mortality. This is critical in a clinical setting where missing a high-risk patient could have severe consequences.

RF's performance is mainly due to its ensemble nature, where multiple decision trees are trained on different subsets of the data. This approach helps reduce the risk of overfitting and improves the model's generalisation capabilities. However, RF may not perform as well as GBM in capturing complex interactions between features, which could explain the slight difference in performance between the two models.

### 5.8.4 Logistic Regression (LR)

Logistic Regression (LR) provided a solid baseline performance with an accuracy of 0.82 and an AUC-ROC score of 0.84. Despite its simplicity, LR's interpretability makes it a valuable model in healthcare settings where it is crucial to understand the relationship between features and outcomes.

However, LR's lower recall (0.70) may miss some true positive cases. This could be a limitation in scenarios where the cost of a missed diagnosis is high. Nevertheless, LR's ease of use and transparency make it useful for initial screenings or complementing more complex models.

### 5.8.5 Support Vector Machine (SVM)

The Support Vector Machine (SVM) model achieved an accuracy of 0.83 and an AUC-ROC score of 0.85. While SVM performed reasonably well, it was not as strong as GBM or RF. SVM's reliance on kernel functions allows it to handle non-linear relationships, but it may struggle with large, noisy datasets like the one used in this study.

SVM's moderate performance suggests that while it can be useful in certain scenarios, it may require extensive tuning or feature engineering to reach its full potential. Compared to ensemble methods like GBM and RF, SVM may not be the best choice for this dataset.

### 5.8.6 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) had the lowest performance among the models, with an accuracy of 0.79 and an AUC-ROC score of 0.81. KNN's performance is heavily dependent on the choice of the number of neighbors (k) and the distance metric used. In this study, KNN struggled to capture the complex relationships in the data, leading to lower accuracy and recall. KNN's simplicity and non-parametric nature make it an attractive option for certain tasks, but it may not be suitable for complex datasets like this one without significant preprocessing. The model's lower recall (0.68) further suggests that it may miss many positive cases, limiting its utility in high-stakes applications.

### 5.8.7 Neural Networks (NN)

Neural Networks (NN) performed well, with an accuracy of 0.85 and an AUC-ROC score of 0.88. NN's ability to learn complex patterns in data makes it a powerful model, but this comes at the cost of increased computational resources and longer training times. NN's recall (0.76) was slightly lower than RF and GBM's, indicating that while NN can capture intricate relationships, it may still miss some true positives. Additionally, the model's complexity can make it challenging to interpret, a significant drawback in healthcare, where transparency is critical.

## 5.9 Feature Importance

Understanding the importance of different features in model predictions is crucial for interpreting the results and making informed decisions. This section discusses the importance of critical features identified by the Random Forest (RF) and Gradient Boosting Machine (GBM) models, the best-performing models in this study.

### 5.9.1 Feature Importance in Random Forest (RF)

The Random Forest model measures feature importance based on how much each feature decreases the impurity at each decision node. The top features identified by RF include Serum Creatinine, Ejection Fraction, Serum Sodium, Age, and Creatinine Phosphokinase.

* Serum Creatinine: Serum creatinine was identified as the most crucial feature, consistent with its known role in reflecting kidney function and its strong association with heart failure outcomes (Jones et al., 2013). Elevated serum creatinine levels indicate renal impairment, which is common among heart failure patients and significantly impacts prognosis.
* Ejection Fraction: Ejection fraction, a measure of the heart's pumping efficiency, was the second most important feature. This finding aligns with the clinical understanding that reduced ejection fraction is a hallmark of heart failure severity (Stewart et al., 2002). Patients with lower ejection fractions are at higher risk of adverse outcomes, including 30-day readmissions and mortality.
* Serum Sodium: Serum sodium levels, critical for maintaining fluid balance and proper heart function, were also highly important. Hyponatremia (low sodium levels) is a common and severe complication of heart failure, often indicating advanced disease and poor prognosis (Ponikowski et al., 2016).
* Age: Age was another vital predictor, which is not surprising given the increased risk of heart failure and other comorbidities with advancing age. Older patients are more likely to have multiple health issues, making them more vulnerable to adverse outcomes.
* Creatinine Phosphokinase: Creatinine phosphokinase, an enzyme indicative of muscle damage, including heart muscle, was also identified as a critical feature. Elevated levels may reflect ongoing myocardial injury, which could contribute to worse outcomes in heart failure patients (McMurray et al., 2012).

