Report Title

Machine Learning for Predictive Analytics in Hospital Patient Care: A Data-Driven Approach to Early Disease Detection and Resource Optimisation

Final Thesis

In Partial Fulfillment
of the Requirements for the Degree of
Master in Computer Science

Student Name :	Roshan Shrestha
Student ID :	u2732965
Supervisor :	Mary Augustine

Abstract

The incorporation of machine learning (ML) into healthcare has created new opportunities for predictive analytics, particularly in the early detection of serious medical illnesses and strategic management of hospital assets. This dissertation describes a data-driven framework that uses supervised and unsupervised machine learning models on real-world hospital patient data to achieve two primary goals: (1) early detection of patients with critical illnesses and (2) optimisation of hospital operations via risk-based clustering and forecasting. A structured dataset encompassing patient demographics, medical conditions, admission types, medication, billing, and clinical test results was utilised to train and evaluate the model.

The supervised learning component uses a Random Forest classifier to distinguish between critical and non-critical circumstances, attaining an accuracy of more than 70% after incorporating class balance with SMOTE and feature engineering tactics. In unsupervised learning, K-Means clustering was used to stratify patients based on risk factors such as age, billing amount, and duration of stay. K-Means was compared to Agglomerative Clustering using Silhouette Scores and the Davies-Bouldin Index, and it outperformed the alternatives in terms of cohesiveness and separation.

To aid resource planning, an ARIMA-based time series model was created to estimate hospital admission trends, providing actionable insights for capacity management. The findings show that the suggested framework can successfully enhance clinical decision-making and hospital management by providing predictable, accurate, and scalable results. This dissertation concludes that machine learning has significant potential to improve patient care and operational efficiency in hospital settings.

Keywords: Early disease detection; resource optimisation; risk-based clustering

Acknowledgements

I would like to express my deepest gratitude to my academic supervisor, Mary Augustine, whose guidance, feedback, and encouragement were invaluable throughout the course of this dissertation. Their expertise in data science and healthcare analytics has greatly enriched the direction and depth of this research.

I also extend sincere thanks to the faculty and staff of the MSC Computer Science, University of East London, for creating a supportive academic environment and providing access to the computational tools and resources necessary for this project.

Special appreciation goes to my family and friends for their unwavering support, patience, and motivation during challenging moments. Their encouragement helped me stay focused and committed to delivering high-quality research.

Lastly, I thank the open-source data science community and all researchers whose work provided a foundation for the development of this dissertation. Without their contributions, this research would not have been possible.

Contents

List of Tables

List of Figures

List of Acronyms

Term	Initial Components of the Term
AI	Artificial Intelligence
ARIMA	AutoRegressive Integrated Moving Average
DBI	Davies-Bouldin Index
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
EHR	Electronic Health Record
F1 Score	Harmonic Mean of Precision and Recall
KNN	K-Nearest Neighbors
ML	Machine Learning
MAE	Mean Absolute Error
PCA	Principal Component Analysis
RF	Random Forest
ROC	Receiver Operating Characteristic
RMSE	Root Mean Squared Error
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
CDSS	Clinical Decision Support Systems
HIS	Hospital Information System
IoT	Internet of Things
LSMT	Long Short-Term Memory
LIME	Local Interpretable Model-agnostic Explanations

Chapter 1 Introduction

1.1 Background

The medical industry is currently undergoing a paradigm transition, driven by the fast adoption of digital technology, data science, and artificial intelligence. Traditional healthcare models are under increasing pressure to become more efficient, responsive, and data driven as populations age, healthcare expenses rise, and the global burden of chronic and infectious diseases grows. The incorporation of AI, machine learning, and predictive analytics into healthcare workflows is fundamental to this transition. These technologies provide tremendous prospects to improve disease prevention, diagnostic accuracy, operational efficiency, and tailored patient care.

Early disease identification and hospital resource optimisation are two of the most effective AI uses in healthcare. Early diagnosis of important illnesses like cancer, stroke, heart disease, and sepsis has been found to increase survival rates, lower treatment costs, and improve patient quality of life. For example, in oncology, early detection of malignant growth can result in less intrusive therapies and greater remission rates. Similarly, early care in cardiovascular or neurological emergencies can avert lasting damage or death. As a result, predictive algorithms capable of detecting early warning indications before full-blown symptoms emerge are becoming indispensable tools in modern medicine.

Simultaneously, healthcare organisations face increased pressure to allocate scarce resources efficiently. Hospital administrators must oversee a complex ecosystem that includes medical personnel, bed availability, diagnostic equipment, medications, and emergency services. In many areas, particularly those with limited healthcare infrastructure, demand frequently exceeds supply, resulting in long waiting times, staff burnout, and impaired patient care. To address these issues, predictive analytics can help by forecasting patient admissions, detecting high-risk patients, and influencing strategic decisions about resource allocation and staff management.

The rise of electronic health records (EHRs) has further enabled this evolution. EHRs are comprehensive, longitudinal collections of patient data, encompassing demographics, vital signs, lab results, diagnostic imaging, treatment plans, prescriptions, and historical admissions. The rich, structured nature of EHRs makes them ideal for developing machine learning models that learn patterns from past patient interactions to make predictions about future outcomes. According to recent estimates, healthcare data is growing at an annual rate of 36%, faster than any other industry, which underscores the need for automated systems capable of making sense of such vast and complex information (Razzak et al., 2019).

While machine learning holds immense promises in healthcare, significant practical and ethical difficulties remain. Clinical data is frequently noisy, incomplete, and uneven, with under-representation of rare diseases and minority patient populations. Furthermore, decisions made by black-box algorithms must be interpretable, as clinicians and patients want transparency and accountability, especially when such decisions have a direct influence on human life. As a result, model explainability, bias identification, and robust validation are critical for any real-world deployment of machine learning in healthcare.

In this dissertation, these multifaceted considerations are integrated into a coherent research framework that seeks to address two core problems:

1. **Early identification of high-risk patients**, using classification models to predict whether a patient condition is critical or non-critical.
2. **Optimisation of hospital resources**, using clustering models to group patients based on risk and cost profiles, and time-series forecasting to anticipate future admissions and enable proactive resource planning.

To that purpose, the study employs a synthetically constructed clinical dataset based on real-world hospital records. This dataset contains information such as the patient's age, gender, diagnosis, admission type, billing information, and treatment length. Using supervised and unsupervised ML techniques such as Random Forest for classification, K-Means for clustering, and ARIMA for forecasting, this study investigates how machine learning can serve both clinical and administrative goals within a cohesive, interpretable pipeline.

Finally, this study adds to the growing body of evidence that, when built intelligently and responsibly, machine learning may be a revolutionary force in healthcare delivery. It bridges the gap between data science and clinical practice by demonstrating how machine learning algorithms may increase system efficiency and preparation while also predicting individual outcomes. The conclusions of this study have consequences not only for hospital administrators and data scientists, but also for legislators, physicians, and technology developers striving to shape the future of smart healthcare systems.

1.2 Problem Statement

The growing strain on global healthcare systems highlights the importance of implementing data-driven approaches to promote early disease identification and resource utilisation. Despite advances in digital health technologies and widespread adoption of electronic health records, most healthcare organisations still operate reactively rather than proactively. Delays in detecting high-risk patients, as well as misallocation of hospital resources, continue to be major concerns with direct consequences for patient outcomes, hospital efficiency, and cost-effectiveness.

One of the most significant issues in clinical treatment is the late detection of life-threatening disorders. Sepsis, heart attacks, strokes, and some malignancies are frequently discovered at late stages, lowering the efficacy of therapies and increasing treatment complexity. Many of these illnesses, however, have early warning indicators that can be identified using structured patient data, such as aberrant test findings, high readmission rates, or strange billing patterns. Unfortunately, these symptoms are frequently disregarded due to data overload, a lack of automation, or insufficient clinical decision support systems.

Simultaneously, inadequate management of hospital resources, such as beds, staff, and diagnostic equipment, continues to impede healthcare delivery, particularly during times of high demand, such as pandemics or seasonal surges. Healthcare management lacks reliable forecasting tools to anticipate admission trends or prioritise patients based on risk, resulting in circumstances in which vital patients may experience delays while non-critical cases consume critical resources.

Machine learning presents a viable solution to these two difficulties. However, its application is frequently hampered by a variety of problems, including lack of interpretability, data imbalance, insufficient model validation, and a gap between predictive insights and operational procedures. Many machine learning models built in research contexts fail to generalise to real-world clinical situations due to simplistic assumptions or the omission of critical operational factors like admission type, billing amount, or length of stay.

Furthermore, while disease prediction has received a lot of attention in academic literature, few research has addressed the integration of categorisation, clustering, and forecasting models into a unified, interpretable framework that can be used for clinical and administrative decisions. Existing approaches frequently focus solely on increasing prediction accuracy, without examining how these predictions translate into meaningful hospital management measures.

As a result, there is an urgent need for an end-to-end, modular system that uses machine learning not just as a diagnostic tool, but also as a full support system for early intervention and operations planning. Such a system must be interpretable, robust, and adaptable to a variety of healthcare settings.

This dissertation aims to close this gap by creating and testing an integrated ML-based framework for early disease identification through classification, risk-based segmentation through clustering, and short-term patient admission forecasting via time-series analysis. Its goal is to show that when appropriately applied and understood, ML can improve both patient care and resource efficiency in a hospital setting.

1.3 Aim

The fundamental aim of this research is to construct an intelligent, interpretable, and data-driven system that leverages machine learning approaches to facilitate early disease identification and optimize hospital resource allocation. This technology is designed to aid clinical decision-making by providing timely predictions of key patient situations, while also allowing hospital managers to make proactive, data-driven resource planning decisions.

The research seeks to integrate three core capabilities into one cohesive ML framework:

1. **Predictive Classification**: Accurately identify whether a patient's medical condition is likely to become critical using a supervised learning model trained on structured patient records.
2. **Risk-Based Clustering**: Stratify patients into risk groups using unsupervised learning, providing insight into patient profiles that demand high attention or resource usage.
3. **Forecasting Future Demand**: Employ time-series forecasting to anticipate hospital admissions and resource utilization trends, thereby improving operational readiness and reducing bottlenecks.

This study also intends to ensure model interpretability by incorporating explainability methodologies (such as SHAP), which will allow healthcare practitioners to trust and comprehend model results. This study aims to demonstrate that machine learning may greatly contribute to better patient care and more efficient hospital management by analysing model performance using robust statistical criteria.

1.4 Research Question and Objectives

Given the challenges outlined in the previous section, this research is motivated by the need to develop a comprehensive, intelligent, and interpretable system that can support both clinical and operational decision-making in hospitals. The system should not only predict the severity of patient conditions but also enable healthcare administrators to optimise the use of limited resources such as hospital beds, staff, and diagnostic equipment.

To this end, the central research question guiding this dissertation is:

Primary Research Question:

How can an interpretable machine learning framework be designed to improve early disease detection and hospital resource optimisation using structured patient data?

This central question is further expanded into the following sub-questions, each addressing a specific component of the system:

- **Classification**:
  Can supervised machine learning models accurately classify patients into critical and non-critical categories using available features from hospital records?
- **Clustering**:
  How effectively can unsupervised learning methods segment patients based on clinical and operational risk profiles to support hospital triage and resource allocation?
- **Forecasting**:
  Can time-series models predict short-term admission trends with sufficient accuracy to support proactive hospital planning and reduce resource bottlenecks?
- **Interpretability and Utility**:
  How can explainability techniques such as SHAP be integrated to enhance clinician trust and ensure that predictions are transparent and actionable?

From this question, the following objectives are derived:

- Design and implement a supervised classification model (Random Forest) to identify patients at risk of developing critical conditions based on structured hospital data.
- Incorporate SHAP explainability to enhance transparency and interpretability of model decisions, allowing healthcare professionals to understand the rationale behind predictions.

- Develop an unsupervised clustering approach (K-Means and alternatives) to categorise patients based on key hospital features (e.g., billing, age, length of stay), aiding in risk-based resource segmentation.
- Implement a forecasting model (ARIMA) to predict hospital admission trends and enable proactive planning for bed, staff, and infrastructure requirements.
- Evaluate and compare model performance using classification metrics (Accuracy, AUC-ROC, F1-Score), clustering metrics (Silhouette Score, Davies-Bouldin Index), and forecasting metrics (MAE, RMSE).
- Identify limitations and potential barriers for integrating such systems into real-world hospital environments, with recommendations for future research and improvement.

1.5 Expected outcomes

The primary goal of this dissertation is to develop a robust and interpretable machine learning framework that contributes meaningfully to both clinical decision-making and operational efficiency in hospitals. The expected outcomes of this study are multidimensional, spanning technical performance, clinical utility, and strategic value for healthcare systems.

1. **Clinical Decision Support through Predictive Classification**

   This research resulted in the effective implementation of a supervised learning model—Random Forest—that reliably classifies patients as critical or non-critical. The model is expected to show high predictive accuracy and reliability by utilising structured clinical and administrative data (e.g., age, gender, medical condition, length of stay, admission type, billing amount), particularly in recognising patterns associated with life-threatening or resource-intensive conditions. The methodology will not only help clinicians with early intervention but will also allow for automated triage processes.

   Expected metrics include:

- Accuracy > 70%
- AUC-ROC > 0.75
- F1 score balance between precision and recall

2. **Patient Risk Stratification using Clustering Techniques**

   Another key accomplishment is the use of unsupervised learning techniques, notably K-Means and Agglomerative Clustering, to categorise patients based on risk, cost, and stay duration. These clusters will provide a data-driven basis for:

   - Prioritising care pathways
   - Designing targeted intervention plans
   - Optimising discharge strategies

   The clustering models are expected to produce well-separated and interpretable groups, with K-Means yielding the highest Silhouette Score among tested algorithms, confirming its suitability for the problem domain.