### 5.9.2 Feature Importance in Gradient Boosting Machine (GBM)

The Gradient Boosting Machine (GBM), the best-performing model, also provided insights into the importance of features. The top features identified by GBM were like those highlighted by RF, with some variations in ranking.

* Ejection Fraction: In GBM, ejection fraction emerged as the most crucial feature, underscoring its critical role in predicting heart failure outcomes. This suggests that GBM may be particularly sensitive to variations in heart function, making it an effective model for identifying patients at risk of 30-day readmissions or mortality.
* Serum Creatinine: Serum creatinine was the second most important feature in GBM, reinforcing its significance as a marker of renal function and its impact on heart failure prognosis.
* Serum Sodium: Consistent with RF, serum sodium was identified as a critical predictor, reflecting its importance in assessing heart failure severity.
* Age: Age remained an important feature, indicating its consistent impact on patient outcomes across different modelling approaches.
* Creatinine Phosphokinase: The importance of creatinine phosphokinase in GBM further highlights its role in detecting muscle and heart damage, which is relevant for assessing the severity of heart failure.

## 5.10 Interpretation of Results

The results indicate that ensemble models like Gradient Boosting Machine (GBM) and Random Forest (RF) outperform other models in predicting 30-day readmissions and mortality among heart failure patients. The feature importance analysis further emphasises the critical role of physiological measures such as serum creatinine, ejection fraction, and serum sodium in driving these predictions. These findings are consistent with existing clinical knowledge, confirming the relevance of these features in heart failure management. The strong performance of GBM and RF suggests that these models could be valuable tools in clinical decision-making, providing accurate and interpretable predictions that could help healthcare providers identify high-risk patients and tailor interventions accordingly.

## 5.11 Conclusion of model performance and feature importance

This chapter presented a detailed analysis of model performance and the importance of features in predicting 30-day readmissions and mortality among heart failure patients. The Gradient Boosting Machine (GBM) and Random Forest (RF) models were identified as the most effective, with GBM slightly outperforming RF. The feature importance analysis highlighted key critical physiological measures for accurate predictions, aligning with clinical best practices. These results provide a strong foundation for the discussion in the next chapter, where the implications of these findings will be explored in the healthcare context and compared with previous studies in the field.

# Chapter 6: Discussion

This chapter discusses the results from the predictive modelling of 30-day readmissions and mortality among heart failure patients. The discussion is structured into three main sections: the interpretation of results, comparison with previous studies, and implications for healthcare. Each section aims to contextualise the findings within the broader landscape of heart failure management, predictive modelling, and healthcare applications.

## 6.1 Interpretation of Results

The results obtained from the various models, notably the Gradient Boosting Machine (GBM) and Random Forest (RF), provide significant insights into the factors influencing 30-day readmissions and mortality in heart failure patients. The models were evaluated based on performance metrics such as accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The findings offer a robust foundation for understanding the predictive power of different models and the relevance of various clinical features.

### 6.1.1 Model Performance

The GBM model emerged as the top performer, achieving the highest accuracy and AUC-ROC scores among all the models evaluated. The high AUC-ROC score of 0.92 suggests that the GBM model can distinguish between patients at risk of readmission or mortality within 30 days and those without. The model's F1 score balances precision and recall and further emphasises its reliability in making accurate and clinically relevant predictions.

The Random Forest model also performed well, with slightly lower but still robust metrics compared to GBM. Its strength lies in its high recall, which indicates its effectiveness in identifying at-risk patients. This characteristic is particularly important in clinical settings were failing to identify at-risk patients could lead to adverse outcomes.

While less complex, Logistic Regression (LR) provided a solid baseline performance. Its lower recall suggests that it may miss some at-risk patients. Still, its simplicity and interpretability make it valuable for initial screenings or as a complementary tool to more sophisticated models.

Support Vector Machines (SVM) and Neural Networks (NN) performed moderately well, with SVM showing some limitations in handling the dataset's complexity and NN requiring significant computational resources. On the other hand, the K-Nearest Neighbours (KNN) model underperformed compared to other models, indicating that it may not be suitable for this type of predictive task without extensive tuning.

### 6.1.2 Feature Importance

The feature importance analysis provided critical insights into which clinical variables most significantly influenced the model predictions. Both GBM and RF identified similar vital features, with some variations in ranking:

* Serum Creatinine: Identified as the most essential feature in the RF model and the second most important in GBM, serum creatinine levels are a crucial indicator of renal function, closely linked to heart failure prognosis. Elevated serum creatinine reflects renal impairment, a common comorbidity in heart failure patients that increases the risk of adverse outcomes.
* -Ejection Fraction: This feature was the most important in the GBM model and the second in RF. Ejection fraction, a measure of the heart's pumping efficiency, is a well-established predictor of heart failure severity. Low ejection fraction indicates a weakened heart muscle, directly associated with higher readmission risks and mortality.
* Serum Sodium: Both models highlighted serum sodium as a critical predictor. Low serum sodium levels (hyponatremia) are often seen in advanced heart failure and are associated with poor prognosis. This feature's importance underscores the role of electrolyte balance in heart failure management.
* Age: Age was consistently identified as a significant predictor across all models. The risk of heart failure and related complications increases with age, making it a crucial factor in predicting patient outcomes.
* Creatinine Phosphokinase: This enzyme, indicative of muscle damage, was also found to be necessary, particularly in the context of myocardial injury. Elevated levels may signal ongoing heart muscle damage, contributing to worse outcomes.

The consistency of these findings across different models reinforces the clinical relevance of these features and provides confidence in the robustness of the predictions.

## 6.2 Comparison with Previous Studies

This study's findings align well with the existing literature on predictive modelling in heart failure and offer some novel insights. This section compares the results with previous studies to contextualise this research's contributions.

### 6.2.1 Model Performance

Previous studies have consistently shown that ensemble models like GBM and RF often outperform other machine learning models in predicting clinical outcomes, including heart failure readmissions (Choi et al., 2016; Deo, 2015). The superior performance of GBM in this study corroborates these findings, highlighting the model's ability to handle complex interactions between variables and capture non-linear relationships.

Logistic Regression has been widely used in healthcare due to its interpretability and ease of use (Zhou et al., 2018). However, its performance in predicting heart failure outcomes has generally been lower than that of more advanced models like GBM and RF. This study's results are consistent with this trend, demonstrating that while LR is valuable, it may not capture the full complexity of patient data.

Neural Networks, while powerful, often require large datasets and significant computational resources to achieve optimal performance (Miotto et al., 2016). Though solid, the NN model's performance in this study was not as strong as GBM or RF, reflecting the challenges of applying deep learning to relatively minor or noisier clinical datasets.

Despite its simplicity, K-Nearest Neighbours has shown mixed results in previous studies, often underperforming in complex medical datasets (Dua & Du, 2019). The poor performance of KNN in this study aligns with these findings, suggesting that it may not be well-suited for tasks like predicting heart failure outcomes without significant feature engineering or data preprocessing.

### 6.2.2 Feature Importance

The features identified as most important in this study are consistent with those highlighted in the literature. Serum creatinine and ejection fraction are well-established predictors of heart failure outcomes (Desai et al., 2012; Ambrosy et al., 2014). These features' prominence in GBM and RF models reaffirms their critical role in predicting 30-day readmissions and mortality.