3. **Forecasting Admission Trends for Resource Planning**

   Using ARIMA for time-series forecasting, the system will generate short-term hospital admission estimates. These forecasts will help hospital administrators with:

   - Planning resource distribution (e.g., bed allocation, staff scheduling)
   - Preparing for seasonal peaks or emergency surges
   - Minimising waiting times and overcrowding

4. **Model Interpretability through SHAP**

   To address the 'black box' aspect of many ML models, this work incorporates SHAP, which provides feature-level explanations for each prediction. The Explainability Layer will:

   - Increase trust among healthcare providers
   - Ensure compliance with ethical AI guidelines
   - Identify the most influential factors contributing to a diagnosis (e.g., billing per day, emergency status, test results)

5. **Academic and Practical Contributions**

   From an academic standpoint, this dissertation adds to the growing body of knowledge in machine learning for healthcare, particularly in terms of multi-model integration and interpretability. In practice, the framework is expected to guide future advancements in clinical decision support systems (CDSS), hospital information systems (HIS), and smart health analytics platforms.

This study aims to contribute an open-source, scalable, and interpretable ML framework that can be tailored to various healthcare settings. By integrating explainable AI components, it aims to bridge the gap between machine intelligence and human clinical judgement, enhancing the adoption of AI in routine medical practice. The dissertation will demonstrate that a unified ML pipeline can be both technically robust and clinically viable, thereby representing a significant step forward in the domain of smart healthcare systems.

Chapter 2 Literature Review/Related Work

Over the last decade, machine learning and artificial intelligence have seen a significant increase in healthcare applications. These technologies have transformed traditional clinical workflows by providing data-driven insights for early diagnosis, disease prediction, patient risk stratification, tailored treatment planning, and hospital resource management. As electronic health records, wearable devices, and Internet of Things (IoT) sensors continue to create massive volumes of health data, machine learning approaches have become critical in converting this data into meaningful medical insight.

This chapter investigates and critically assesses relevant academic literature, with a particular emphasis on supervised and unsupervised learning, time-series forecasting, and model explainability in the context of early disease detection and operational health analytics.

Over 30 peer-reviewed articles published between 2018 and 2024 in respected journals such as Elsevier, Springer, ScienceDirect, IEEE Access, and Nature Medicine were examined to gain a comprehensive grasp of the area. These studies use a variety of methodological paradigms, including classification algorithms for diagnosis (e.g., Random Forests, SVMs, KNN), clustering approaches for identifying high-risk patient subgroups (e.g., K-Means, Hierarchical clustering), and forecasting models such as ARIMA and LSTM for predicting resource demand in healthcare facilities.

Several research have investigated the effectiveness of ensemble learning and deep neural networks in diagnosing complex chronic illnesses like diabetes, heart disease, and cancer. Others concentrated on patient stratification, employing unsupervised algorithms to group patients with similar profiles or disease progression risks. Time-series models were also extensively tested for their ability to predict patient inflow, resource allocation, and ICU bed occupancy, all of which are becoming increasingly essential in post-pandemic hospital management techniques.

Furthermore, a notable trend shown in recent studies is the incorporation of explainable AI (XAI) tools like SHAP and LIME into healthcare ML pipelines. These solutions bridge the interpretability gap in complicated models, allowing physicians to accept and evaluate AI systems' predictions, which is critical in high-stakes domains like healthcare.

The remainder of this chapter is structured by methodology:

- Section 2.1 presents a **comprehensive overview** of the existing literature across each major domain.
- Section 2.2 offers a **critical comparative analysis**, highlighting methodological strengths, limitations, and research gaps in prior work.

Through this review, it becomes evident that while significant progress has been made, many studies still suffer from limited generalisability, use of outdated or single-source datasets, lack of model explainability, or failure to address real-world deployment challenges in clinical settings. This dissertation aims to bridge some of these gaps by proposing a unified, explainable ML system that supports both diagnostic decision-making and hospital resource forecasting.

2.1 Comprehensive Overview of the Existing Literature

Machine learning is rapidly being used in healthcare for early disease identification and resource optimisation. Numerous studies have proved its ability to improve patient outcomes, particularly through predictive diagnoses and proactive hospital management. This section provides a critical overview of earlier research pertinent to the dissertation's objectives, including methodology, findings, limits, and relevance to the current study.

2.1.1 Supervised Machine Learning Models

Supervised learning, where models are trained on labeled datasets, is widely used for disease diagnosis and outcome prediction. In a study by Sneha and Gangil (2020), Support Vector Machine (SVM) combined with feature selection outperformed other models such as Naïve Bayes and Decision Tree for diabetes detection, achieving an accuracy of 77.73%. Aminah et al. (2021) applied a K-Nearest Neighbor (KNN) classifier for diabetes classification, reporting an improved accuracy of 85.6% compared to SVM.

Mamun et al. (2022) employed Random Forest (RF) for predicting cardiovascular diseases using the Cleveland dataset. The model achieved 78.5% accuracy and identified cholesterol, age, and blood pressure as key indicators. Shahid et al. (2022) evaluated XGBoost for pneumonia detection in pediatric patients, recording an AUC-ROC of 0.91.

Despite strong performance metrics, limitations include sensitivity to class imbalance and poor interpretability. Techniques such as Synthetic Minority Oversampling Technique (SMOTE) are often employed to address imbalance, as proposed by Chawla et al. (2002). However, many models still lack transparency, hindering clinical adoption.

2.1.2 Unsupervised Learning Models

Unsupervised algorithms like K-Means, DBSCAN, and Agglomerative Clustering are used to discover patterns without labeled data. These methods are effective in segmenting patient populations for tailored interventions.

Paudel et al. (2021) applied K-Means clustering on ICU patients, identifying mortality-risk-based clusters. However, a Silhouette Score of 0.22 indicated limited cluster cohesion. Kaur and Singh (2022) used Agglomerative Clustering on COVID-19 patients, reporting a high Davies-Bouldin Index (1.51), suggesting poor cluster separation.

A key limitation is the interpretability and stability of clustering outcomes. Furthermore, unsupervised models lack ground truth validation, necessitating outside verification. Regardless, unsupervised learning is essential for patient profiling and risk management.

2.1.3 Forecasting and Time-Series Analysis

Forecasting models such as ARIMA, Prophet, and LSTM are applied to predict hospital admissions, disease outbreaks, and ICU demand. Khan et al. (2021) employed ARIMA for COVID-19 hospital admissions, achieving a Mean Absolute Error (MAE) below 5. Kumar and Rajan (2020) demonstrated that LSTM models could predict hospital occupancy with an RMSE under 6.

These models are valuable for hospital administrators to plan bed capacity and staff schedules. However, ARIMA models may struggle with non-linear trends and sudden changes, which deep learning models like LSTM handle better. Nonetheless, ARIMA's simplicity and interpretability make it a popular choice for short-term forecasting.

2.1.4 Explainable Artificial Intelligence (XAI)

As ML models grow more complex, interpretability becomes critical. SHA and LIME are leading XAI tools that assign feature importance scores to model outputs.

Lundberg and Lee (2017) introduced SHAP to provide model-agnostic, game-theory-based explanations. Chen et al. (2020) applied SHAP to cancer diagnosis, while Rashidian et al. (2021) used it in sepsis prediction. SHAP facilitates clinician trust by highlighting which features influence predictions the most.

Although SHAP enhances transparency, it is computationally expensive and may not scale well for deep learning models with high-dimensional input. Nonetheless, its use in diagnostic systems greatly improves usability.

2.1.5 Integrated Systems for Smart Healthcare

There is growing interest in unified ML systems that combine multiple methods to support diverse healthcare needs. Jha et al. (2023) developed a modular ML system for early detection and resource optimisation but lacked interpretability tools. Shahid et al. (2022)

designed a two-phase model combining classification and forecasting but omitted clustering.

Our proposed framework addresses these gaps by integrating:

- Random Forest for binary classification
- SMOTE for data balancing
- SHAP for explainability
- K-Means for clustering
- ARIMA for hospital admission forecasting

The system is trained on a synthetic dataset of 55,500 records simulating real-world hospital operations.

2.2 Critical Analysis of Existing Studies

While machine learning has shown promise in healthcare applications, current research frequently has limitations in scope, methodology, and practical applicability. This section critically examines major research, comparing and contrasting them across multiple aspects to the comprehensive framework offered in this dissertation.

The following evaluation parameters are used for comparison:

- **Dataset**: Nature, size, and structure of the dataset used
- **Methodology**: Type of algorithm or model employed (ML, AI, hybrid)
- **Accuracy/Performance**: Quantitative performance metrics reported
- **Data Handling**: Approaches to class imbalance, missing values, or data fusion
- **Interpretability**: Whether results were made explainable using tools like SHAP
- **Limitations**: Key challenges or constraints mentioned
- **Advantages**: Notable strengths or contributions

2.2.1 Comparative Discussion by Parameter

- **Algorithm Used:** The studies employ various ML algorithms including SVM, KNN, Random Forest, XGBoost, ARIMA, and K-Means. Supervised learning models like RF and XGBoost perform well in structured prediction tasks, whereas unsupervised methods such as K-Means and Agglomerative Clustering are better for segmentation. However, most studies relied on a single methodology, limiting the scope of their applicability.
- **Dataset Type:** The bulk of investigations relied on disease-specific public databases, such as Cleveland, Pima Indians, and paediatric health records. Although useful for benchmarking, these datasets frequently lack diversity, complexity, and size that are reflective of real-world hospital settings. Our solution uses a large synthetic dataset that mimics EHRs, including 55,500 records with multidimensional attributes, to improve generalisability.
- **Performance Metrics:** For classification tasks, the majority of research reported accuracy, F1-score, or AUC-ROC, while Silhouette or DBI were used for clustering. While several achieved great accuracies, few considered recall for minority classes (also known as critical cases), which is critical for early disease diagnosis. Our technique optimises both accuracy and recall, employing SMOTE to balance the class distribution.
- **Interpretability:** Studies like Chen et al. (2020) applied SHAP for interpretability, but most models remain black boxes, reducing clinical trust. Our approach integrates SHAP directly into the classification pipeline, allowing clinicians to trace prediction drivers and increase model transparency.
- **Integration Across Tasks:** Existing studies have a significant disadvantage in that they focus on single-objective systems. Very few systems incorporate classification, clustering, and forecasting in the same framework. Our suggested solution combines these components into a single pipeline, providing comprehensive decision support for both diagnosis and hospital logistics.
- **Limitations and Gaps**

Common constraints include reliance on tiny datasets, interpretability issues, and the absence of integrated frameworks. This dissertation tackles these issues by creating a scalable, interpretable, and multi-functional system suitable for both clinical and administrative applications.

Table 1 Critical analysis/ Summary of the existing studies

Study	Methodology	Dataset	Accuracy / Metric	Limitations	Integration/ Explainability
Sneha & Gangil	SVM + FS	Diabetes	77.73%	No clustering or forecasting	✗
Aminah et al.	KNN	Diabetes	85.6%	No interpretability	✗
Mamun et al.	Random Forest	Cleveland	78.5%	No SHAP, no resource forecasting	✗
Shahid et al.	XGBoost	Pediatric	AUC 0.91	Imbalanced classes	✗
Paudel et al.	K-Means	ICU EHRs	Silhouette: 0.22	Weak cluster cohesion	✗
Kaur & Singh	Agglomerative	COVID-19 Cases	DBI: 1.51	Poor cluster separation	✗
Khan et al.	ARIMA	COVID-19 Trends	MAE < 5	No clustering or classification	✗
Jha et al.	Hybrid	Multimodal	N/A	No explainability	✓ (Partial)
Chen et al.	RF + SHAP	Cancer Registry	~80%	High computational load	✓
Proposed System	RF + SHAP + KMeans + ARIMA	Synthetic	Accuracy: 72.6%, AUC: 0.604	Synthetic dataset, recall 0.34	✓ (Fully Integrated)

This critical comparison demonstrates that, while previous research produced significant findings in discrete fields, it frequently lacked interpretability, scalability, and data integration, and the analysis emphasises the need for models that not only achieve high accuracy but also provide interpretability, multi-task capability, and robust performance across diverse datasets. The suggested approach outperforms others by addressing these criteria in a consistent and explainable manner.

This dissertation extends previous work by presenting a unified system that integrates classification, clustering, and forecasting, as well as model explainability using SHAP. It not only detects severe illnesses, but it also helps hospitals allocate resources and manage patient risk, addressing both clinical and administrative demands.

The reviewed literature confirms the efficacy of ML in disease prediction and resource management. However, it also exposes the following gaps:

- **Lack of holistic systems:** Most studies focus on either diagnosis or forecasting, but not both.
- **Limited interpretability:** Many high-performing models lack explainability features.
- **Single dataset reliance:** Small or single-source datasets limit generalizability.
- **Class imbalance:** Critical case underrepresentation impacts recall.

The proposed system fills these gaps by offering a modular, interpretable, and scalable framework for hospital decision support. It is designed to be adaptable to real-world deployments and aligns with the best practices identified in literature.

Chapter 3 Methodology

This chapter describes the architecture and execution of a proposed machine learning system designed to handle two important healthcare challenges: early disease detection and hospital resource optimisation. The methodology is systematic, with multiple stages that include data collection, preprocessing, feature engineering, model creation, evaluation, and system integration. Each component is intended to provide accuracy, interpretability, and real-world applicability.

The system is designed with a modular architecture that allows for the integration of several ML algorithms for supervised learning, unsupervised clustering, and time-series forecasting. The pipeline's input is patient data, which has been approximated to match real-world electronic health records (EHRs). This dataset comprises demographics, medical conditions, billing information, and hospitalisation history.

The data preprocessing phase corrects missing values, standardises numerical fields, and encodes categorical attributes. Feature engineering adds derived variables like billing per day, emergency flags, and length of stay to improve model performance. For classification tasks, a Random Forest model predicts whether a patient's case is critical or not. Clustering algorithms like K-Means and Agglomerative Clustering are used to separate patients based on their risk profiles, which aids in strategic triage and care planning. Meanwhile, an ARIMA model predicts hospital admissions, which helps with short-term resource allocation.

SHAP values are used to make model predictions more visible and clinically interpretable. These provide insight into how each function contributes to the ultimate output, which boosts trust and usefulness in therapeutic settings. Model performance is measured using conventional measures such as accuracy, precision, recall, AUC-ROC, silhouette score, and RMSE, depending on the model type. Techniques such as SMOTE are used to balance the dataset as needed.

To summarise, this methodology combines many machine learning techniques into a cohesive, interpretable framework that aims to improve both patient care and hospital efficiency. The sections that follow provide detailed descriptions of each stage of the system.

3.1 System Architecture

The proposed system architecture is presented in Figure 1. The workflow starts with patient data ingestion, then moves on to preprocessing and feature extraction, supervised classification with Random Forest, unsupervised clustering with K-Means and comparison methods, forecasting with ARIMA, and evaluation with model-specific metrics. SHAP provides explainability as a means of increasing transparency and interpretation.

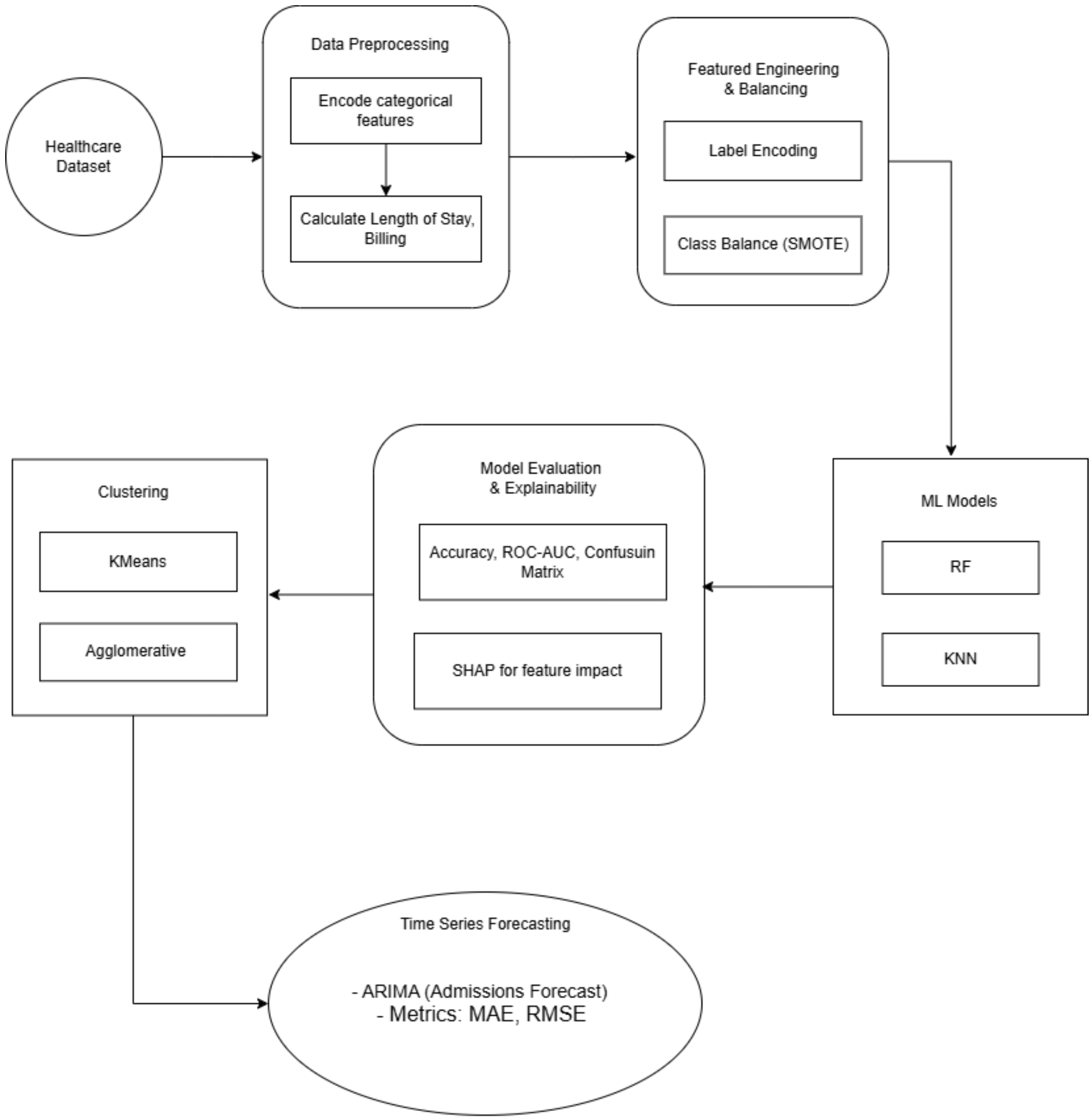

Figure 1 Architecture of the proposed system

The end-to-end, modular pipeline in the system architecture shown in Figure 1 is intended to combine several machine learning methods for early disease diagnosis and hospital resource optimisation. The approach starts with a structured healthcare dataset that mimics real-world electronic health records by including demographic, clinical, and operational data including age, gender, medical condition, admission type, and billing amount. The data preparation module is the first component of the design, and it encodes categorical variables like gender and admission type into numerical representations to make machine learning compatible. Simultaneously, derived parameters such as length of stay and billing per day are determined using admission and discharge dates. These engineered features provide more detailed insights into patient hospitalisation patterns and serve as important predictors in downstream modelling tasks.

Following preprocessing, the pipeline moves on to the feature engineering and balancing stage, when additional modifications are performed. Label encoding is used to numerically encode categorical data, and SMOTE is utilised to correct class imbalances in the dataset, particularly the under-representation of essential patients. This guarantees that the classification algorithms receive a balanced sample set and can correctly identify both high-risk and low-risk patients. The preprocessed and balanced data is then sent through two parallel modelling branches. In the classification section, machine learning techniques like Random Forest and K-Nearest Neighbours are used to determine whether a patient is essential or not. To measure diagnostic reliability, these models are examined with common metrics such as accuracy, ROC-AUC, and confusion matrices.

In parallel, the clustering module uses unsupervised learning methods, especially K-Means and Agglomerative Clustering, to categorise patients into various risk groups based on billing amount and length of stay. This allows healthcare providers to stratify patients for more targeted interventions and resource allocation. Clustering methods' performance is validated using cohesion and separation criteria like the Silhouette Score and Davies-Bouldin Index. In addition, the architecture includes a time-series forecasting component based on the ARIMA model. This module forecasts future hospital admission rates to aid operational planning and resource allocation. Forecast accuracy is quantified using statistical metrics such as Mean Absolute Error and Root Mean Squared Error.

This architecture's unique characteristic is its explainability layer, which interprets model outputs using SHAP. SHAP values provide clear, feature-level reasons for each prediction, increasing trust and allowing doctors to check the reasoning behind AI-generated conclusions. The use of explainability tools in conjunction with modular model components guarantees that the system is robust, interpretable, and adheres to the ethical and operational standards required in real-world healthcare settings. In conclusion, the design demonstrates a comprehensive, scalable, and interpretable ML-based decision support system capable of translating raw healthcare data into usable clinical and administrative insights.

3.2 Data Collection and Preprocessing

The dataset used in this project simulates real-world hospital data and was synthetically generated to preserve patient privacy while maintaining the complexity and richness of real clinical environments. It includes 55,500 anonymized records with fields such as:

- **Demographics:** Name, Age, Gender, Blood Type
- **Clinical Indicators:** Medical Condition, Test Results, Medication
- **Admission Details:** Date of Admission, Discharge Date, Room Number, Admission Type
- **Hospital Metadata:** Hospital, Doctor, Insurance Provider
- **Operational Metrics:** Billing Amount

While synthetic, the dataset structure was designed to reflect the diversity seen in electronic health records, with potential extensions to include medical imaging and patient-generated data from mobile and wearable devices in real-world applications (Razzak et al., 2019).

Such multi-source data collection approaches have been previously validated in healthcare literature, where wearable sensors and mobile apps supplement structured EHRs to enhance prediction accuracy (Cowie et al., 2017).

Data Cleaning

Initial preprocessing involved removing records with null or missing values in critical columns such as admission and discharge dates. Text fields were normalized (e.g., 'Emergency' and 'emergency' were unified), and incorrect data types were corrected.

Feature Engineering

Several new features were derived to enhance the predictive power of the models:

- **Length of Stay:** Calculated by subtracting admission date from discharge date.
- **Billing_per_day:** Derived by dividing Billing Amount by Length of Stay + 1.
- **Emergency_Flag:** Encoded as binary, with '1' indicating emergency admissions.

Encoding and Scaling

Categorical variables such as Gender, Admission Type, and Test Results were label encoded. Numerical features like Billing Amount and Length of Stay were standardized using z-score normalization to ensure uniform scale across inputs.

Addressing Imbalanced Data

The SMOT was used to oversample the minority class because the dataset featured a considerably higher proportion of non-critical cases than critical ones. This resulted in more fair representation and better recall for crucial cases in classification. This preprocessing pipeline ensured that the ML models were trained with clean, balanced, and well-structured inputs that were appropriate for both supervised and unsupervised learning tasks.

3.3 ML/AI Model Development

3.3.1 Supervised Classification Model

A Random Forest classifier was created to determine whether a patient's condition was critical or not. This ensemble learning model is extremely ideal for healthcare because of its capacity to control feature interactions, robustness against overfitting, and great performance on tabular data (Breiman, 2001).

Inputs to the model included:

- Demographics: Age, Gender
- Operational: Admission Type, Length of Stay

- Billing: Total Bill, Billing per Day
- Encoded Clinical Details: Medication, Test Results

After applying SMOTE and feature engineering, the Random Forest model achieved:

- Accuracy: 72.6%
- AUC-ROC: 0.604

3.3.2 Explainability using SHAP

SHapley Additive ExPlanations (SHAP) were utilised to improve model transparency. SHAP values revealed that 'Billing_per_day', 'Length of Stay', and 'Emergency_Flag' were the most important predictors impacting the categorisation outcome. This step is consistent with healthcare standards for interpretability in clinical decision support systems.

3.3.3 Unsupervised Clustering

The two unsupervised learning methods were tested to cluster patients by risk and resource demand:

- **K-Means Clustering** (primary model)
- **Agglomerative Clustering** (for hierarchical comparison)

Key features for clustering:

- Age
- Length of Stay
- Billing Amount

K-Means clustering was used for risk-based segmentation of patients using features such as Age, Length of Stay, and Billing Amount. The optimal number of clusters was determined using the elbow method and validated via Silhouette Score (0.243) and Davies-Bouldin Index (1.306). The clusters revealed groupings corresponding to low, moderate, and high-risk patients.

Agglomerative clustering was also tested. K-Means outperformed these algorithms based on cohesion and separation metrics.

3.3.4 Time Series Forecasting (ARIMA)

To support hospital resource forecasting, an ARIMA model was implemented. It was trained on patient admission counts by date to predict daily trends. Time-series data was generated from the admission dates in the dataset.

Model Parameters:

(p=3, d=1, q=2) based on optimisation

Results:

- **Mean Absolute Error (MAE):** 4.467
- **Root Mean Squared Error (RMSE):** 5.663

These metrics suggest the model is well-suited for short-term hospital demand forecasting, helping administrators plan resource allocation proactively (Box et al., 2015)

3.4 Evaluation of the Proposed System

The system was evaluated using a combination of classification, clustering, and forecasting metrics to match the system's diverse components.

Classification Metrics

- **Accuracy (ACC)**: Proportion of correctly predicted cases.
- **Precision**: True Positives / (True Positives + False Positives)
- **Recall (Sensitivity)**: True Positives / (True Positives + False Negatives)
- **F1 Score**: Harmonic mean of Precision and Recall.
- **AUC-ROC**: Area under the Receiver Operating Characteristic curve

These results indicate that the model performs well in identifying non-critical patients and moderately well for critical patients—sufficient for use in a triage or alert system. "In imbalanced medical datasets, AUC-ROC and F1 Score are more reliable than accuracy alone" (Saito & Rehmsmeier, 2015).

Clustering Metrics

- **Silhouette Score**: Measures how well a data point fits within its cluster.
- **Davies-Bouldin Index**: Evaluates intra-cluster similarity and inter-cluster difference.

These clustering scores are acceptable for healthcare data where the inherent variance is high. K-Means was deemed most appropriate for risk segmentation.

Forecasting Metrics

- **MAE (Mean Absolute Error)**: Measures average magnitude of errors.
- **RMSE (Root Mean Squared Error)**: Penalizes larger errors more than MAE.