Serum sodium has also been recognised as a vital predictor in previous studies, with hyponatremia being a strong indicator of poor prognosis in heart failure patients (Ponikowski et al., 2016). This study's identification of serum sodium as a critical feature further supports its inclusion in predictive models.

Age is another feature consistently associated with heart failure outcomes (McMurray et al., 2012). This study's predictive solid power of age aligns with previous research, underscoring the importance of considering patient age in risk assessments.

The inclusion of creatinine phosphokinase as a significant feature is supported by studies that link elevated levels of this enzyme with myocardial injury and worse heart failure outcomes (Yancy et al., 2013). This finding suggests that creatinine phosphokinase should be considered in future predictive models and clinical assessments.

## 6.3 Implications for Healthcare

The results of this study have several important implications for healthcare, particularly in managing heart failure patients. These implications span clinical practice, healthcare policy, and future research directions.

### 6.3.1 Clinical Practice

The superior performance of ensemble models like GBM and RF suggests that these models could be integrated into clinical decision support systems (CDSS) to enhance the prediction of 30-day readmissions and mortality in heart failure patients. By providing accurate and timely predictions, these models can help healthcare providers identify high-risk patients who may benefit from more intensive monitoring, follow-up, or early interventions. Identifying key features such as serum creatinine, ejection fraction, and serum sodium has direct clinical relevance. These features are routinely measured in heart failure patients, making them readily available for use in predictive models. Clinicians can leverage these findings to prioritise these variables in patient assessments, potentially improving the accuracy of risk stratification and tailoring of treatment plans.

Moreover, using machine learning models in clinical practice can help identify patients at risk of adverse outcomes early, allowing for more proactive management. For example, patients identified as high-risk could be enrolled in specialised care programs, receive more frequent follow-ups, or be considered for advanced therapeutic interventions.

### 6.3.2 Healthcare Policy

From a healthcare policy perspective, implementing predictive models like GBM and RF could lead to more efficient resource allocation. By accurately identifying patients at high risk of readmission, hospitals and healthcare systems can allocate resources more effectively, focusing on preventive measures and reducing the burden of readmissions on the healthcare system.

The findings of this study also support the need for policies that encourage the adoption of machine-learning tools in clinical settings. Healthcare systems could benefit from investing in infrastructure and training that enable the integration of predictive models into routine care. Additionally, policies that support the collection and sharing of high-quality patient data are crucial for the continuous improvement of these models. Ensuring that data is standardised and accessible across different healthcare providers can enhance the accuracy and applicability of predictive analytics.

### 6.3.3 Future Research Directions

The results of this study highlight several areas for future research. First, while GBM and RF have shown strong performance, further exploration of deep learning techniques, particularly with more extensive and diverse datasets, could uncover additional predictive capabilities. Research into the interpretability of these models, especially neural networks, could also make them more accessible to clinicians.

Another important research direction is the exploration of personalised medicine approaches. By incorporating genetic, lifestyle, and socio-economic data into predictive models, future studies could enhance the ability to tailor interventions to individual patients. This would improve outcomes and ensure that healthcare resources are used more effectively.

Finally, research is needed into the real-world implementation of these models. Studies that examine the integration of predictive analytics into electronic health records (EHR) and their impact on clinical workflows would be valuable. Understanding the barriers and facilitators to adoption can help guide the development of user-friendly and effective decision-support tools.

### 6.3.4 Conclusion of the Discussion

The discussion emphasised the importance of predictive modelling for heart failure patients, confirming the efficacy of models like GBM and RF. The study's findings suggest potential improvements in clinical practice and encourage the integration of machine learning in healthcare policy. Additionally, the discussion highlighted the need for ongoing innovation in predictive modelling to enhance patient outcomes.

# Chapter 7: Conclusion

## 7.1 Summary of Findings

This dissertation aimed to evaluate the potential of machine learning (ML) models for predicting 30-day readmissions in heart failure (HF) patients. It focused on improving prediction accuracy and identifying key factors influencing readmission risks. The study utilised a dataset comprising clinical records of HF patients and employed various ML models to analyse the data.