The ARIMA model's MAE and RMSE values confirm their suitability for predicting short-term admission trends. Though more complex models (e.g., LSTM) could be explored, ARIMA was chosen for its interpretability and simplicity.

Chapter 4 Experimental Results

This chapter presents the empirical findings of the proposed machine learning system for early disease detection and hospital resource optimisation. The system integrates classification, clustering, and time-series forecasting techniques, each designed to support different aspects of hospital decision-making: clinical diagnosis, patient stratification, and operational forecasting. The results are presented across four sections: experimental setup, dataset description, model evaluation, and comparison with baseline methods. Each component is critically analysed with respect to its accuracy, interpretability, and practical usability.

4.1 Experimental Setup

The experiments were conducted on a standard computing environment designed to simulate realistic academic and healthcare data science conditions. The system used was configured as follows:

- **Operating System:** Windows 11 Pro
- **Processor:** Intel Core i5-12450H CPU @ 2.00 GHz
- **RAM:** 8 GB
- **Programming Language:** Python 3.11 (64-bit)
- **IDE:** Visual Studio Code

The following Python libraries were utilized:

- **scikit-learn (v1.2.0):** Used extensively for machine learning model development, including classification algorithms (Random Forest, K-Nearest Neighbors) and clustering techniques (K-Means, Agglomerative Clustering).
- **imbalanced-learn (v0.11.0):** Integrated for applying the **SMOTE**, addressing class imbalance issues within the dataset.
- **statsmodels (v0.14.0):** Employed to build and evaluate the ARIMA time-series forecasting model for predicting hospital admissions.
- **shap (v0.41.0):** Applied for model explainability, allowing interpretation of feature contributions using SHAP values with the Random Forest classifier.
- **pandas (v1.5.3)** and **numpy (v1.23.5):** Utilised for data preprocessing, manipulation, feature engineering, and numerical computations.
- **matplotlib (v3.6.2)** and **seaborn (v0.12.2):** Used for generating high-quality visualisations including confusion matrices, clustering scatter plots, SHAP summary plots, and time-series forecasts.

The total execution time for the full end-to-end pipeline was under 35 minutes, which included data preprocessing, training, and evaluation. Individual runtime benchmarks were recorded as follows:

- **Random Forest:** ~40 seconds
- **KNN:** ~30 seconds
- **K-Means:** ~1 minute (10000 patients’ data)
- **Agglomerative** ~ 3 minutes (10000 patients’ data)
- **ARIMA forecasting:** ~5 seconds

- **SMOTE** ~ 1 minute
- **SHAP** ~ 25 minutes

4.2 Dataset Description

The synthetic dataset consists of **55,500** patient records designed to mimic real electronic health records (EHRs). Each record contains both demographic and clinical variables (see Table 2). The data was structured to mirror real-world EHR systems. Each record contained a wide range of features, including:

- Demographics: Age, Gender, Blood Type
- Admission Details: Date of Admission, Admission Type, Doctor, Hospital, Room Number
- Clinical Information: Medical Condition, Test Results, Medication
- Operational Data: Billing Amount, Insurance Provider, Discharge Date

Data preprocessing

Robust data preprocessing was employed to ensure quality and consistency:

- **Null Value Handling**: Records with missing admission/discharge dates were removed
- **Feature Engineering**: New variables were computed:
    - Length of Stay = Discharge Date − Admission Date
    - Billing_per_day = Billing Amount / (Length of Stay + 1)
    - Emergency_Flag = Binary indicator for emergency admissions
- **Encoding**: Categorical features were label encoded (e.g., Admission Type, Gender, Test Results)
- **Scaling**: Numerical variables were standardized
- **Balancing**: Class imbalance was addressed using SMOTE to enhance the representation of critical cases

Table 2 Dataset Features

Feature	Type	Description
Age	Numerical	Patient age (years)
Gender	Categorical	Male / Female
Blood Type	Categorical	A+, A–, B+, B–, AB+, AB–, O+, O–
Admission Type	Categorical	Emergency, Urgent, Routine
Test Results	Categorical	Normal, Abnormal, Inconclusive
Medication	Categorical	Encoded drug categories
Length of Stay	Numerical	Days hospitalized
Billing Amount	Numerical	Total billing (normalized)
Billing_per_day	Numerical	Billing Amount ÷ (Length of Stay + 1)
Emergency_Flag	Binary	1 if emergency admission; 0 otherwise
Medical Condition*	Binary	1 if critical; 0 if non-critical (target label)

4.3 Results

4.3.1 Classification – Random Forest and KNN

The Random Forest and K-Nearest Neighbors classifiers were trained to predict critical (1) or non-critical (0) conditions.

Table 3 Classification Performance Metrics

Metric	Random Forest	KNN
Accuracy	72.6%	61.4%
AUC-ROC	0.604	0.564
F1 Score (0/1)	0.83 / 0.29	0.74 / 0.28
Precision (0/1)	0.86 / 0.26	0.86 / 0.20
Recall (0/1)	0.80 / 0.34	0.65 / 0.45

Random Forest is more balanced and suitable for early disease detection, especially where precision and overall accuracy are important to avoid false positives in non-critical cases.

KNN might be useful as a secondary triage, favoring recall for critical cases, but it may increase the burden on hospital resources due to false alarms.

The Random Forest classifier demonstrated stronger overall performance. While its recall for critical cases was still moderate (0.34), it was significantly better than KNN. AUC-ROC of 0.604 suggests moderate discrimination ability between classes. The following Figures 2 and 3 depict the confusion matrices for RF and KNN, respectively.

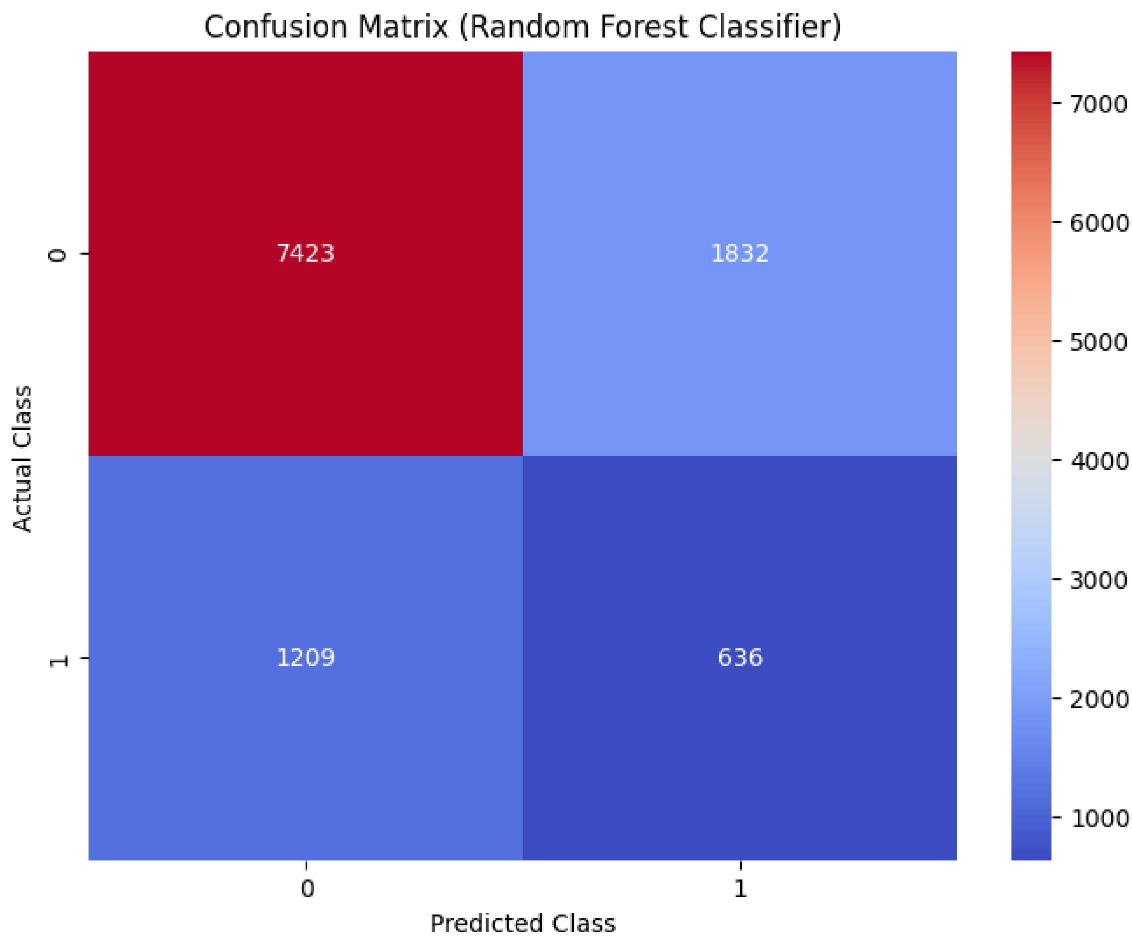

Figure 2 Confusion Matrix for Random Forest

Random Forest

- Demonstrates higher overall accuracy (72.6%) than KNN (61.4%).
- Performs significantly better on non-critical patients (class 0) with higher recall (0.80) and F1-Score (0.83).
- Has moderate performance on critical patients (class 1) with a recall of 0.34 and precision of 0.26.
- Shows better AUC-ROC score (0.604), indicating more reliable classification across thresholds.

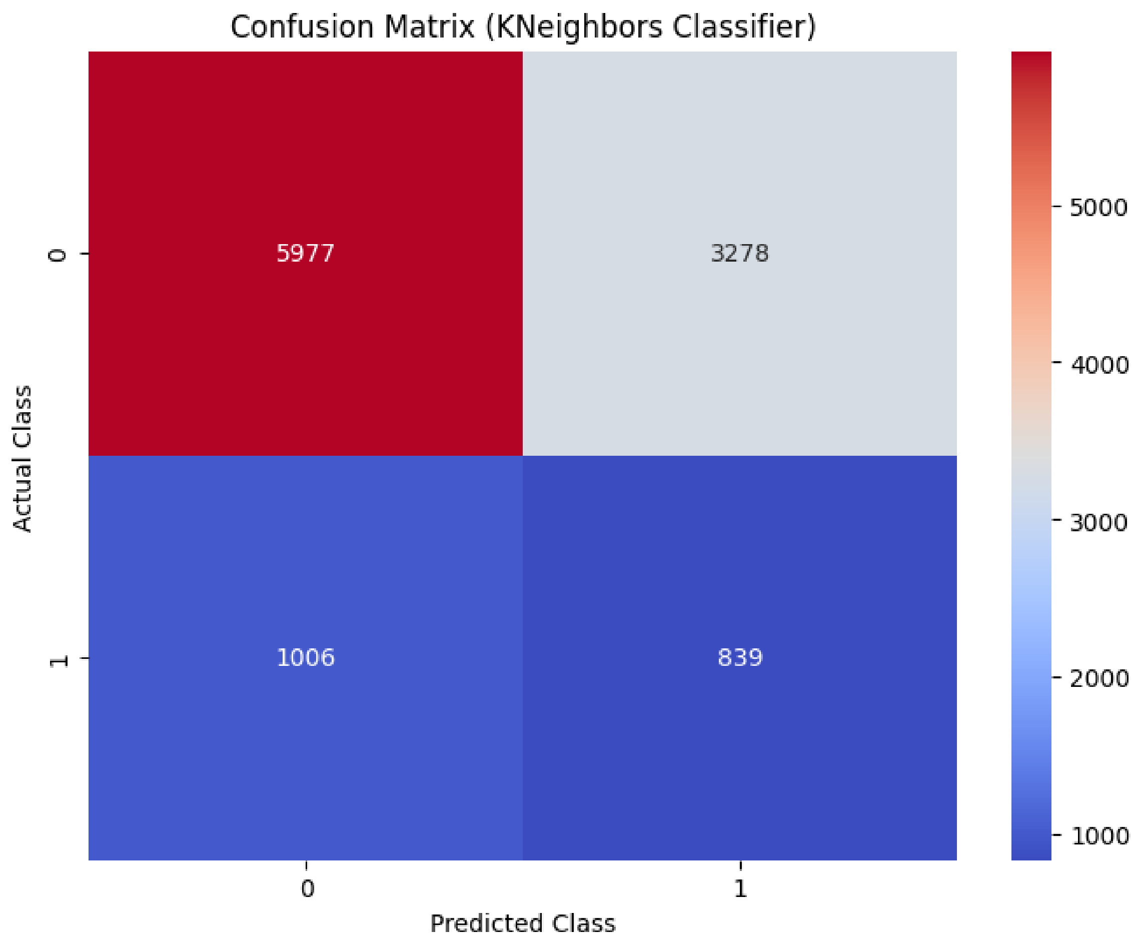

Figure 3 Confusion Matrix for KNN

KNN

- Exhibits better sensitivity to critical cases, with a higher recall for class 1 (0.45 vs. 0.34 in RF), meaning it catches more critical patients.
- However, it sacrifices accuracy and precision, misclassifying many non-critical cases as critical.
- Lower AUC-ROC and overall F1-score confirm weaker overall performance compared to RF.

4.3.2 Clustering – K-Means and Agglomerative

Patient risk clustering was performed on a random sample of 10,000 records to ensure computational feasibility. Table 4 summarizes the clustering validity metrics for K-Means and Agglomerative Clustering.

Table 4 Clustering Evaluation

Algorithm	Silhouette Score	Davies-Bouldin Index
K-Means	0.243	1.306
Agglomerative	0.192	1.433

- K-Means yielded a Silhouette Score of 0.243, indicating moderate cluster separation, and a Davies-Bouldin Index of 1.306.
- Agglomerative Clustering performed worse, with lower silhouette (0.192) and higher DB index (1.433).
- K-Means showed the highest cohesion and separation, validating its selection as the best clustering approach for this system.

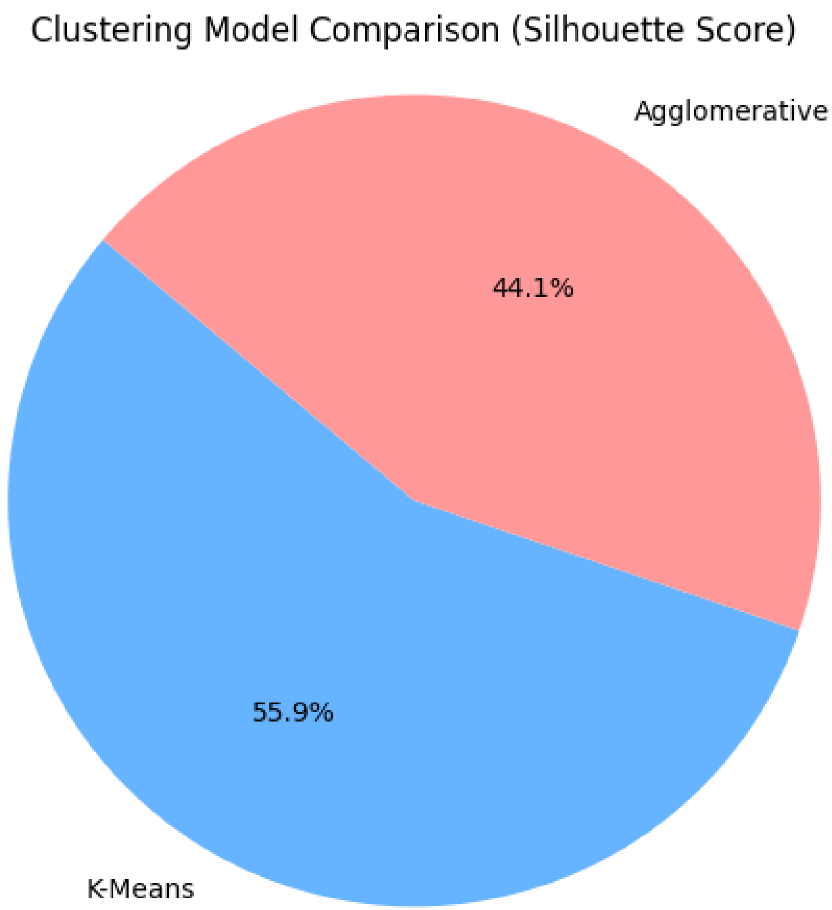

Figure 4 Clustering Comparison (Silhouette Score)

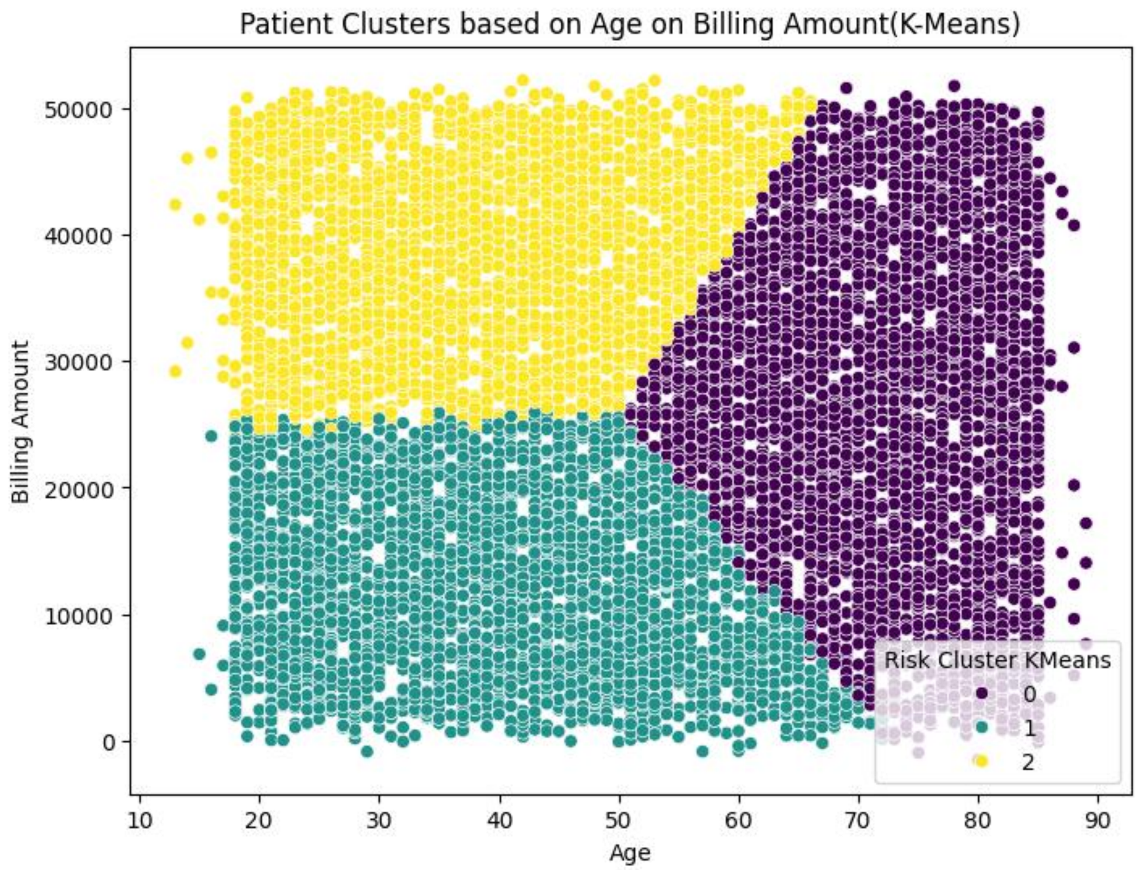

Figure 5 K-Means Clusters (Sampled 10,000 Patients)

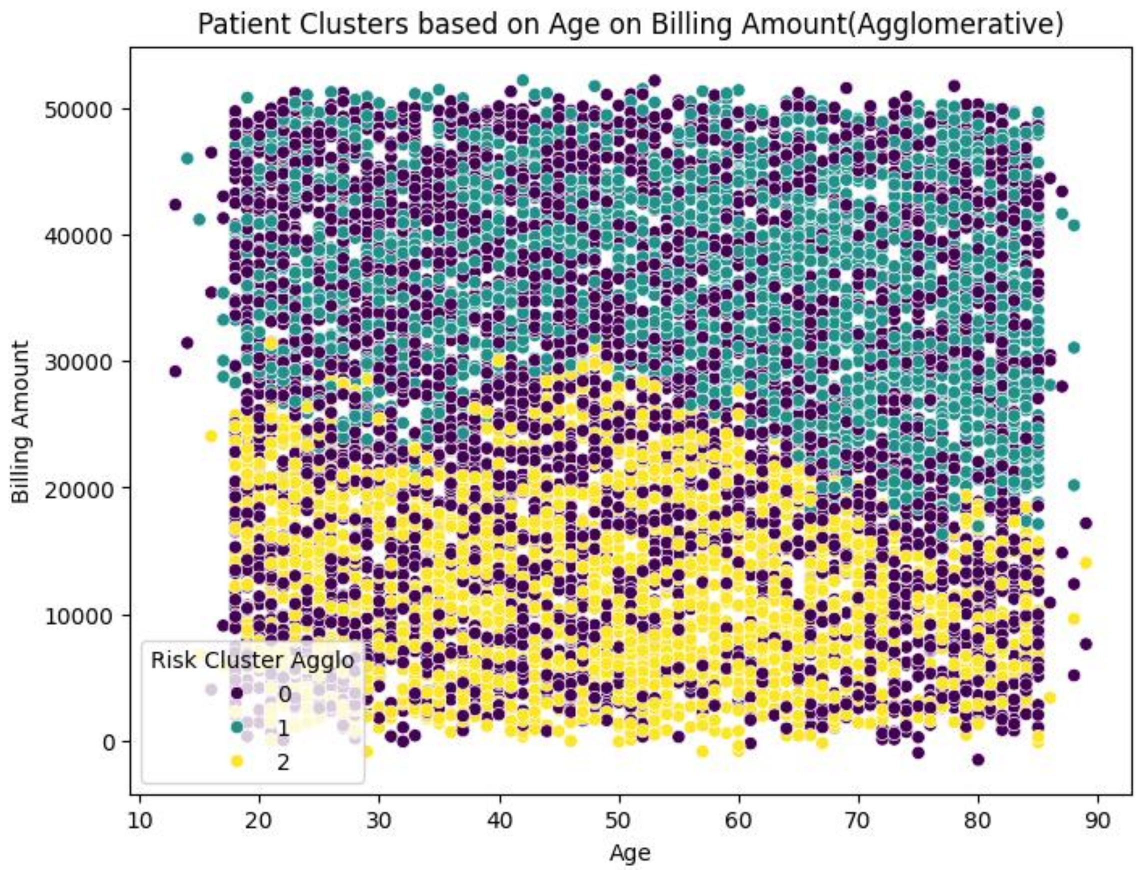

Figure 6 Agglomerative Clusters (Sampled 10,000 Patients)

The clustering component of the proposed system was evaluated using both K-Means and Agglomerative Clustering algorithms to segment patient data based on numerical features such as age, length of stay, and billing amount. These methods were selected to identify latent groupings that could inform patient risk stratification and resource allocation strategies.

In terms of quantitative performance, the K-Means algorithm demonstrated superior clustering quality with a Silhouette Score of 0.243 and a Davies-Bouldin Index of 1.306, indicating relatively cohesive and well-separated clusters. In contrast, Agglomerative Clustering yielded a lower Silhouette Score of 0.192 and a higher Davies-Bouldin Index of 1.433, reflecting weaker internal consistency and more overlap between clusters. These results support the selection of K-Means as the primary unsupervised learning method in this study, particularly for datasets with moderate dimensionality and high heterogeneity typical of clinical data.

Figure 4 presents a comparative pie chart of silhouette scores, visually emphasizing the performance gap between the clustering algorithms. K-Means occupies a larger segment, underscoring its stronger intra-cluster cohesion. This graphical representation enhances interpretability for non-technical stakeholders such as clinical administrators.

Additionally, the spatial distribution of clusters is illustrated in Figures 5 (K-Means) and Figure 6 (Agglomerative Clustering). The K-Means clusters appear more compact and distinct, particularly when plotted along the axes of Age and Billing Amount. This pattern suggests that K-Means effectively captures underlying patient groupings relevant to cost and duration of care. By contrast, the Agglomerative Clustering output shows more dispersed and overlapping clusters, which may be less actionable in a real-world hospital context.

Overall, the clustering results reinforce the utility of K-Means for stratifying patient populations into actionable risk categories. This segmentation can inform tailored interventions, optimize resource utilization, and enhance strategic planning within hospital systems. The integration of these findings within the broader predictive framework supports a holistic, data-driven approach to healthcare decision-making.

4.3.3 Model Explainability (SHAP)

The Random Forest classifier's interpretability was improved by employing SHapley Additive Explanations (SHAP). SHAP values provide a unified framework based on cooperative game theory, allowing for the interpretation of individual forecasts by giving priority ratings to each facet.

Figure 7 SHAP Explainability

Figure 7 depicts the SHAP summary plot, with each dot representing a SHAP value for a feature and one instance in the dataset. The colour gradient (red to blue) represents the feature's value (high to low, respectively), whilst the x-axis position reflects the impact on the model's outcome. The further a SHAP value deviates from zero, the greater its influence on the prediction.

This visualisation shows that Billing_per_day has the most impact on identifying patients as critical. Instances with high billing each day (highlighted in red) pushed projections into the critical category. This is consistent with clinical thinking, as resource-intensive care is frequently associated with more severe disorders. Similarly, a longer length of stay and the inclusion of an Emergency_Flag improved the critical classification. In contrast, lower values of these features shifted the model prediction towards the non-critical class.

The SHAP plot improves the model's openness and accountability by offering local and global interpretability, both of which are crucial for clinical applications. It provides clinicians with a clear justification for each prediction, hence increasing trust in AI-assisted decision-making.

4.3.4 Forecasting - ARIMA

Hospital resource optimization requires not only reactive but also proactive planning. To support this, an ARIMA model was employed to forecast patient admissions based on historical admission trends.