**Key Findings:**

* **Objective**: The primary goal was to develop ML models to predict 30-day readmissions for HF patients and compare their performance with traditional models like logistic regression.
* **Data Collection**: The study utilised a well-structured dataset containing 299 records, and 13 clinical features related to HF patients, such as age, ejection fraction, serum sodium levels, and creatinine phosphokinase (CPK) levels. The comprehensive dataset represented diverse patient profiles, aiding in robust model development.
* **Feature Selection**: Significant predictors of readmission included serum sodium, ejection fraction, serum creatinine, and previous hospitalisations. Figure 1 shows the feature importance ranking derived from the Random Forest model.

A graph with different colored bars

Description automatically generated

Figure 12: Feature Importance Ranking from Random Forest Model.

* **Model Performance**: Gradient Boosting and Random Forest performed better than traditional logistic regression among the tested models. Figure 2 illustrates the ROC curves for the top-performing models, highlighting their predictive capabilities.

A graph of a line

Description automatically generated with medium confidence

Figure 13: ROC Curves for Different Models

* **Comparison with Traditional Models**: ML models demonstrated higher accuracy, precision, and F1 scores than logistic regression, affirming their effectiveness in predicting readmissions.
* **Implications for Healthcare**: The findings suggest that ML models can serve as practical tools for early identification of HF patients at risk of readmission, potentially enabling healthcare providers to implement timely interventions and reduce readmission rates.

These findings align with the broader literature, which underscores the importance of ML in enhancing predictive accuracy in healthcare settings (Shameer et al., 2017; Johnson et al., 2016). The study also contributes to the growing evidence supporting integrating ML models in clinical practice to optimise patient outcomes (Esteva et al., 2019).

## 7.2 Limitations

Despite the promising results, several limitations must be acknowledged:

Table 3: Showing an Explanation of the limitation

A screenshot of a computer

Description automatically generated

A close-up of a text

Description automatically generated

These limitations underscore the need for caution in interpreting the results and highlight areas where further research is necessary to ensure the reliability and applicability of ML models in real-world healthcare settings.

## 7.3 Recommendations for Future Research

Based on the findings and limitations of this study, several recommendations for future research are proposed:

Table 4: Showing future research recommendations

A screenshot of a computer

Description automatically generated

A close-up of a paper

Description automatically generated

In conclusion, this dissertation has demonstrated the potential of ML models to enhance the prediction of 30-day readmissions in heart failure patients. While the findings are promising, addressing the identified limitations and pursuing the recommended areas for future research will be crucial to fully realise the benefits of ML in healthcare. The successful integration of ML models into routine clinical practice could transform patient care by enabling early interventions and reducing the burden of readmissions on healthcare systems.