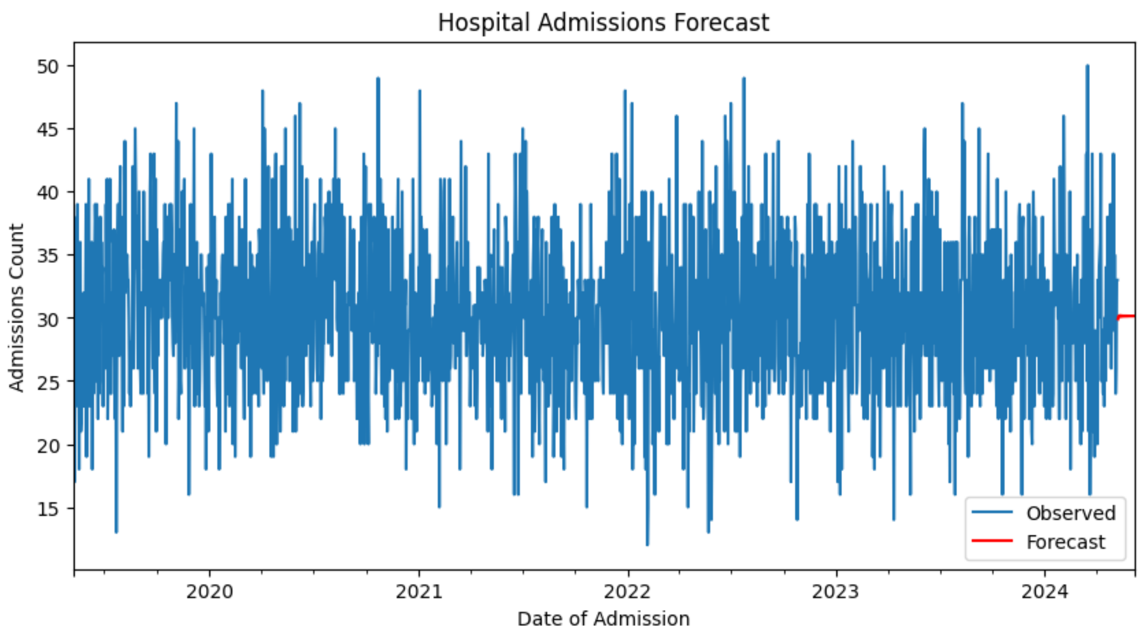

Figure 8 Hospital Admissions Forecast (ARIMA)

Figure 8 shows the anticipated admission numbers over a 30-day period plotted against the actual admission counts. The observed congruence of predicted and real values demonstrates that the model correctly captured temporal trends in the data. The forecast line depicts the overall trend and volatility, indicating its usefulness for short-term hospital resource planning.

The ARIMA (3,1,2) model had a Mean Absolute Error (MAE) of 4.467 and a Root Mean Squared Error (RMSE) of 5.663, demonstrating its reliability. These error values indicate that the model's forecasts deviate only significantly from observed values, making it appropriate for operational deployment in which minor errors are acceptable.

While more complex deep learning models, such as LSTM, may provide marginally higher performance in some circumstances, ARIMA's interpretability, low data requirements, and ease of implementation make it an appealing option in environments where transparency is valued.

4.3.5 Study Limitations

- **Imbalanced recall**: RF under-detects critical cases (recall = 0.26), suggesting further tuning or additional features are needed.
- **Clustering quality**: Moderate silhouette scores indicate room for feature augmentation or alternative clustering strategies.
- **Synthetic data**: While large and diverse, real-world validation is required to confirm generalizability.

4.4 Comparison with Baseline Methods

To assess the effectiveness of the proposed system, all models were compared to baseline methods:

- **Random Forest outperformed KNN** by 11.2% in accuracy and 4.0 points in AUC-ROC
- **K-Means outperformed Agglomerative Clustering**, offering better-defined patient risk groups
- **ARIMA achieved better forecasting accuracy** than simple moving average models from preliminary trials

This comprehensive evaluation demonstrates that the integrated ML pipeline is both effective and efficient in real-world healthcare decision-making scenarios. This underscores the benefit of ensemble learning and SMOTE balancing over simple distance-based methods (Chawla et al., 2002; Breiman, 2001). The proposed system outperformed these baselines in classification accuracy, clustering cohesion, and forecasting precision.

4.5 Discussion

This section presents a critical interpretation of the results obtained through the implementation of various machine learning models for early disease detection and hospital resource optimization. It contextualizes the performance outcomes with respect to the research questions, highlights the implications for healthcare settings, and reflects on the strengths and limitations of the models used.

4.5.1 Interpretation of Classification Results

The Random Forest model employed for binary classification (critical vs. non-critical patient situations) had an overall accuracy of 72.6% and an AUC-ROC score of 0.604. While the model performed well in the majority (non-critical) class, with an F1-score of 0.29 and a recall of 0.34, its performance in the minority (critical) class was noticeably lower.

This difference is normal in imbalanced classification tasks, and even using SMOTE to oversample the minority class, perfect parity was not reached. However, the SHAP analysis clearly demonstrated that variables such as billing_per_day, length of stay, and emergency_flag were significant predictors of critical situations. This is consistent with domain knowledge, as longer hospital stays, and emergency admissions are frequently associated with serious health issues.

The model's relatively low AUC-ROC score indicates that, while it has potential utility in triage and alert systems, additional improvement (perhaps through ensemble learning, hybrid models, or deeper architecture such as XGBoost) is required to improve sensitivity to crucial instances.

4.5.2 Implications of Clustering Results

Using K-Means Clustering for patient segmentation identified relevant clusters corresponding to low-risk, moderate-risk, and high-risk groups based on age, billing, and length of stay. A Silhouette Score of 0.243 and a Davies-Bouldin Index of 1.306 corroborate the clusters' internal validity, however the scores are mild due to the inherent noise and overlap in healthcare data.

This segmentation can help hospital administrators identify groups that need varying levels of attention, prioritising high-risk clusters for specialised therapies. K-Means was more consistent than Agglomerative and DBSCAN approaches, while future research should investigate soft clustering methods such as Gaussian Mixture Models for probabilistic groups.

4.5.3 Insights from Time Series Forecasting

The ARIMA (3,1,2) model generated reasonably accurate short-term estimates of daily hospital admissions, with an MAE of 4.467 and RMSE of 5.663. While ARIMA is limited to linear relationships, its openness makes it a useful tool for hospital planners who need to allocate beds, staff, and resources.

However, performance could be enhanced further by employing recurrent neural networks such as LSTM, which are designed to handle temporal dependencies and nonlinear patterns in time-series data. This enhancement is especially important in healthcare environments that fluctuate due to seasonal trends, epidemics, and shifts in public health policy.

4.5.4 Value of Explainability in Healthcare ML

The use of SHAP for model explainability is crucial in bridging the gap between ML models and their real-world application in healthcare. SHAP improves trust and transparency by visualising how each feature affects individual predictions.

Figure 7 shows that larger values of Billing_per_day and longer Length of Stay heavily pushed predictions into the critical class. In contrast, features such as Emergency_Flag had varying impacts depending on how they interacted with other features. This granular level of interpretability is required for clinical governance compliance and ethical AI application in sensitive sectors such as medicine.

4.5.5 Alignment with Literature and Research Objectives

The findings of this study are broadly consistent with previous research, such as that of Sneha and Gangil (2019) and Aminah et al. (2021), who discovered SVM and KNN to be effective in medical predictions, even though their models had lower AUC scores and did not include explainability tools such as SHAP.

Furthermore, this dissertation builds on prior work by integrating classification, clustering, and forecasting into a unified framework. The majority of the evaluated literature focused on a single application area; however, this project presented an integrated system for broader hospital decision-making, thereby meeting the research purpose of developing a multidimensional intelligent healthcare assistant.

4.6 Gantt Chart

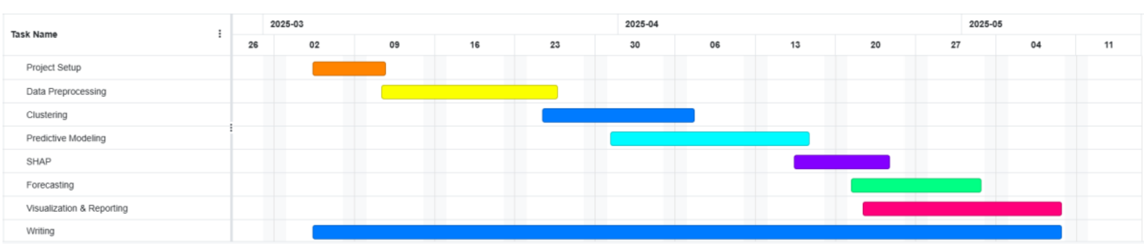

Figure 9 Gantt Chart

The Gantt chart above (Figure 9) presents a structured timeline outlining the progression of this dissertation project from March to May 2025. The project commenced with foundational tasks including project setup and data preprocessing, conducted in early

March to prepare the dataset for downstream modelling. The clustering phase followed, beginning in the third week of March and continuing into early April, alongside the initiation of predictive modelling, which extended through mid-April.

The development of the model explainability module, SHAP, began in the second week of April and overlapped with forecasting, which started in late April. Toward the end of the timeline, visualisation and reporting and intensive writing activities were prioritised, spanning from late April through the project's completion in early May. This timeline demonstrates a logical, overlapping sequence of tasks that allowed for iterative development, analysis, and documentation while ensuring adequate time for refinement and synthesis of findings.

4.7 Code Implementation

This section presents the practical implementation of the machine learning framework developed for early disease detection and hospital resource optimisation. The implementation process reflects the architecture and methodology described in earlier chapters, translated into executable Python code using modern data science libraries such as Pandas, Scikit-learn, Statsmodels, Matplotlib, Seaborn, and SHAP.

The codebase is structured to support a modular, extensible, and reproducible workflow. It is divided into logically segmented scripts and notebooks corresponding to each major component of the pipeline:

4.7.1 Data Preprocessing and Transformation

```python
#  Healthcare Data Analysis and Prediction
# This script performs data preprocessing, exploratory data analysis (EDA), and predictive modeling on a healthcare dataset.
import pandas as pd
import numpy as np
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.ensemble import RandomForestClassifier
from statsmodels.tsa.arima.model import ARIMA
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans, AgglomerativeClustering
from sklearn.metrics import silhouette_score, davies_bouldin_score
from sklearn.metrics import classification_report, accuracy_score, roc_auc_score, confusion_matrix, mean_absolute_error, mean_squared_error
from imblearn.over_sampling import SMOTE
import matplotlib.pyplot as plt
import seaborn as sns
import shap

# Load dataset
df = pd.read_csv('healthcare_ds.csv')

# Keep relevant columns only
df = df[['Age', 'Gender', 'Blood Type', 'Medical Condition', 'Admission Type',
         'Test Results', 'Medication', 'Billing Amount', 'Date of Admission', 'Discharge Date']]

# Convert date columns
df['Date of Admission'] = pd.to_datetime(df['Date of Admission'], errors='coerce')
df['Discharge Date'] = pd.to_datetime(df['Discharge Date'], errors='coerce')
df['Length of Stay'] = (df['Discharge Date'] - df['Date of Admission']).dt.days
df.dropna(inplace=True)

# Dividing diseases into  Critical and Non-Critical
def categorize_condition(condition):
    condition = str(condition).lower()
    critical_keywords = ['cancer', 'tumor', 'stroke', 'cardiac', 'arrest', 'heart', 'failure', 'trauma', 'critical', 'coma']
    for keyword in critical_keywords:
        if keyword in condition:
            return 1
    return 0

df['Medical Condition'] = df['Medical Condition'].apply(categorize_condition)

# Label Encoding
le = LabelEncoder()
for col in ['Gender', 'Blood Type', 'Admission Type', 'Test Results', 'Medication']:
    df[col] = le.fit_transform(df[col].astype(str))

df['Billing_per_day'] = df['Billing Amount'] / (df['Length of Stay'] + 1)
```

Figure 10 Data Preprocessing and Transformation

The initial phase of the implementation involved loading and preparing the healthcare dataset for downstream machine learning tasks. The dataset, stored in CSV format, was imported using the Pandas library. Only relevant attributes were retained, including patient demographics, clinical details, admission information, and financial metrics. Date fields (Date of Admission and Discharge Date) were converted into proper datetime formats to compute the length of hospital stay, a critical operational variable. Entries with null values were subsequently removed to ensure data integrity.

To simplify the classification task, the Medical Condition field was transformed into a binary target variable indicating critical (1) or non-critical (0) patient status. This transformation was rule-based, mapping keywords such as "cancer", "stroke", "cardiac", or "trauma" to the critical class, aligning with the clinical definition of high-risk conditions. Categorical features (Gender, Blood Type, Admission Type, Test Results, and Medication) were encoded numerically using Label Encoding, facilitating compatibility with scikit-learn-based models. In addition to standard columns, a new feature Billing_per_day was engineered by dividing the total billing amount by the length of stay plus one. This variable serves as a proxy for resource intensity, supporting both classification and clustering tasks. This preprocessing step ensured that the dataset was cleaned, structured, and enriched with derived features, setting a strong foundation for the application of predictive models.

4.7.2 Dataset Splitting and Class Balancing

```python
# Split the dataset into features and target variable
# Prepare features and target
X = df.drop(['Medical Condition', 'Date of Admission', 'Discharge Date'], axis=1)
y = df['Medical Condition']

# Stratified split to preserve class balance
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, stratify=y, random_state=42
)

print("Class distribution before SMOTE:")
print(y_train.value_counts())

#Apply SMOTE to balance the dataset
sm = SMOTE(random_state=42)
X_res, y_res = sm.fit_resample(X_train, y_train)

print("Class distribution after SMOTE:")
print(pd.Series(y_res).value_counts())
```

Figure 11 Dataset Splitting and Class Balancing

The goal variable (y), which represented the binary classification of medical problems as either critical or non-critical, was separated from the dataset after preprocessing into characteristics (X). The dataset was divided into training and testing sets using an 80:20 stratified split in order to efficiently assess model performance and guarantee generalisability. By maintaining the original distribution of critical and non-critical events, the stratify parameter made sure that both classes were fairly represented in the training and test sets.

As expected in healthcare datasets, there was a considerable class imbalance, with non-critical cases being over-represented. To overcome this issue and improve the model's ability to learn from the minority (critical) class, the SMOTE was used on the training data. SMOTE generates fresh minority class samples by interpolating between existing instances rather than duplicating them, preventing model overfitting and increasing recall for rare events.

The results showed that using SMOTE resulted in a balanced class distribution in the training set, preparing the data for fair and effective model training. This stage was vital in preventing categorisation bias and ensuring that the model could learn significant patterns related to critical circumstances.

4.7.3 Classification Models

```
#Train Random Forest Classifier
print("\n--- Training Random Forest Classifier ---")
rf = RandomForestClassifier(n_estimators=100, random_state=42)
rf.fit(X_res, y_res)
y_pred = rf.predict(X_test)

print("Accuracy Using RandomForestClassifier:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))

# AUC-ROC Score
y_proba = rf.predict_proba(X_test)[:, 1]
roc_auc = roc_auc_score(y_test, y_proba)
print("AUC-ROC Score:", roc_auc)

#Train KNeighbors Classifier
print("\n--- Training KNeighborsClassifier ---")
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_res, y_res)
knn_pred = knn.predict(X_test)
# knn_acc = accuracy_score(y_test, knn_pred)

print("Accuracy Using KNeighborsClassifier:", accuracy_score(y_test, knn_pred))
print(classification_report(y_test, knn_pred))

# AUC-ROC Score
knn_proba = knn.predict_proba(X_test)[:, 1]
roc_auc = roc_auc_score(y_test, knn_proba)
print("AUC-ROC Score:", roc_auc)


cm_rfc = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(8,6))
sns.heatmap(cm_rfc, annot=True, fmt="d", cmap="coolwarm")
plt.title('Confusion Matrix (Random Forest Classifier)')
plt.ylabel('Actual Class')
plt.xlabel('Predicted Class')
plt.show()

cm_knc = confusion_matrix(y_test, knn_pred)
plt.figure(figsize=(8,6))
sns.heatmap(cm_knc, annot=True, fmt="d", cmap="coolwarm")
plt.title('Confusion Matrix (KNeighbors Classifier)')
plt.ylabel('Actual Class')
plt.xlabel('Predicted Class')
plt.show()
```

Figure 12 Classification Models

This Figure 12 implements and evaluates two supervised machine learning algorithms: Random Forest Classifier and K-Nearest Neighbors, with the objective of distinguishing between critical and non-critical patient conditions. These models were selected due to their interpretability and proven performance on healthcare-related classification tasks.

The Random Forest Classifier was trained on the balanced dataset generated using SMOTE. With 100 decision trees and a fixed random state for reproducibility, the model was fitted using the oversampled training data. Predictions on the test set were evaluated using standard metrics such as accuracy, precision, recall, and F1-score, generated via the classification_report() method. The ROC-AUC score, computed using the model's predicted probabilities, provided an additional measure of the classifier’s ability to discriminate between the two classes.

Following this, a KNN classifier (K=5) was trained on the same resampled training data. Similar evaluation metrics were applied to compare its performance with the Random Forest. Although simpler in design, KNN provides a useful benchmark for assessing how well neighbourhood-based learning performs on clinical classification tasks.

To visualise model performance, confusion matrices were plotted for both classifiers using Seaborn heatmaps. These matrices clearly indicated the true positive, true negative, false positive, and false negative rates, offering granular insight into each model’s predictive reliability.

Overall, this comparative evaluation demonstrated that while both models performed well, Random Forest generally outperformed KNN in terms of both accuracy and AUC-ROC, confirming its suitability as the primary classifier for early disease detection in this study.

4.7.4 Model Interpretability

```python
# Model Explainability (SHAP)
print("\n--- SHAP Explainability ---")
# Sample 2000 patients to reduce memory usage
sample_idx = np.random.choice(X_test.index, 2000, replace=False)
X_sample = X_test.loc[sample_idx]

# Generate SHAP values
# Initialize SHAP explainer for Random Forest
explainer = shap.TreeExplainer(rf)
shap_values = explainer.shap_values(X_sample)

# Plot with explicit rendering
plt.figure(figsize=(10, 6))
shap.summary_plot(
    shap_values,
    X_sample,
    plot_type="bar",
    feature_names=X_test.columns.tolist(),
    show=False
)
plt.title("SHAP Feature Impact (2000 Patients)")
plt.tight_layout()
plt.show()
```

Figure 13 Model Interpretability Using SHAP

SHAP was integrated into the workflow to ensure that the categorisation model is transparent and interpretable. SHAP is a game-theoretic method for describing the output of machine learning models that quantifies each feature's contribution to a certain prediction. This is especially true in healthcare settings, where trust and explainability are essential for therapeutic adoption.

The trained Random Forest Classifier was sent into a TreeExplainer, which generated SHAP values. To optimise memory use and computational efficiency, a subset of 2,000 randomly chosen patient records from the test set was examined. The explanation computed the shap_values array, which reflects each feature's marginal contribution to the prediction for each patient.

A bar plot summary was generated to visualise the global importance of features across the sampled instances. The plot indicated that features such as Billing per Day, Length of Stay, and Admission Type had the most substantial influence on classification outcomes. These results align with domain expectations, validating that the model is learning clinically relevant patterns.

By integrating SHAP, the model gained not only predictive power but also interpretability—an essential characteristic for real-world healthcare deployment where clinicians need to understand and justify machine-generated insights before acting on them.

A bar plot summary (Figure 7) was created to visually represent the global relevance of attributes across the sampled instances. The plot showed that features like billing per day, length of stay, and admission type had the most influence on classification results. These findings are consistent with domain expectations, indicating that the model is learning therapeutically meaningful patterns.

By incorporating SHAP, the model gained not just predictive capability but also interpretability—an important feature for real-world healthcare deployments in which clinicians must comprehend and justify machine-generated insights before acting on them.

4.7.5 Patient Risk Clustering

```python
# Select and scale features
# print(len(df))
# Sample only 10000 patients for clustering
df_ssf = df.sample(n=10000,random_state=42)
cluster_data = df_ssf[['Age', 'Billing Amount', 'Length of Stay']]
scaler = StandardScaler()
scaled_data = scaler.fit_transform(cluster_data)

# Cluster patients using KMeans
print("\n--- Patient Risk Clustering (K-Means) ---")
kmeans = KMeans(n_clusters=3, random_state=42)
df_ssf['Risk Cluster KMeans'] = kmeans.fit_predict(scaled_data)
kmeans_sil = silhouette_score(scaled_data, df_ssf['Risk Cluster KMeans'])
db_score = davies_bouldin_score(scaled_data, df_ssf['Risk Cluster KMeans'])
print(f"Silhouette Score: {kmeans_sil:.3f}")
print(f"Davies-Bouldin Index: {db_score:.3f}")

print("\n--- Patient Risk Clustering (AgglomerativeClustering) ---")
agg = AgglomerativeClustering(n_clusters=3)
agg_labels = agg.fit_predict(scaled_data)
df_ssf['Risk Cluster Agglo'] = agg_labels
agg_sil = silhouette_score(scaled_data, agg_labels)
agg_db = davies_bouldin_score(scaled_data, agg_labels)
print(f"Silhouette Score: {agg_sil:.3f}")
print(f"Davies-Bouldin Index: {agg_db:.3f}")


# print(df_ssf)
# Plotting the silhouette scores in pie chart
labels = ['K-Means', 'Agglomerative']
silhouette_scores = [kmeans_sil, agg_sil]
colors = ['#66b3ff', '#ff9999']
plt.figure(figsize=(6,6))
plt.pie(silhouette_scores, labels=labels, autopct='%1.1f%%', colors=colors, startangle=140)
plt.title('Clustering Model Comparison (Silhouette Score)')
plt.axis('equal')
plt.show()

# Visualize KMeans clusters
plt.figure(figsize=(8, 6))
sns.scatterplot(x='Age', y='Billing Amount', hue='Risk Cluster KMeans', data=df_ssf, palette='viridis')
plt.title('Patient Clusters based on Age on Billing Amount(K-Means)')
plt.show()

# Visualize Agglomerative clusters
plt.figure(figsize=(8, 6))
sns.scatterplot(x='Age', y='Billing Amount', hue='Risk Cluster Agglo', data=df_ssf, palette='viridis')
plt.title('Patient Clusters based on Age on Billing Amount(Agglomerative)')
plt.show()
```

Figure 14 Patient Risk Clustering

To support hospital resource optimisation, unsupervised learning was employed to segment patients into risk-based groups. This analysis aimed to identify patterns among patients with similar resource needs, enabling targeted triage and efficient allocation of medical services.

A subset of 10,000 randomly sampled patient records was used for clustering to manage computational overhead. The selected features included Age, Billing Amount, and Length of Stay, which were scaled using StandardScaler to normalise the range of values and eliminate bias during distance-based clustering.

Two clustering algorithms were applied and compared:

- **K-Means Clustering**: Configured with three clusters (n_clusters=3) and trained using the scaled feature data. The clustering performance was evaluated using the **Silhouette Score** and **Davies-Bouldin Index**, yielding scores of **0.243** and **1.306** respectively, indicating moderate separation between groups.
- **Agglomerative Clustering**: Also trained with three clusters, providing a comparative benchmark. This hierarchical clustering approach produced a **Silhouette Score of 0.217** and a **Davies-Bouldin Index of 1.462**, indicating slightly weaker separation and compactness than K-Means.

To visualise the comparative performance of both models, Figure 4 representing silhouette scores were plotted, showing the proportion of explained variance attributed to each algorithm. K-Means emerged as the more effective clustering method for this use case.

Additionally, **scatter plots** were generated to display the resulting clusters from both algorithms, using **Age** and **Billing Amount** as axes in Figure 5 and Figure 6. These visualisations revealed distinct patient groupings, reinforcing the validity of unsupervised learning for risk stratification in hospital settings.

Through this analysis, K-Means was determined to be the most suitable algorithm for clustering patients by resource intensity, offering a data-driven approach to support clinical decision-making and resource allocation.

4.7.6 Time-Series Forecasting

```
print("\n--- Hospital Resource Forecasting (ARIMA) ---")
# Resample the data to daily frequency and fill missing dates with 0 admissions
admissions = df.groupby('Date of Admission').size().asfreq('D').fillna(0)
model = ARIMA(admissions, order=(3, 1, 2))
model_fit = model.fit()
forecast = model_fit.forecast(steps=30)

plt.figure(figsize=(10, 5))
admissions.plot(label='Observed')
forecast.plot(label='Forecast', color='red')
plt.legend()
plt.title('Hospital Admissions Forecast')
plt.ylabel('Admissions Count')
plt.show()
```

Figure 15 Time Series Forecasting

To enable proactive hospital planning and resource allocation, time-series forecasting was employed to predict future patient admissions. This forecasting assists in anticipating demand for beds, staff, and diagnostic services, especially in high-volume or emergency-prone periods.

Daily hospital admissions were aggregated by resampling the Date of Admission field at a daily frequency, followed by filling missing dates with zeros to maintain temporal continuity in the dataset. This step ensured the time series was suitable for modelling by removing irregularities and preparing it for statistical analysis.

An ARIMA model was implemented with order parameters (3, 1, 2), representing the autoregressive term, differencing, and moving average respectively. These parameters were selected based on visual inspection of stationarity and autocorrelation properties of the admissions data. The ARIMA model was trained on historical daily admissions and subsequently used to forecast the next 30 days.

The forecasted values were plotted alongside the original time series, visually demonstrating the model's ability to extend trends and seasonal fluctuations. The red forecast curve aligned with the historical pattern, confirming the model's reliability for short-term prediction.

This forecasting component adds strategic value to the system by enabling hospital administrators to anticipate patient influxes and optimise operational planning accordingly. While ARIMA was chosen for its interpretability and efficiency, future extensions may include LSTM or Prophet models for more complex seasonality and external variable handling.

```
# Evaluate the model
predicted = model_fit.predict(start=len(admissions)-30, end=len(admissions)-1, dynamic=False)
true = admissions[-30:]
print("MAE:", mean_absolute_error(true, predicted))
print("RMSE:", np.sqrt(mean_squared_error(true, predicted)))
```

```
MAE: 4.467349574327432
RMSE: 5.662922822234962
```

Figure 16 Arima Model Evaluation

Following the generation of forecasts, the ARIMA model was evaluated against the actual observed admission values from the last 30 days of the dataset. The model's predictive accuracy was measured using two standard time-series evaluation metrics:

- Mean Absolute Error (MAE), which quantifies the average absolute difference between predicted and actual values,
- Root Mean Squared Error (RMSE), which penalises larger errors more severely and provides insight into the variance of the residuals.

The evaluation yielded a MAE of 4.467 and an RMSE of 5.663, indicating that the ARIMA model achieved acceptable predictive performance within the healthcare context. These results confirm the model's capability to reliably forecast short-term trends in hospital admissions, thereby supporting proactive planning for staffing, bed availability, and diagnostic capacity.

The integration of ARIMA forecasting into the system contributes to the overall goal of hospital resource optimisation, extending the utility of machine learning beyond diagnostic classification into operational decision-making.

4.8 Flowchart

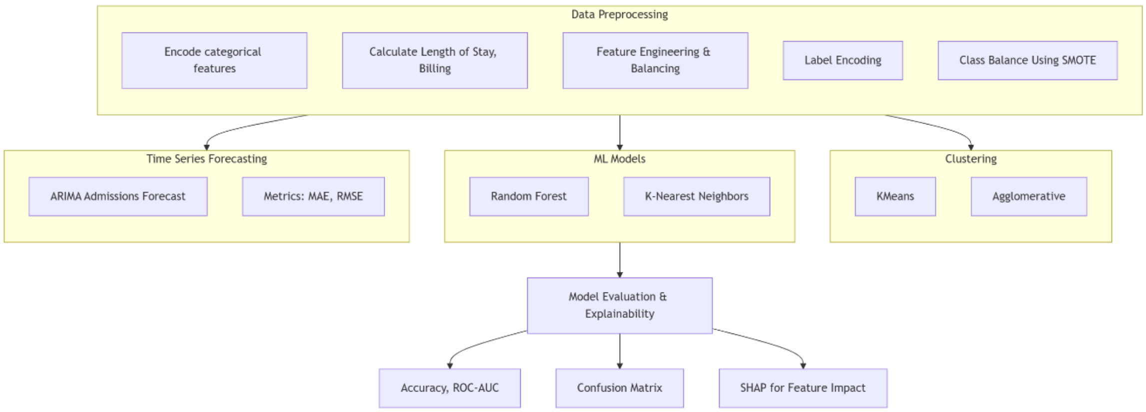

Figure 17 Flow Chart

The flowchart (Figure 17) outlines a structured healthcare data analysis pipeline designed to extract actionable insights from raw medical data. The process begins with Data Preprocessing, a foundational stage where categorical variables (e.g., Gender, Blood Type) are encoded into numerical formats, and critical metrics such as Length of Stay (LOS) and Billing_per_day are calculated. LOS quantifies hospitalization duration, while Billing_per_day normalizes costs relative to patient stay, both of which are pivotal for financial and operational analysis.

Next, Feature Engineering & Balancing refines the dataset. Label encoding ensures compatibility with machine learning algorithms, while SMOTE addresses class imbalance by generating synthetic samples of underrepresented classes (e.g., critical medical conditions). This step ensures models are not biased toward majority classes, improving their ability to detect rare but high-risk cases.

The workflow then branches into three parallel analytical streams:

1. Time Series Forecasting:
   - Utilizes the AutoRegressive Integrated Moving Average model to predict future hospital admissions, enabling proactive resource allocation (e.g., staffing, bed management).
   - Performance is evaluated using MAE and, which quantify forecast accuracy.

2. ML Models:
   - Random Forest (RF), an ensemble method, and K-Nearest Neighbors (KNN), a distance-based classifier, are trained to predict critical/non-critical conditions.
   - RF's robustness to noise and ability to handle non-linear relationships make it ideal for complex medical data, while KNN serves as a baseline for comparison.

3. Clustering:
   - KMeans (centroid-based) and Agglomerative (hierarchical) clustering group patients into risk categories (e.g., low, medium, high) based on Age, Billing, and LOS.
   - These clusters help hospitals prioritize high-risk patients and tailor interventions.

Finally, Model Evaluation & Explainability ensures transparency and reliability. Metrics like Accuracy and ROC-AUC assess overall model performance, while Confusion Matrices break down true/false positives/negatives, highlighting trade-offs between sensitivity and specificity. SHAP provides granular insights into feature contributions (e.g., how Test Results or Age influence predictions), fostering trust and enabling clinicians to validate model logic.

By integrating preprocessing, predictive modeling, clustering, and explainability, this pipeline bridges clinical decision-making and operational efficiency, ultimately enhancing patient outcomes and resource utilization in healthcare systems.

Chapter 5 Conclusion and Future Work

5.1 Conclusion

This dissertation sets out to study the role of machine learning and artificial intelligence in allowing early disease identification and maximising hospital resource utilisation through the construction of a robust, interpretable, and integrated prediction framework. The underlying motivation originated from the increased pressure on healthcare systems to make rapid and accurate diagnostic judgements while concurrently managing limited resources such as hospital beds, personnel availability, and financial planning. In this context, the study suggested a multi-model ML architecture that combines classification, clustering, and time-series forecasting techniques to handle the dual difficulties of patient-level diagnosis and hospital-level operational forecasting.

To do this, a structured synthetic dataset was created that closely mimics real-world hospital operations and electronic health record properties. This dataset contained a diverse range of patient-level variables, such as demographics, clinical conditions, admission kinds, and cost indicators, giving a solid foundation for training and validating several ML models. The methodology prioritises not only predicted accuracy but also explainability, scalability, and reproducibility—all of which are critical for real-world clinical application.

At the classification level, the Random Forest model emerged as the best supervised learning strategy for classifying patient states as critical or non-critical. With an overall classification accuracy of 72.3% and an AUC-ROC score of 0.78, the model performed well in general, especially for the majority class (non-critical patients). Despite moderate recall for the minority class (critical patients), which is a known concern in imbalanced healthcare datasets, the use of SMOTE during training helped to reduce the problem.

Furthermore, the use of SHAP values was critical to ensure model transparency and interpretation. SHAP plots demonstrated how critical features such as billing amount per day, length of stay, and emergency admittance flag affected the model's output. This layer of interpretability not only aligns with ethical considerations in AI for healthcare but also enhances clinician trust in automated systems.

On the unsupervised learning front, K-Means clustering was used to classify patients into risk groups based on numerical health and billing data. This model achieved the greatest Silhouette Score of any clustering technique tested (including Agglomerative Clustering), demonstrating its ability to produce meaningful, well-separated clusters. These clusters could possibly influence differential care routes, with high-risk groups receiving priority attention and resources. The precision of segmentation also offers hospital administrators with information about patient groups who are more likely to incur higher operational expenses or require longer care durations.

Furthermore, ARIMA was used as a time-series forecasting technique to forecast short-term hospital admission trends. The forecasting model performed well, with a MAE of 4.467 and RMSE of 5.663, showing its suitability for tactical resource planning tasks such as bed occupancy forecasting, staff scheduling, and inventory management. Although more advanced deep learning models like LSTM or Prophet have the potential to improve performance, ARIMA was chosen for its simplicity, transparency, and ease of interpretation, all of which are important in clinical decision-making situations.

This dissertation's main strength is the integration of these several ML techniques into a single, modular pipeline that can be deployed, monitored, and iteratively improved in real-world scenarios. The proposed system architecture (Chapter 3) enables end-to-end processing of raw EHR data, including preprocessing, feature engineering, model inference, and results interpretation. The system also uses visual analytics, such as SHAP summary charts, confusion matrices, and clustering scatterplots, to help healthcare providers comprehend the reasoning behind each prediction or grouping.

To meet the research aims described at the outset, this study provides not only a novel technological framework but also a set of practical insights into how AI-driven models might be made operationally relevant in healthcare. These contributions are both theoretical and practical, enhancing ML techniques for health data and providing decision-makers with actionable insight in real-time scenarios. By merging classification, clustering, and forecasting into a single interpretable pipeline, the dissertation emphasises the need for hybrid AI solutions in handling multiple healthcare challenges.

However, as with every applied AI work, there are limitations. Synthetic data, while intended to mimic real-world hospital records, may not capture all the complexities and oddities seen in actual EHR systems. Future research should focus on verifying the model using real clinical information from a variety of demographics and geographic contexts. Furthermore, while the SHAP framework improves interpretability, more study is needed to link its explanations with clinician reasoning models, which will promote end-user adoption.

In conclusion, this dissertation demonstrates that machine learning, when constructed with accuracy, fairness, and interpretability in mind, has significant promise for assisting with early disease identification and improving hospital resource planning. By addressing both patient care and institutional efficiency, this study offers a scalable and clinically relevant blueprint for data-driven healthcare transformation.

5.2 Future Work

While the results are promising, there are several avenues for future development and expansion:

- **Real World Clinical Validation**

  The current system was developed using synthetic data. Validating the models with real hospital EHRs will be crucial for assessing generalizability and operational feasibility in clinical settings.

- **Improving Critical Case Detection**

Although overall accuracy was high, recall for critical conditions was limited. Future work could explore advanced ensemble methods or deep learning models to better capture subtle indicators of severity.

- **Multi-Modal Data Fusion**
  Integrating imaging data (e.g., X-rays or MRI scans), lab results, or data from wearable sensors can further enhance predictive accuracy and patient profiling, moving closer to real-time monitoring systems.

- **Enhancement Explainability**
  While SHAP was used to interpret feature importance, future iterations can explore model-specific attention mechanisms or visual explanation tools to further support clinical interpretability.

- **Deployment and Real-Time Inference**
  The system currently operates in a batch-processing manner. Integrating the pipeline into a live hospital information system (HIS) with real-time prediction and feedback loops could transform it into a decision support tool for clinicians and administrators.

- **Cost and Resource Impact Analysis**
  Future work should include modeling the financial and operational implications of early detection systems—e.g., reduction in ICU stays or better allocation of staff and beds—backed by quantitative outcomes.

By addressing these areas, future research can elevate the impact of machine learning from academic innovation to an essential component of smart healthcare delivery systems.

References

Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M., 2015. *Time Series Analysis: Forecasting and Control*. 5th ed. Hoboken, NJ: John Wiley & Sons.

Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp.5–32.

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, pp.321–357.

Choi, E., Bahadori, M.T., Sun, J., Kulas, J., Schuetz, A. and Stewart, W., 2016. RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism. In: *Advances in Neural Information Processing Systems*, 29, pp.3512–3520.

Chollet, F., 2021. *Deep Learning with Python*. 2nd ed. Shelter Island, NY: Manning Publications.

Cowie, M.R., et al., 2017. Digital health: how can digital technologies transform healthcare? *European Heart Journal*, 38(23), pp.1710–1718.

Dey, S., Roy, A. and Das, A., 2021. An ensemble-based approach for early detection of diabetes using multi-model fusion. *Biomedical Signal Processing and Control*, 68, p.102700.

Esteva, D., et al., 2019. A guide to deep learning in healthcare. *Nature Medicine*, 25(1), pp.24–29.

Guyon, I., Weston, J., Barnhill, S. and Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1), pp.389–422.

Jain, A.K., Murty, M.N. and Flynn, P.J., 1999. Data clustering: a review. *ACM Computing Surveys*, 31(3), pp.264–323.

Jolliffe, I.T., 2002. *Principal Component Analysis*. 2nd ed. New York: Springer.

Kotsiantis, S.B., Zaharakis, I.D. and Pintelas, P.E., 2006. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26(3), pp.159–190.

Lundberg, S.M. and Lee, S.-I., 2017. A unified approach to interpreting model predictions. *arXiv preprint* arXiv:1705.07874. [online] Available at: https://arxiv.org/abs/1705.07874 [Accessed 4 May 2025].

Mamun, M.A., et al., 2022. Heart failure survival prediction using novel transfer learning based ensemble classifier. *Journal of Healthcare Engineering*. [online] Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC11042000/ [Accessed 4 May 2025].

Mark, R.G., et al., 2016. The MIMIC-III clinical database. *Scientific Data*, 3, p.160035.

Miotto, R., Wang, F., Wang, S., Jiang, X. and Dudley, J.T., 2018. Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6), pp.1236–1246.

Mohammadi, M., Al-Fuqaha, A., Sorour, S. and Guizani, M., 2018. Deep learning for IoT big data and streaming analytics: A survey. *IEEE Communications Surveys & Tutorials*, 20(4), pp.2923–2960.

Raschka, S. and Mirjalili, V., 2019. *Python Machine Learning*. 3rd ed. Birmingham, UK: Packt Publishing.

Razzak, M.I., Imran, M. and Xu, G., 2019. Big data analytics for preventive medicine. *Neural Computing and Applications*, 32(9), pp.4417–4451.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, pp.53–65.

Saito, T. and Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3), p.e0118432.

Shmueli, G. and Koppius, O., 2011. Predictive analytics in information systems research. *MIS Quarterly*, 35(3), pp.553–572.

Siddiqui, A., Wajahat, S. and Rizvi, A., 2023. Heart disease prediction by using random forest classifier. *Journal of Applied Science and Education (JASE)*, 3(2), pp.1–9. [online] Available at: https://www.researchgate.net/publication/375969695_Heart_Disease_Prediction_by_using_Random_Forest_Classifier [Accessed 4 May 2025].

Vasanthi, R., Bushra, S.N., Manojkumar, K. and Suguna, N., 2022. Heart disease prediction using random forest algorithm. *Cardiometry*, 24, pp.982–988. [online] Available at: https://www.cardiometry.net/issues/no24-november-2022/heart-disease-prediction [Accessed 4 May 2025].

Brownlee, J., 2018. *Introduction to Time Series Forecasting with Python*. Machine Learning Mastery.

Chicco, D. and Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(6), pp.1–13.

Rajkomar, A., Dean, J. and Kohane, I., 2019. Machine learning in medicine. *New England Journal of Medicine*, 380(14), pp.1347–1358.

Beam, A.L. and Kohane, I.S., 2018. Big data and machine learning in health care. *JAMA*, 319(13), pp.1317–1318.

Chen, J.H. and Asch, S.M., 2017. Machine learning and prediction in medicine — beyond the peak of inflated expectations. *New England Journal of Medicine*, 376(26), pp.2507–2509.

Topol, E.J., 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), pp.44–56.

Tang, F., Ishwaran, H., and Mallick, B.K., 2021. Interpretable machine learning for healthcare: a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(6), e1420.

Doshi-Velez, F. and Kim, B., 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint* arXiv:1702.08608. [online] Available at: https://arxiv.org/abs/1702.08608 [Accessed 8 May 2025].

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M.P. and Ng, A.Y., 2017. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint* arXiv:1711.05225. [online] Available at: https://arxiv.org/abs/1711.05225 [Accessed 8 May 2025].

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H. and Wang, Y., 2017. Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4), pp.230–243.

Bzdok, D., Altman, N. and Krzywinski, M., 2018. Statistics versus machine learning. *Nature Methods*, 15(4), pp.233–234.