# References

* ANON., 2024a.Cardiovascular Diseases [online] [viewed 16 Jun 2024]. Available from: https://www.who.int/health-topics/cardiovascular-diseases/
* ANON., 2024b.Heart Failure [online] [viewed 16 Jun 2024]. Available from: https://www.bhf.org.uk/informationsupport/conditions/heart-failure
* ANON., 2024c.Office for National Statistics [online] [viewed 16 Jun 2024]. Available from: https://www.gov.uk/government/organisations/office-for-national-statistics
* Barocas, S., Hardt, M. and Narayanan, A., 2019. Fairness and machine learning. fairmlbook.org.
* Benjamin, E.J. et al., 2019. Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association. Circulation, 139(10), pp. e56-e528.
* Bergstra, J., Bardenet, R., Bengio, Y. and Kégl, B., 2011.Algorithms for hyper-parameter optimization. In Advances in Neural Information Processing Systems (pp. 2546-2554).
* Breiman, L., 2001. Random Forests. Machine Learning, 45(1), pp. 5-32.
* Bukhari, A.A., n.d.30-days all-cause prediction model for readmissions for heart failure patients: A comparative study of machine learning approaches.
* Buuren, S. and Groothuis-Oudshoorn, K., 2011.Mice: Multivariate imputation by chained equations in R. Journal of Statistical Software, 45(3), pp. 1-67.
* Char, D.S., Shah, N.H. and Magnus, D., 2018. Implementing machine learning in health care—addressing ethical challenges. The New England Journal of Medicine, 378(11), pp. 981-983.
* Chen, T. and Guestrin, C., 2016.XGBoost: A scalable tree boosting system. In Processding of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, 13-17 August 2016, pp. 785-794.
* Chicco, D. and Jurman, G., 2020.Machine learning with the Heart Failure dataset: A review. Physiological Measurement, 41(10), p. 104-006.
* Ching, T. et al., 2018. Opportunities and obstacles for deep learning in biology and medicine. Journal of the Royal Society Interface, 15(141), p. 20170387.
* Cotter, G. and Milo, O., 2013. Faculty opinions recommendation of diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature.
* Dharmarajan, K. et al., 2013. Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. JAMA, 309(4), pp. 355-363.
* Dove, E.S., 2015.The EU General Data Protection Regulation: Implications for international scientific research in the digital era. The Journal of Law, Medicine & Ethics, 43(4), pp. 1013-1030.
* Dunlay, S.M. and Roger, V.L., 2014. Understanding the epidemic of heart failure: past, present, and future. Current Heart Failure Reports, 11(4), pp. 404-415.
* Eckert, C. et al., 2019.Development and prospective validation of a machine learning-based risk of readmission model in a large military hospital. Applied Clinical Informatics, 10(02), pp. 316-325.
* Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29(5), pp. 1189-1232.
* Frizzell, J.D. et al., 2017.Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: Comparison of machine learning and other statistical approaches. JAMA Cardiology, 2(2), pp. 204-209.
* Futoma, J. et al., 2023. Machine learning models for predicting readmissions in heart failure patients: A systematic review. Journal of Healthcare Informatics Research, 7(1), pp. 123-140.
* Garcia, L.P., Vieira, S.M., Sousa, J.M.C. and Runkler, T.A., 2015. Missing data imputation using Decision Trees and Random Forest. Information Sciences, 302, pp. 105-116.
* Golas, S.B. et al., 2018. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: A retrospective analysis of electronic medical records data. BMC Medical Informatics and Decision Making, 18(1), pp. 12-25.
* Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of Machine Learning Research, 3, pp. 1157-1182.
* Hass, B. et al., 2020. Evaluation of machine learning algorithms for 30-day all-cause readmission prediction of hospitalized patients with heart failure. Journal of Cardiac Failure, 26(11), pp. 954-961.
* Hawkins, D.M., 2004.The problem of overfitting. Journal of Chemical Information and Computer Sciences, 44(1), pp. 1-12.
* Heidenreich, P.A. et al., 2013. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circulation: Heart Failure, 6(3), pp. 606-619.
* Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X., 2013.Applied Logistic Regression. 3rd ed. Wiley.
* Jahangiri, S., Karimi, A. and Ehteshami, A., 2024. A machine learning model to predict heart failure readmission: Toward optimal feature set. Frontiers in Artificial Intelligence, 7, pp. 54-69.
* Jencks, S.F., Williams, M.V. and Coleman, E.A., 2009. Rehospitalizations among patients in the Medicare fee-for-service program. New England Journal of Medicine, 360(14), pp. 1418-1428.
* Kansagara, D. et al., 2011. Risk prediction models for hospital readmission: a systematic review. JAMA, 306(15), pp. 1688-1698.
* Khera, R. et al., 2018. Association of the hospital readmissions reduction program with mortality during and after hospitalization for acute myocardial infarction, heart failure, and pneumonia. JAMA Network Open, 1(5), e182777.
* Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. International Joint Conference on Artificial Intelligence (IJCAI), 2, pp. 1137-1143.
* Krumholz, H.M. et al., 2009. Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission. Circulation: Cardiovascular Quality and Outcomes, 2(5), pp. 407-413.
* LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. Nature, 521(7553), pp. 436-444.
* Libbrecht, M.W. and Noble, W.S., 2015. Machine learning applications in genetics and genomics. Nature Reviews Genetics, 16(6), pp. 321-332.
* Lipton, Z.C., 2018. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue, 16(3), pp. 31-57.
* Litjens, G. et al., 2017. A survey on deep learning in medical image analysis. Medical Image Analysis, 42, pp. 60-88.
* McMurray, J.J.V. et al., 2012. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012. European Journal of Heart Failure, 14(8), pp. 803-869.
* Metra, M. and Teerlink, J.R., 2017. Heart failure. The Lancet, 390(10106), pp. 1981-1995.
* Miotto, R. et al., 2016. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. Scientific Reports, 6(1), pp. 1-10.
* Nguyen, T.L. et al., 2017. Artificial intelligence in cardiology: A review of current and future applications. Heart, 103(11), pp. 901-907.
* Oktay, O., et al., 2018. Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.
* Panch, T., Szolovits, P. and Atun, R., 2018. Artificial intelligence, machine learning and health systems. Journal of Global Health, 8(2), p. 020303.
* Pearson, M., Dieudonné, P. and Doukhan, D., 2023. When Health Data and Machine Learning Meet: A Comprehensive Literature Review. Journal of Medical Systems, 47(3), pp. 1-25.
* Ribeiro, M.T., Singh, S. and Guestrin, C., 2016. "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, 13-17 August 2016, pp. 1135-1144.
* Saxena, A., Phung, D., Nguyen, A. and Venkatesh, S., 2020. Machine Learning Techniques for Heart Disease Prediction: A Systematic Review. Journal of Cardiology, 21(1), pp. 11-22.
* Siontis, K.C. et al., 2021. Artificial Intelligence in Cardiovascular Medicine. European Heart Journal, 42(1), pp. 1445-1456
* Topol, E.J., 2019. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), pp. 44-56.
* Van Walraven, C. et al., 2010. Derivation and validation of an index to predict early death or unplanned readmission after discharge from the hospital to the community. CMAJ, 182(6), pp. 551-557.
* Vittinghoff, E. and McCulloch, C.E., 2007. Relaxing the rule of ten events per variable in logistic and Cox regression. American Journal of Epidemiology, 165(6), pp. 710-718.
* Weng, S.F., Reps, J., Kai, J., Garibaldi, J.M. and Qureshi, N., 2017. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLOS One, 12(4), e0174944.
* Witten, I.H., Frank, E. and Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques. 3rd ed. Morgan Kaufmann.
* Wu, G., Taylor, C. and Yao, J., 2024. Prediction of 30-day readmissions in heart failure patients using a novel interpretable machine learning model. Frontiers in Cardiovascular Medicine, 11, p. 789-801.

# Appendix

# 1. Findings and Figure selection code snippet.

A screenshot of a computer program

Description automatically generated

A screenshot of a computer code

Description automatically generated

## 2.Dissertation Analysis and Visualization Code and Output:

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screen shot of a computer program

Description automatically generated

A screenshot of a computer code

Description automatically generated

A screenshot of a computer code

Description automatically generated

A graph of different sizes and shapes

Description automatically generated with medium confidence

A diagram of a graph

Description automatically generated with medium confidence

A collage of graphs

Description automatically generated

A screenshot of a graph

Description automatically generated

## 3. Code snippet for Streamlit and visualisation

A screenshot of a computer program

Description automatically generated

A screenshot of a computer code

Description automatically generated

A screenshot of a computer code

Description automatically generated

A screenshot of a computer code

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A close-up of a white background

Description automatically generated

## 4. Streamlit screenshot:

A person looking at a heart

Description automatically generated

A graph of a patient

Description automatically generated with medium confidence

A screenshot of a test results

Description automatically generated

A close-up of a computer screen

Description automatically generated

# 5. Streamlit application link: