Optimising Heart Attack Prediction: Comparing Deep Learning and Traditional Machine Learning Techniques

MSc Research Project
MSc in Data Analytics

Melbin Biju
Student ID: x23165839

School of Computing
National College of Ireland

Supervisor: Dr David Hamill

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Melbin Biju
Student ID:	x23165839
Programme:	MSc in Data Analytics **Year:** 2024
Module:	Research Project
Supervisor:	Dr David Hamill
Submission Due Date:	12th December 2024
Project Title:	Optimising Heart Attack Prediction: Comparing Deep Learning and Traditional Machine Learning Techniques
Word Count**:** 8839	Page Count: 21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:

Date: 12th December 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Optimising Heart Attack Prediction: Comparing Deep Learning and Traditional Machine Learning Techniques

Melbin Biju
Student ID: x23165839

Abstract

This research focuses on understanding the effect of deep learning techniques for predicting heart attack risk in comparison with the traditional machine learning models. Heart disease, especially heart attack, is a global leading cause of mortality and early risk prediction is important for saving the life of patients and better health outcomes. This study aims at comparing the performance of deep learning model, namely Multilayer Perceptron (MLP), with traditional machine learning models such as Logistic Regression, Random Forest, XGBoost and Support Vector Machine (SVM). Based on a dataset comprising demographic, clinical, and medical features, selected models were trained and tested to assess risks of heart attacks. Different metrics such as accuracy, precision, recall, F1 score, and AUC were used in the evaluation of the models. The findings showed that performance of the MLP model was higher compared to other traditional machine learning models, especially in recall, which is necessary for the identification of high-risk patients and minimising false negatives. Implementation of hyperparameter tuning further increased the model performance and strengthened the applicability of deep learning models in clinical practice. This research adds to the current knowledge in using Artificial Intelligence (AI) for the prediction of heart attacks, emphasising the real-world applicability of the models, and effective resource management.

1 Introduction

Myocardial infarction (MI), commonly known as a heart attack, occurs when a person experiences a blockage in the blood flow in a specific part of the heart (Elsheikh et al., 2023). With millions experiencing heart attacks each year, heart disease remains one of the leading causes of death around the globe. According to the World Health Organisation (WHO), in 2019, there were 18 million deaths around the globe caused by heart disease, with approximately 85% of these deaths owing to heart attacks and strokes (World Health Organisation, 2021). Heart attacks not only affect the heart but also other tissues and organs. Hence, identifying and stopping heart attacks early is important for the reasons of lowering mortality rates and improving patient's outcomes (Fang et al., 2019). Conventional diagnostic approaches, comprising clinical evaluations, blood tests, and electrocardiograms (ECGs), require considerable time and are typically expensive, which is hard for healthcare systems to provide prompt and correct diagnoses to a widespread population. In addition, the complicated nature of heart disease and the changes in individual patient data contribute to ongoing challenges related to diagnostic uncertainty.

In the past few years, innovations in artificial intelligence (AI) and machine learning (ML) have suggested the ability to better address these challenges by delivering improved and reduced-cost tools for heart attack prediction. Traditional statistical techniques cannot expose the complex patterns and risk factors that machine learning algorithms can reveal in large datasets of patient data. Machine learning techniques, especially Multilayer Perceptron (MLP) within Feedforward Neural Network (FNN) in deep learning have appeared as a strong means for managing nonlinear relationships and making predictions with diverse variables. Although traditional machine learning techniques are useful, they frequently have trouble handling non-linear, high-dimensional data in healthcare, making deep learning a suitable solution for this circumstance.

Various traditional machine learning algorithms were used in the past to accurately classify heart attack prediction. However, there is a lack of implementation of advanced machine learning techniques like deep learning and the performance comparison of both traditional and advanced machine learning techniques in this case. Hence, the goal of this research inquiry is to assess the power of deep learning techniques for the prediction of heart attacks and to measure their performance relative to traditional machine learning algorithms, consisting of Logistic Regression, Random Forest, XGBoost, and Support Vector Machines (SVM). Specifically, the study will investigate the ability of deep learning models to improve the accuracy, precision, recall, and total effectiveness of heart attack risk prediction compared with conventional approaches. By carrying out this approach, the research aims to expand the current body of knowledge about AI-driven healthcare solutions by doing so, it seeks to give insights that can enhance clinical practices and the quality of patient care and resource utilisation. Predicting heart attack risks with precision supports clinical decision-making and also enables individualised treatment strategies, making certain that patients receive prompt, targeted interventions in accordance with their particular risk profiles. The potential to lower healthcare expenditures and lighten the load on emergency services renders this research particularly important within the current limited healthcare resources scenario.

Research Question. *To what extent the deep learning techniques help to improve predicting risk of heart attack for maximising benefit to patients and healthcare providers?*

Sub Research Question. *How the performance of deep learning techniques can be compared to traditional machine learning algorithms for predicting heart attack?*

This research will develop and compare several predictive models using a comprehensive dataset of patient information in order to resolve these questions. Assessment of each model includes accuracy, precision, recall, and F1 score and comparing them with other models. In addition, this study will look into ensemble methods and hyperparameter tuning in order to optimise the performance of these models.

This research project report is structured as follows: Introduction section provides the background and motivation for the study, states the research questions and gives the research objectives. The Literature Review section presents the related literature on the prediction of heart attack and the various machine learning algorithms used in the analysis. Methodology section explains the research methodology adopted in the study for data collection, model building and evaluation procedures and Design Specification section outlines the design and implementation of the predictive models used in deep learning and traditional machine learning models. The results and assessment of the comparative analysis of these models are discussed in the Results and Evaluation section. In the Discussion and Conclusion section, a critical analysis takes place of the results in the context of the research questions and the literature. Then in the Conclusion, an analysis of the implications of the research, and recommendations for future research and clinical work.

2 Related Work

Acute myocardial infarction (AMI) is a leading cause of death worldwide, therefore, predicting AMI is a critical area of research which potentially could save many lives. With advancements in machine learning and deep learning techniques, researchers are considering whether these technologies can help to advance early diagnosis and provide preventable care. Studies have identified key risk factors such as hypertension and diabetes, however, there are challenges, such as small datasets, poor representation of genetic risk factors and model interpretability. The objective of this literature review is to present what is been written in the literature regarding heart disease prediction, including clinical risk factors, comparing machine and deep learning models, and determining the effect of ensemble methods and hyperparameter tuning on the performance of the model. This review will identify the gaps and indicate better ways of using diverse datasets focusing on heart attacks, along with powerful modelling techniques to address these gaps and improve clinical applicability.

2.1 Heart Attack Risk Factors

The study by (Mehri and Hammami, 2017) highlighted the importance of clinical and biochemical risk factors associated with acute myocardial infarction (AMI). The key strength of this research was the various analyses performed on AMI patients and healthy controls, especially when showing differences in blood pressure, glucose levels, and lipid profiles. This has helped us understand the risk levels where there were some independent predictors of AMI, such as smoking, diabetes, and hypertension. However, one of the major limitations in this research was the small sample size used for the analysis and therefore, restricts the generalisation of results for the entire population. Furthermore, they failed to explore the genetic variables in the analysis. This limitation points to a direction of future research by including the genetic factors thereby ensuring accurate results of the study.

Previous research has been conducted on the role that the social and psychological variables attached to men and women had on heart attacks (Nyström et al., 2022). According to their analysis, stress levels, cholesterol and triglycerides were much higher for women, and they were more prone to heart attacks than men. Although they contributed meaningful insights on the gender difference in heart attacks, it could not be generalised based on their small sample size. Furthermore, they didn’t fully explain how lifestyle risk factors affected stress levels and cardiovascular health. In future research, large and diverse samples would be more appropriate for a more in-depth understanding of the effects of gender on the occurrence of heart attacks.

According to (Yusuf et al., 2020), investigations into the modifiable risk factors for cardiovascular disease and mortality in 21 countries at different levels of income among

155,722 individuals have shown that a significant portion of the heart disease cases and deaths were attributed to a range of modifiable risk factors including hypertension and tobacco use. The study incorporated high-income (HIC), middle-income (MIC) and low-income countries (LIC), thus enabling a globally representative perspective that previous research lacks. The reliability of this research was increased with standardised data collection; however, generalisability was undermined by the limited representation of HICs. The study further highlighted that hypertension was a major risk factor and should be brought under target intervention to address specific local issues, particularly in LICs and MICs, to promote the development of context specific policies for reducing the cardiovascular risks globally.

2.2 Comparative Studies in Predicting Heart Disease

A comparative analysis between machine learning and deep learning models for heart disease prediction was completed by (Bharti et al., 2021). Traditional machine learning algorithms such as Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, Decision Tree and deep learning were used for building models within this study. The architecture of the deep learning model consisted of three dense layers. The first and second dense layers used 128 and 64 units whereas the third layer used 32 units. To prevent overfitting issues, dropout layers were also added after the first and second dense layers with values of 0.2 and 0.1 respectively. The result showed that the deep learning model outperformed other models with an accuracy of 94.2%. The KNN model had the highest accuracy (84.8%) among the traditional machine learning models. This study also highlighted the importance of data preprocessing like feature selection and normalisation in predicting heart disease. However, the use of small dataset size made this study not generalised among all populations. The lack of a detailed explanation of hyperparameter tuning for the deep learning model and the issue of model interpretability was a significant drawback of this study for the enhanced performance and healthcare applications in clinical decision-making.

Additionally, (Singh et al., 2023) performed a comparative study on the performance of traditional machine learning algorithms such as Naïve Bayes, Logistic Regression, and Decision Tree with deep learning technique Multi-Layer Perceptron (MLP) for early prediction of heart disease. Based on their research, the Decision Tree model showed the highest accuracy (98.04%) whereas MLP had the second highest accuracy (95.51%). Although this research showed the highest accuracy, the dataset used for this study was small, only having 1025 instances. The small size of the dataset led to overfitting of the model, particularly for high-capacity models like MLP. Another major drawback of this study was the lack of transparency of the dataset, specifically in demographic diversity limiting the generalisability of the model in a wide range of populations. These potential issues could be overcome by using large, diverse datasets and by addressing overfitting challenges.

Research by (Gupta & Seth, 2023) analysed the efficiency and applicability of machine learning and deep learning approaches in predicting heart disease based on UCI heart disease and Framingham heart study datasets. From performance analysis of various models including Decision Trees, Random Forests, K Nearest Neighbors, Support Vector Machines and Multi-layer perceptions, it was found that Random Forest showed an accuracy of 97.13% over the Framingham dataset. However, due to a small sample size in the UCI dataset (303 records), such results implied limited generalisability, as well as a lack of discussion about model interpretability, especially of the black box nature of deep learning. Future research could focus on the analysis of the model's robustness by utilising multiple datasets and advanced ensemble techniques.

2.3 Ensemble Methods and Hyperparameter Tuning

The inquiry by (Asif et al., 2023) evaluated ensemble learning algorithms such as ExtraTrees Classifier, Random Forest, XGBoost, CatBoost by implementing different hyperparameter optimisation techniques to enhance the prediction of heart diseases. The study combined three Kaggle datasets having similar features to review a comprehensive data regarding heart disease. Based on the analysis, they found that hyperparameter optimisation methods like GridSearchCV and RandomizedSearchCV significantly influenced the performance of models. The proposed ExtraTrees Classifier model showed the highest accuracy (98.15%) using GridSearchCV, outperforming other models and RandomizedSearchCV method. The limitations of this study included focusing only on the specific set of ensemble techniques and overlooking advanced methods like deep learning to capture the non-linear relationships in the dataset. The limited size and diversity of the dataset made this result not generalised to different demographics. The authors suggested that future research could integrate clinical data with genetic variables to enhance the performance and reliability of the model.

The importance of implementing feature selection and ensemble techniques was explained by (Gupta & Seth, 2022) to improve the prediction of coronary heart disease. The algorithms considered for this study were Random Forest, Decision Tree, Support Vector Machine (SVM), Gradient Boosting, and K-Nearest Neighbors. They selected the eight best features - systolic blood pressure, glucose, age, cigarettes per day, total cholesterol, diastolic blood pressure, prevalent hypertension, and sex - using the Chi-Square test which improved the model interpretability and reduced computational load. After applying feature selection to the dataset, these features were used with three ensemble methods – stacking, majority voting, and bagging. Among these methods, majority voting showed the highest increase in accuracy about 7.42% after applying feature selection, with a final accuracy of 98.38%. The stacking and bagging methods resulted in an increase in accuracy of 2.11% and 0.14% respectively. This revealed the importance of feature selection in enhancing the performance of ensemble methods. However, the reliance on a single dataset and failure to implement advanced techniques like deep learning limited the acceptance and reliability of the model. Overall, although this study established the importance of feature selection and ensemble for the prediction of heart disease, there exist many other advanced and diverse methods that can be adopted in the subsequent studies for improving the prediction.

A study by (Firdaus, Nugroho & Soesanti, 2021) explored deep neural networks (DNNs) for predicting heart disease using a stepwise approach, including data pre-processing and hyperparameter tuning using random search, grid search and Bayesian optimisation. The best performance of the model was achieved by Bayesian optimisation compared to other optimisation methods. Despite this, it didn't sufficiently handle the biases in the Cleveland heart disease dataset and what they imply about model generalisation. Moreover, performance metrics such as accuracy and F1-score were provided, however, other machine learning techniques were not compared, which limits the context of its findings. Overall, this study provided useful information and could be more powerful if those dataset limitations were addressed by implementing a comprehensive comparative analysis.

2.4 Traditional Machine Learning Techniques

According to (Obasi & Shafiq, 2019), the study proposed a robust Random Forest model for predicting heart attacks. The strength of their study was the use of a comprehensive dataset containing 18 features with a considerable number of records. In this study, it showed 92.44% accuracy of the Random Forest model for predicting heart disease. More importantly, the implementation of the K-fold cross-validation further strengthens the reliability of the model, overcoming the problem of overfitting and for better model validation. However, they only included three algorithms - Random Forest, Linear Regression, and Naïve Bayes limited this study. They could consider superior algorithms like Support Vector Machines (SVM) and Gradient Boosting for better comparison of the performance of the models. The clinical applications of the model should be further discussed for how this effectively integrates with the healthcare system. Although the proposed model showed promising results, the use of other algorithms and discussion on clinical practice would lead to a more complete understanding of how the model can be used practically.

(Jindal et al., 2021) examined the cardiovascular disease prediction using K-Nearest Neighbors (KNN), Logistic Regression, and Random Forest on a popular UCI heart disease dataset. It reported 88.5% accuracy for K-Nearest Neighbors (KNN), and Logistic Regression algorithms and found these outperformed Support Vector Machine (SVM) and Decision Tree based on the previous studies. While these algorithms proved effective, this study didn't present a comparison of the machine learning techniques. The dataset used only contains 303 instances which affects the diversity and generalisability of this work. Including cross-validation, hyperparameter tuning and metrics other than accuracy could make this study more reliable. They failed to discuss the practical deployment of the model in the real world for the advancement of heart disease patients. Hence, future research could consider a comprehensive study of implementing various machine learning techniques to enhance heart disease prediction.

The study by (Nandal, Goel & Tanwar, 2022) explored different machine learning algorithms such as Support Vector Machines (SVM), Logistic Regression, Naïve Bayes and XGBoost for heart attack prediction based on the dataset from UCI ML Repository. The methodology used includes data collection, pre-processing, exploratory data analysis (EDA) and model development for accurate prediction. According to their findings, XGBoost performed better than other classification models, which showed 91% training accuracy, 89% testing accuracy and AUC score of 0.94. Logistic Regression and Naïve Bayes also showed solid performance, whereas Support Vector Machines (SVM) lagged due to the non-linear nature of the heart disease data, which was managed by the XGBoost algorithm. Although their studies were promising, the small sample size of this study limited its generalisability across diverse populations. They failed to incorporate effective feature selection and dimensionality reduction methods for enhancing heart attack prediction. The authors suggested future work include deep learning methods and hybrid models to capture more complexity in data for improving the prediction of heart diseases.

The literature shows that while machine learning and deep learning have improved heart disease prediction, most studies consider heart disease from a boarder perspective where they lack in the development of heart attack-specific models. While traditional machine learning algorithms such as Random Forest and Support Vector Machines are effective, they tend to struggle with the non-linear and complex nature of most health data. Deep learning techniques have the capability to handle this complexity, however, these techniques are underutilised specifically in the case of predicting heart attacks. This gap demonstrates that targeted deep-learning research on heart attack prediction is required for improved accuracy, model interpretability, and clinical relevance. Thus, my research addresses this by comparing machine learning and deep learning techniques solely for heart attack risk prediction to deliver new insights as well as improved tools for accurate early diagnosis within this field.

To conclude, the literature review examines the current state of research in the prediction of heart attacks, explains the limitations of the traditional machine learning algorithms in the case of non-linear, complex nature of datasets and lack of use of the deep learning techniques in this field. This review identifies the research gap in these existing studies and recommends targeted research to develop more effective heart attack specific predictive models. In the next section, the methodology of this study is described explains how data is collected, pre-processed and compared the performance of traditional machine learning and deep learning models in determining heart attack risks.

3 Research Methodology

This research project focuses on predicting heart attack risks by developing predictive models using both traditional and advanced machine learning algorithms. The primary objective of this research is to understand the effectiveness of deep learning techniques in predicting heart attacks compared to traditional machine learning algorithms. Moreover, it also checks how the performance of deep learning models can be compared to traditional machine learning techniques. To systematically address the aim of this project, it follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodological approach. This approach was selected for this project because of its structured and iterative nature for each phase in the data mining process, allowing the achievement of accurate, effective, and immediate applicable results for clinical practice. Even though KDD (Knowledge Discovery in Databases) also presents a set of structured phases, it does not identify with business context as CRISP-DM does, which is why it is not suitable for developing healthcare projects with particular clinical objectives. SEMMA (Sample, Explore, Modify, Model, Assess) mainly deals with modelling while neglecting the formulation of goals and understanding of problems, which is crucial in cases such as heart attack prediction with high-risk consequences. This project goes through the following processes based on a CRISP-DM approach.

3.1 Business Understanding

The initial phase of CRISP-DM methodology was focused on the importance of predicting heart attack risks and its business perspectives. Heart disease patients are increasing day by day which burdens the capacity of the healthcare system, resulting in millions of deaths annually. Timely intervention is important for reducing mortality rates and better management of the lives of the patients.

To explain the relevance of this research, the literature review was conducted to understand the challenges faced by the healthcare workforce in caring for patients with heart diseases especially with heart attacks. Such difficulties include the need to increase the accuracy of identification of high-risk patients, enhancing the treatment plans and executing effective strategies in monitoring heart disease patients. Based on these requirements, the predictive models were adjusted according to their needs. The insights from understanding the business requirements not only helped to select the methodology choices of this project but also emphasised its impact on clinical decision-making and resource-limited settings where special care for heart disease patients is needed.

3.2 Data Understanding

In the data understanding phase, it was sourced a dataset from Kaggle website, containing a wide range of features connected with heart health. The data consists of 4 databases collected from multiple regions including Cleveland, Hungary, Switzerland, and Long Beach V with the primary purpose of studying and predicting heart diseases in patients. The dataset contains 13 features and a target variable heart disease for prediction. The features include demographic information (age, sex), clinical symptoms and measurements (resting blood pressure, cholesterol, maximum heart rate achieved, and chest pain type), health conditions and medical history (resting electrocardiographic results, fasting blood sugar, exercise-induced angina, depression in ST segment during exercise relative to rest, slope of peak exercise ST segment, and the number of major vessels coloured by fluoroscopy), and thalassemia condition (a blood disorder).

Adding certain additional features to the dataset makes this study specific for predicting heart attacks. The additional features include troponin levels in the blood, previous heart attack history, chest pain type (detailed), smoking history, family history of heart disease, heart rate recovery, and assessing the level of physical activity of each patient. Adding these features made this dataset suitable to predict heart attack-specific problems.

Initial data exploration was conducted to understand the overall structure of the dataset. Summary statistics such as mean, median, standard deviation, etc. for the numerical features and count for the categorical features were checked. Datatypes were verified if they were appropriate to each feature or not. Also checked any need for transformation in the data type of each feature. Basic visualisations were carried out using bar charts, pie charts, and box plots to understand the distribution, outliers, and patterns of the features. The distribution of the target variable heart attack status was plotted using a bar chart and pie chart as shown in Figure 1 below.

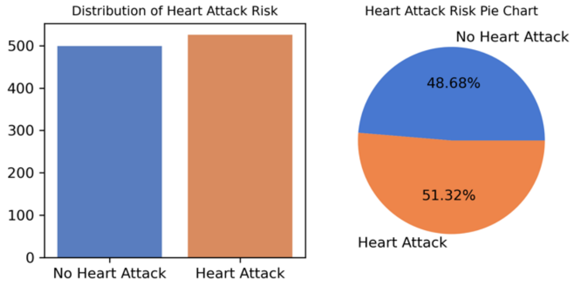

Figure 1: Heart Attack Risk Distribution

There were two classes available in the target variable to understand the heart attack status of the patient. From the pie chart, it showed 51.32 percent of patients experienced heart attacks in the dataset. Almost equal representation of each class of the target variable made this dataset balanced. The distribution of features such as sex, chest pain, fasting blood sugar, resting electrocardiographic results, exercise-induced angina, ST slope, number of major vessels coloured by fluoroscopy, and thalassemia condition with respect to the heart attack status of the patients plotted in Figure 2 below. According to the gender distribution of the dataset, men showed a greater number of heart attack cases compared to women, suggesting gender-based risk differences. The chest pain type feature appeared to be a significant predictor that showed a higher number of heart attacks with patients who had non-anginal chest pain, while asymptomatic associated with lower occurrences of heart attacks. Also, it showed an almost balanced number of patients with typical and atypical angina who experienced heart attacks.

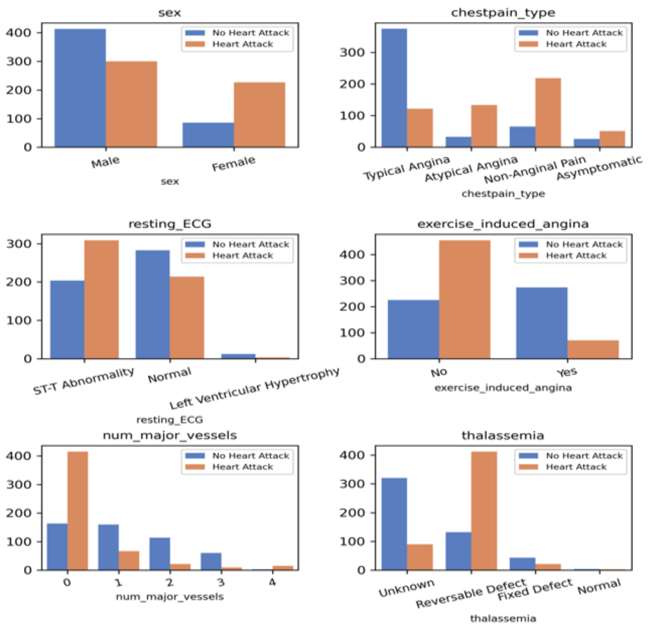

Figure 2: Distribution of Categorical Features w.r.t Heart Attack Risk

Most of the patients with heart attacks had ST-T wave abnormality whereas probable or definite left ventricular hypertrophy was shown by very few patients in the dataset based on the resting electrocardiogram results. The thalassemia feature represented the blood disorder condition with different levels of thalassemia. The trend of this feature suggested that most heart disease patients experienced reversible defects whereas many unknown or missing status of thalassemia were reported in those without heart attacks. The majority of heart disease patients showed downsloping ST segment who went through the highest risk of heart attacks, followed by flat and upsloping in the slope of the peak exercise ST segment feature.

While fasting blood sugar over 120 mg/dL is generally a risk factor for heart diseases, it seemed that many of the patients with heart attacks had normal fasting blood sugar in this dataset. The distribution of the typical exercise-induced angina patients was lower than that of the patients without exercise-induced angina. The number of major vessels coloured by the fluoroscopy analysis was contrast for the general perception of higher chance of heart disease to understand the blood flow to the heart. The risk of heart attacks increases when the number of vessels affected by blockage increases. Most heart disease patients showed zero majority vessels, while a higher number of majority vessels were more common in patients without heart attacks. These suggested other features might be more predictive of heart attacks than these features. Generally, these were against the trend of heart attack patients; therefore, it would be further discussed in detail in the Discussion and Conclusion section.

From Figure 3 below, it shows the distribution of additional features added in relation to heart attacks. Based on the analysis, it was found that most of the patients who had a previous heart attack history were significant contributors to heart attacks. Many patients were affected by heart attacks even though they didn't have a family background of heart disease. The physical activity level was too low for most of the heart attack patients while high physical activity significantly reduces the occurrence of heart attacks. The smokers in the dataset had a higher chance of heart attack compared to the non-smokers. The most common type of chest pain experienced by heart attack patients was pressure where they feel heavy or squeezing sensation followed by radiating, sharp, and dull nature of the chest pain. This analysis highlighted several key risk factors that contribute to heart attacks.

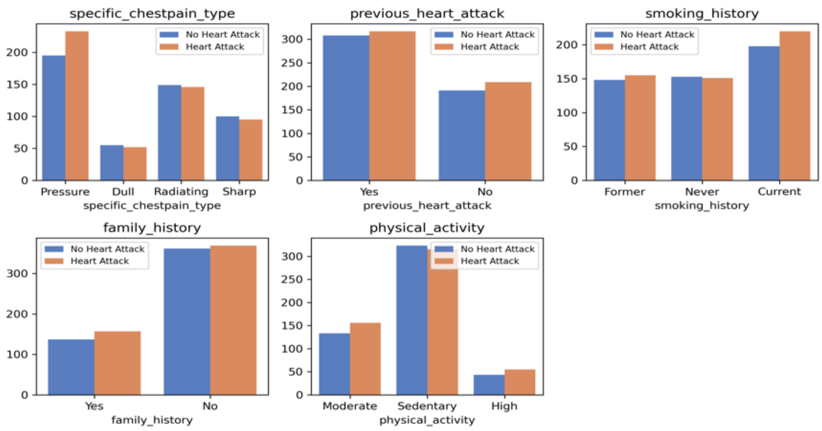

Figure 3: Distribution of Additional Features w.r.t Heart Attacks

The best plot to understand the distribution and to identify the outliers in the dataset is a box plot as shown in Figure 4 below.

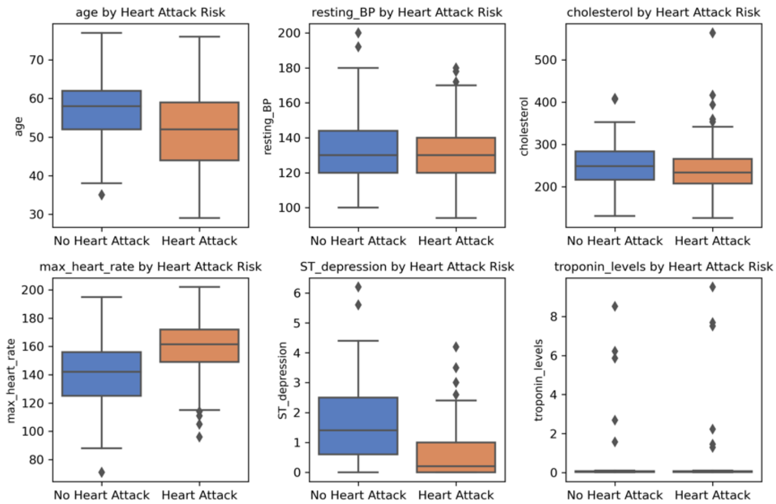

Figure 4: Features with Outliers

Based on this plot, it was identified significant outliers in resting blood pressure, cholesterol, maximum heart rate, ST depression, and troponin level features. The median age of the patients was around in the mid-50s with only a less extreme low outlier in patients with no risk of heart attack in this feature. The long whiskers of this feature showed the data were more spread. Significant high outliers were reported in the amount of troponin levels in the blood. The distribution of resting blood pressure showed several outliers above the value of 170. The interquartile range of this feature spans between 120 and 140 values. The cholesterol feature had several high values above 400. The majority values of this feature were less than 300, with a median of around 250. The median value of the maximum heart rate achieved was close to 150, with values generally between 120 and 170. Finally, ST depression feature showed most of the values concentrated at lower levels between 0 and 1.5, but a few high outliers pushed the range upwards.

3.3 Data Preparation

To ensure the reliability and robustness of the models, preparing the data suitable for machine learning and deep learning models is an important step in predicting heart attacks. Initially, it was checked for any missing values and duplicate values present in the dataset. The outliers in several features were treated using the capping (or winsorizing) method where the extreme values were replaced with the nearest values within a specified percentile range. The reason to select the capping method over the z-score method, and IQR (Interquartile Range) method was because of its ability to balance simplicity, preserve data structure, and avoid important information loss, especially in the case of small datasets. The z-score method accurately identifies the outliers only if the data is normally distributed whereas IQR method is typically used to detect outliers, not to cap them often results in removing outliers. Since not all the features of this dataset were not normally distributed, it was selected as a capping method for treating outliers.

The age variable was binned for better perspectives of analysis to reduce the noise level and for better model performance by group-based trends. Age was divided into 6 categories and then these categories were encoded each age group to an integer value. This conversion was done in stages of age increase where the younger patient data was represented by small integer values. Old age groups might show different behaviors compared to younger age groups with respect to heart attacks. Also, it was able to keep the numerical order of the age groups by using ordinal encoding.

The encoding of the categorical features was done using label encoding, one-hot encoding, and binary encoding. The physical activity level and smoking history features were label encoded because of its importance in maintaining the order of the categories in the feature. The sedentary, moderate, and high categories of physical activity showed varying degrees of relationships between the categories. The chest pain type, resting electrocardiographic results, thalassemia condition, and specific chest pain type (detailed) features do not show an inherent order between the values, for that reason it was selected one-hot encoding method for these features. Binary features such as previous heart attack and family history were encoded using a map function because it only had two distinct categories. There were some of the variables found high correlation with other variables. So, feature selection was implemented by variance inflation factor (VIF) to treat multicollinearity issues in the dataset.

At last, the dataset was split into training and testing sets in order to assess the generalisation capability of the models. The data was split to 80:20 where 80% was used for training while the remaining 20% was used for testing the model. After splitting the data, all features were made equally contributing to the model through data standardisation which enhanced the model performance.

3.4 Modelling

The modelling phase of CRISP-DM methodology provides a structured approach to develop predictive models for heart attack risk. The primary objective of this section is to compare the effect of deep learning Multilayer Perceptron (MLP) model with traditional machine learning techniques for predicting heart attacks.

In the modelling phase, there were some suitable traditional machine learning algorithms selected for predicting heart attacks. Logistic Regression, Random Forest, XGBoost, and Support Vector Machine (SVM) were selected for developing traditional predictive models based on their efficiency in solving classification problems and their capability to handle complex relationships, particularly in medical datasets. All these algorithms are unique in their ways, and their strengths can be applied to predict heart attacks. Logistic Regression, one of the straightforward and interpretable algorithms for binary classification problems, allows healthcare professionals to easily understand the contribution of each feature to predicting heart attacks. This algorithm is relatively efficient to compute makes this a good option for comparison against other techniques which are more advanced. Random forest has the capability to capture complex interactions and non-linear relationships in the features of the dataset without the need of deep feature engineering. This algorithm performs well because it average various decision trees which results reduce the chances of overfitting. Furthermore, it helps to get the feature importance which can be used to understand the risk factors for prediction of heart attack risk. The peculiarity of the XGBoost algorithm is high performance and high speed, especially in classification tasks. This algorithm has the ability to effectively manage imbalanced datasets, which is common among the medical datasets where many of the cases of heart attack are less frequent compared to cases without heart attacks. Support Vector Machine (SVM) algorithm is particularly effective in high dimensional data relative to the number of records available in the dataset. The kernel trick of SVM allows the creation of non-linear decision boundaries helpful in separating heart attack cases from non-cases. Also, it shows robustness to overfitting, especially in high-dimensional datasets.

While many traditional machine learning algorithms were available, the selected algorithms showed distinct advantages for predicting heart attack risk. Some of the other algorithms were not implemented because of various reasons. For example, Naïve Bayes was not implemented in the study because it assumes feature independence which is often not possible with medical data. Additionally, while decision trees are easy to interpret high chances of being overfitted if it is not pruned correctly affect the generalisability of data. The K-Nearest Neighbors (KNN) algorithm faces difficulty in managing imbalanced classes which is very important in medical data connected with heart diseases. This algorithm is sensitive to feature scaling, and it is also highly computational with large datasets. Thus, only selected algorithms were considered capable of handling complex data related to heart attacks and showed more reliable and accurate results than other popular traditional algorithms in machine learning.

To understand the effect of deep learning techniques, Multilayer Perceptron (MLP), a type of deep learning neural network, was implemented in predicting heart attacks. The use of this deep learning approach is highly effective in complex datasets which are often found in medical data like heart disease datasets. MLP algorithm is simple to implement compared to other deep learning techniques. For structured tabular data, MLP is the best choice to enhance the performance of the model. More complex architecture such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) are most suitable for specialised data formats (images or sequences) where MLP is the best choice for this specific task.

In order to obtain better predictive accuracy for each of the models, GridSearchCV, one of the most effective grid-search techniques was used. This was useful in identifying right parameters which can improve the performance of each of the model. For the MLP model, tuning was focused on hidden layer sizes, batch size, number of iterations, and the strength of regularisation. Parameters for tuning for Logistic Regression include the strength of the regularisation and the type of solver while for Random Forests, parameters include the number of trees, the depth of trees, and minimum of samples per split. XGBoost tuning comprised a selection of learning rate, number of estimators, tree depth as well as regularisation parameters. SVM tuning was done by a selection of both the kernel and a related regularisation parameter that was used to avoid overfitting.

3.5 Evaluation

The performance of the classification models was assessed on the test dataset based upon different evaluation metrics including accuracy, precision, recall, F1 score, ROC-AUC score and ROC curves to draw the comparison. Since false negatives are costly, recall was the primary measure used to maximise true positive rates (identifying patients at-risk). Precision and F1-score were the secondary performance metrics that allowed for avoiding both high number of false positives and low recall. ROC-AUC scores and curves showed how well the models could classify between the classes. The results of the deep learning MLP model and traditional machine learning models were compared and evaluated before and after the hyperparameter tuning. A combined ROC curve plot and a bar chart showing the values of metrics such as precision, recall, and F1-score were used to illustrate model performance differences. The accuracy of the models was also plotted before and after tuning was performed. The best model was selected based on performance and then deployed on AWS EC2 instance using Flask, PuTTY, PuTTYgen, and WinSCP.

3.6 Deployment

The best model used for the prediction of heart attack was deployed using Flask framework. Firstly, the Flask application was created and tested on the local environment for its proper working status. An AWS account was created and an EC2 instance was used for deployment of the application. PuTTY and PuTTYgen were used to obtain a proper private key for the purpose of the secure SSH connection while the data transferred to the server was done using the WinSCP. Due to the port configuration of the EC2 instance, changes were made in the Flask code to execute the code properly. The necessary libraries were installed on the server through PuTTY and the application was started by using the code 'python3 app.py'. The deployment was completed by using a web URL for testing the availability and working of the application.

4 Design Specification and Implementation

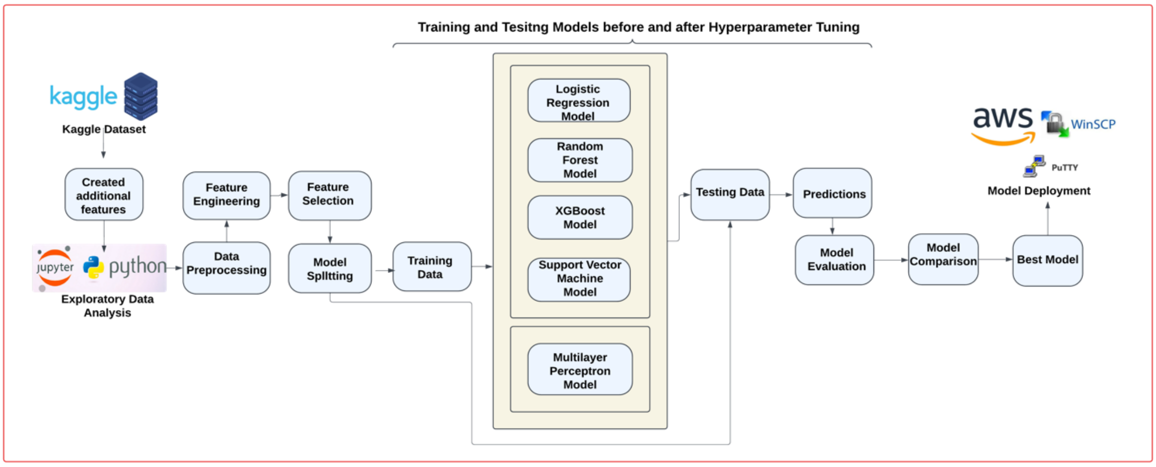

Figure 5: Project Workflow Diagram

The Design Specification and Implementation section describes the project development phase by phase with specific tools and libraries used. The heart disease dataset was obtained from Kaggle and pre-processed using pandas library in Python. Initial exploration of the data was first performed on the dataset to understand the dataset and then additional features were created using numpy.random.choice() function. For better understanding features and getting better plots, the column names were renamed for clarity and encoded features were converted to human readable format. Visualisation of the data was done using the seaborn and matplotlib libraries; the bar chart, pie chart, box plot, line plot, and correlation matrix were used to analyse the relationship and patterns.

Outliers were treated with the capping method by capping values below 5th percentile and above 95th percentile, and continuous features like age were binned into categories. Features were encoded using methods such as one-hot encoding (get_dummies), binary encoding (map), and LabelEncoder from scikit-learn. Feature selection was performed using the variance_inflation_factor function from statsmodels. The dataset was split into training and testing sets (80/20) using scikit-learn's train_test_split, and both sets were standardised with StandardScaler.

A deep learning model was built with the help of MLPClassifier from scikit-learn with two hidden layers (200 neurons in the first hidden layer and 150 neurons in the second one), 'tanh' activation function, L2 regularisation (alpha=0.1), a batch size of 20 and 'adam' optimiser. For hyperparameter tuning, the GridSearchCV was implemented to optimise hidden layer sizes (20, 20) and (150, 150), alpha values ([0.0001, 0.001, 0.01, 0.1]), batch sizes [10, 20, 30] and the max_iter [500, 1000].

Traditional machine learning models developed and tuned used in the study include Logistic Regression, Random Forest, XGBoost, and SVM. Logistic Regression used the C values ([0.001, 0.01, 0.1, 1]) and solvers (liblinear', 'saga'). Random Forest set parameters of n_estimators as [100, 200], max_depth as [2, 4] and min_samples_split as [5, 10]. XGBoost tried to optimise learning_rate with [0.001, 0.01], n_estimators with [100, 150], max_depth with [2, 4] and alpha with [0, 0.1, 1]. SVM tuned kernel types include 'linear' and 'rbf', and C values [0.001, 0.01, 0.1, 1)]. For the model evaluation, 10-fold cross validation was implemented and the tuning of hyperparameters greatly enhanced the efficiency of the model.

Evaluation metrics used to assess model performance included accuracy, precision, recall, F1-score, ROC-AUC score, and ROC curves. ROC plots and bar charts were used to compare metrics to show performance differences and improvements in accuracy were also plotted before and after tuning. The best model was deployed on AWS EC2 instance with the help of Flask, PuTTY, PuTTYgen, and WinSCP.

5 Results and Evaluation

In this section, the results of the proposed Multilayer Perceptron (MLP) model and traditional machine learning algorithms such as Logistic Regression, Random Forest, XGBoost, and SVM for classifying the heart attack risks are discussed. Changes in their predictive capabilities before and after hyperparameter tuning were compared based on evaluation metrics including accuracy, precision, recall, F1 score and AUC.

5.1 Baseline Model Performance

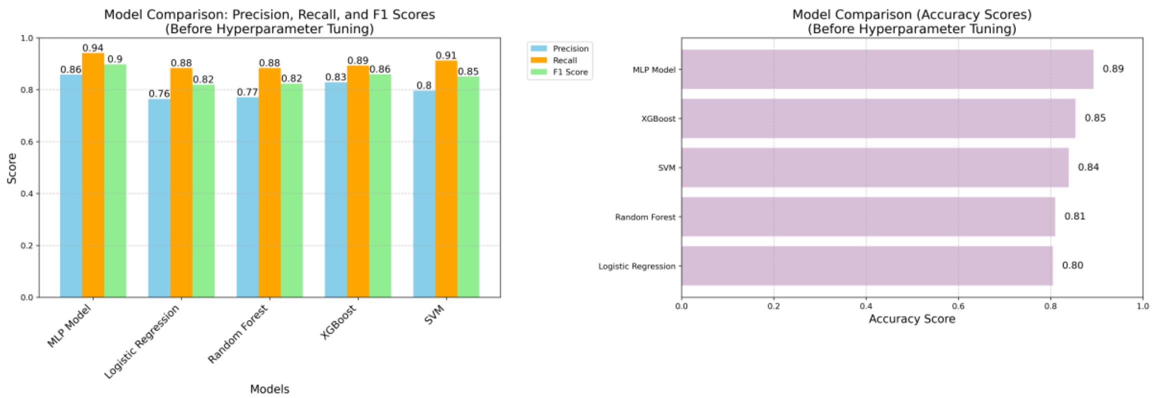

Figure 6: Performance of the Models before Hyperparameter Tuning

From the analysis done in Figure 6 above, it is clear that the MLP model performed the best out of the five models before implementing hyper parameter tuning. This model achieved an accuracy of 89.27%, precision of 85.84%, recall of 94.17%, F1 score of 89.81%, and an AUC of 0.972.

These results suggested that MLP performed exceptionally well in correctly classifying heart attack cases, especially in the recall metric which is vital in reducing the false negatives for timely identification of heart attacks. Although not the best performer, it was able to achieve well-balanced metrics with an accuracy of 85.37%, recall of 89.32%, F1 score of 85.98%, and AUC of 0.936 using the XGBoost model. SVM showed slightly better results in comparison with Random Forest in accuracy (83.90% vs. 80.98%), F1 score (85.07% vs. 82.35%) and AUC (0.898 vs. 0.895), proved its ability to find the right balance between the precision and recall.

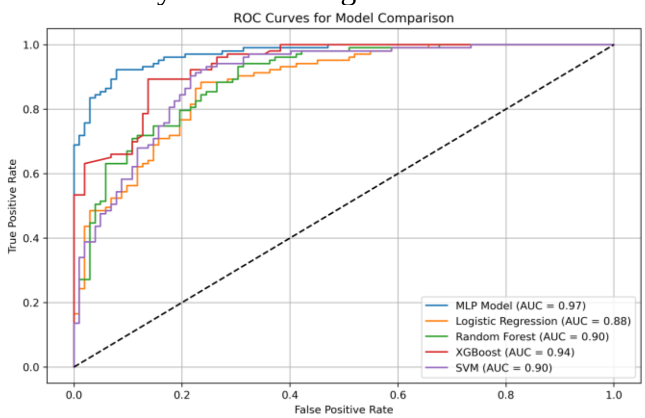

Figure 7: ROC Curves of the Models before Hyperparameter Tuning

A weaker baseline performance was observed for the Logistic Regression model with an accuracy of 80.49% and F1 score of 81.98%. However, considering the values given, its ability to separate classes was reasonable with the AUC of 0.879, while being less effective than other models as shown in Figure 7 above. These results showed MLP as the most effective model, followed by XGBoost, SVM, Random Forest, and Logistic Regression making MLP as the best model for further optimisation in this particular task.

5.2 Impact of Hyperparameter Tuning

The tuning of the hyperparameters showed improvements across all the models, especially in the MLP model as shown in Figure 8 below.

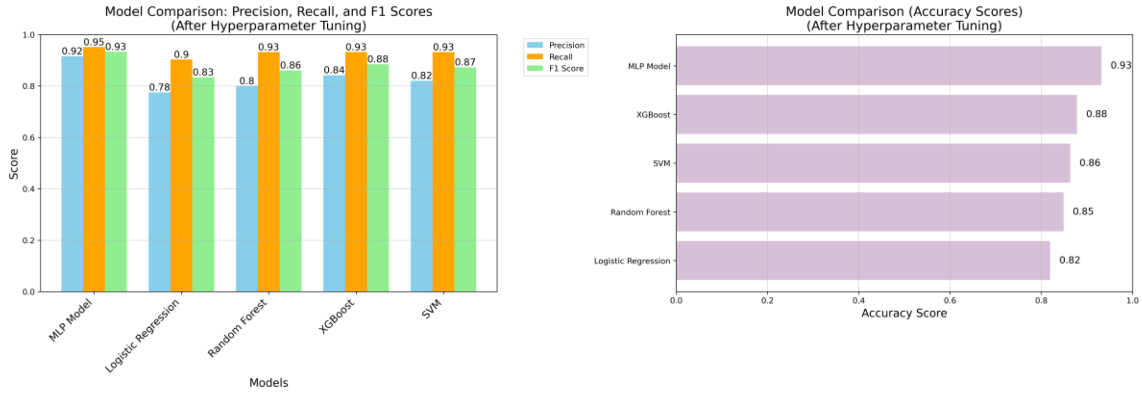

Figure 8: Performance of the Models after Hyperparameter Tuning

After hyperparameter tuning, MLP scored 93.17% accuracy, 91.59% precision, 95.15% recall, 93.33% F1 score and an AUC of 0.984, which reconfirmed MLP as the best model. XGBoost had a noticeable increase in performance showing improvement in accuracy to 87.80%, F1 score to 88.48%, and recall to 93.20%. There was also an increase in the performance of SVM model showing an accuracy of 86.34%, F1 score of 87.27 %, and a recall of 93.20%. For the Random Forest model, it showed 84.88% accuracy, 93.20% recall, and AUC of 0.927 respectively. On the other hand, there was modest improvement noticed for Logistic Regression model. The performance of Logistic Regression model increased to 81.95% of accuracy, 90.29% of recall, and the AUC of 0.880. Despite this improvement, this model had a lower performance compared to the ensemble methods and MLP model.

5.3 Comparative Analysis

A more comprehensive comparison of the ROC curves of the various models after optimisation of hyperparameters was done pointing the supremacy of MLP model. From the ROC curves in Figure 9 below, MLP showed the highest AUC of 0.984 proving its high capacity in correctly classifying the positive and negative cases. Additionally, the highest recall placed MLP in the highest reliable position to minimise the false negative cases, which is important in heart attack prediction where a positive case can be critical if missed. It was followed by XGBoost having an AUC of 0.950 and Random Forest with an AUC of 0.927 which proved to be less powerful in discriminating class than MLP.

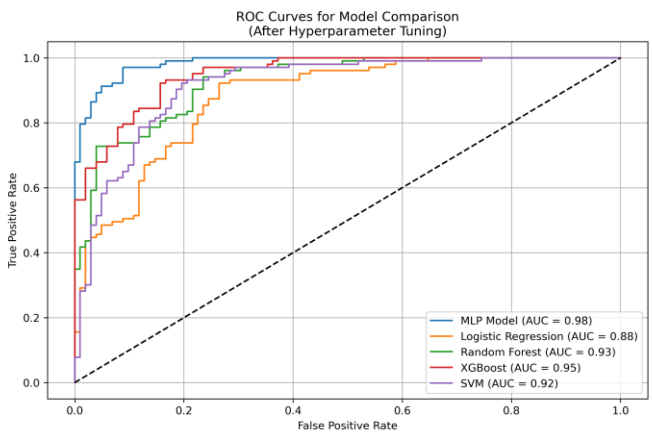

Figure 9: ROC Curves of the Models after Hyperparameter Tuning

Although SVM and Logistic Regression demonstrated better performance, they had slightly low AUC of 0.916 and 0.880 as expected due to their relatively weaker performance. This representation explains how MLP was capable of possessing an optimal balance between sensitivity (recall) and specificity which is vital in the kind of task such as identifying heart attacks.

Therefore, it is possible to conclude that MLP model is the best option for classifying and diagnosing heart attacks because it obtained the best accuracy, recall, and AUC metrics before and after tuning the hyperparameters. It's far better performance in a recall, one of the most important measures when predicting heart attacks, confirmed its appropriateness to this task. After tuning, XGBoost, SVM, and Random Forest performed better but fell slightly below MLP, while Logistic Regression showed improvements, but was not as efficient as the advanced ensemble and neural network methods. Furthermore, the findings show that recall metric is important in the classification of heart diseases, especially in the prediction of heart attacks and thus, the MLP model is more suitable for deployment in clinical settings where early detection of heart attacks is crucial.

6 Discussion and Conclusion

This study indicates that deep learning, especially the Multilayer Perceptron (MLP) model, is more useful in identifying possible heart attack risks than other traditional machine learning models. The MLP model showed the highest results among all the models developed, with a recall of 95.15% and AUC of 0.984 after tuning hyperparameters. Such results have identified its efficiency in reducing cases of false negatives which are vital in identifying and treating heart attack patients. However, some of the findings contradicted common expectations, such as a weak association between fasting blood sugar levels and risk of heart attack and additionally, a high proportion of patients with zero affected vessels, but they experience heart attack, and a relatively low correlation between exercise - induced angina and the heart attack risks. These discrepancies might be due to the features of this dataset. For example, there might be some sampling biases, discrepancies in the way data have been collected, or some confounding factors may not have been measured. Moreover, this dataset is not very large so it might not have adequately represented the complexity of actual heart attack risk profile cases.

The dataset used in this study was obtained from several geographical locations including Cleveland, Hungary, Switzerland, and Long Beach V. However, there were possible weaknesses in the dataset even though it covered different regions. Other synthetic variables were added to the dataset which may not be fully a representation of the actual records of the patients. Variables such as specific chest pain type, troponin levels, previous heart attack history, heart rate recovery, smoking history, family history of heart disease, and physical activity levels were generated fictitiously. Although these additions were intended to fill the gaps and extend the dataset to become more specific to heart attack prediction, they are not actual clinical data. As a result, the observed relation involving such variables might not be replicated in actual settings, which means that the findings would be hard to be generalised. These challenges are even exacerbated by the small dataset size because smaller datasets are prone to overfitting and may not represent a large enough patient population base to capture all the reported clinical cases.

Furthermore, there is not an important association with conventional risk factors such as fasting blood sugar compared to prior studies. This feature was one of the important predictors reflected in the study by Mehri & Hammami (2017) for the prediction of heart diseases. The same applies to such findings as the majority of heart attack patients had zero affected vessels, which may not align with clinical expectations showed chances of biases in data construction, analysis, or the limitations of dataset size. These anomalies point out the need for larger sample-size, clinically validated datasets from real-life to check the reliability and accuracy of the obtained results.

On the one hand, there are many advantages to this research. The presence of additional variables opened up more scope of opportunities in modelling in relation to risk factors of heart attacks and the inclusion of hyperparameter tuning proved improvements in model efficiency. Moreover, the recall-oriented evaluation approach emphasised the minimisation of false negatives which made the MLP model most effective in the early detection of heart attacks. The deployment of the MLP model on the AWS also showed the applicability of the concept in real-world applications, fulfilling the essential requirement of creating scalable, automated diagnostic model in clinical practice.

However, there should be more improvements needed in this study because of the black-box nature of the MLP model. Unlike models like Logistic Regression and Random Forest which provide a clear explanation for its predictions, the MLP model lacks interpretation of the results of the decisions made during the process limiting clinical adoption. It is possible that methods like SHAP or LIME would improve the model interpretability by explaining not only what the model focuses on, but also how exactly it makes its decision. In addition, it is needed to use a much more large and wide-ranging dataset to verify the generalisation and reliability of the model.

Conclusion

To conclude, the efficient model of deep learning technique, namely MLP model, depicted the promising results among all the selected traditional machine learning models for accurately predicting the heart attack at the early stage. The high values of accuracy, recall, precision, F1-Score, and AUC of MLP model showed the potential of this technique to reduce false negatives and maximising the accuracy which is critical in the case of predicting heart attacks. This developed model will help in the accurate prediction of heart attack patients, improving the quality of patient care, and better the utilisation of the limited resources. Although synthetic features enriched the dataset, using a limited sample size and some synthetic features in the dataset indicate the necessity of further research based on large real-world datasets. Integrating the best performing model on AWS showed its potential of application in clinical environments and presented a possible route to implement large scale and efficient diagnostic solutions.

References

Asif, D., Bibi, M., Arif, M.S. and Mukheimer, A. (2023). Enhancing heart disease prediction through ensemble learning techniques with hyperparameter optimization. *Algorithms*, 16, 308. https://doi.org/10.3390/a16060308.

Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S. and Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. *Computational Intelligence and Neuroscience*, 2021, Article ID 8387680. https://doi.org/10.1155/2021/8387680.

Elsheikh, E., AlKhudair, A.M., AlGanem, S.N., AlDandan, H.R., AlGhareeb, A.H., AlSalman, M.A., AlKhamis, A.A., AlHajri, H.M. and Alsubaie, G.A., 2023. Perception of heart attack risk factors and their complications among the adult population in the Eastern region of Saudi Arabia. *Cureus*, 15(12), p.e49860. DOI: 10.7759/cureus.49860.

Fang, J., Luncheon, C., Ayala, C., Odom, E., and Loustalot, F. (2019) 'Awareness of heart attack symptoms and response among adults — United States, 2008, 2014, and 2017', *Morbidity and Mortality Weekly Report*, 68(5), pp. 101–106.

Firdaus, F.F., Nugroho, H.A. and Soesanti, I. (2021). Deep neural network with hyperparameter tuning for detection of heart disease. In: *2021 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob)*, Bandung, Indonesia, 08-10 April 2021. IEEE. DOI: 10.1109/APWiMob51111.2021.9435250.

Gupta, P. and Seth, D. (2023). Comparative analysis and feature importance of machine learning and deep learning for heart disease prediction. *Indonesian Journal of Electrical Engineering and Computer Science*, 29(1), pp. 451-459. DOI: 10.11591/ijeecs.v29.i1.pp451-459.

Gupta, P. and Seth, D.D. (2022). Improving the prediction of heart disease using ensemble learning and feature selection. *International Journal of Advanced Soft Computing Applications*, 14(2), pp. [insert page range]. DOI: 10.15849/IJASCA.220720.03.

Jindal, H., Agrawal, S., Khera, R., Jain, R. and Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. *IOP Conference Series: Materials Science and Engineering*, 1022, 012072. IOP Publishing. DOI: 10.1088/1757-899X/1022/1/012072.

Mehri, S. and Hammami, M. (2017). Clinical and biochemical factors associated with acute myocardial infarction: Risk factors for acute myocardial infarction. In: *2017 International Conference on Engineering & MIS (ICEMIS)*, 08-10 May 2017, Monastir, Tunisia. IEEE. DOI: 10.1109/ICEMIS.2017.8273109.

Nandal, N., Goel, L. and Tanwar, R. (2022). Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis [version 1; peer review: 1 approved]. *F1000Research*, 11, 1126. Available at: https://doi.org/10.12688/f1000research.123776.1.

Nyström, A., Strömberg, S., Jansson, K., Olsen Faresjö, Å. and Faresjö, T. (2022). Cardiovascular risks before myocardial infarction: differences between men and women. *BMC Cardiovascular Disorders*, 22, 110. https://doi.org/10.1186/s12872-022-02555-3.

Obasi, T. and Shafiq, M.O. (2019). Towards comparing and using machine learning techniques for detecting and predicting heart attack and diseases. In: *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA, 09-12 December 2019. IEEE. DOI: 10.1109/BigData47090.2019.9005488.

Singh, G., Guleria, K. and Sharma, S. (2023). Machine learning and deep learning models for early detection of heart disease. In: *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, Greater Noida, India, 03-04 November 2023. IEEE. DOI: 10.1109/ICCCIS60361.2023.10425392.

Yusuf, S., Joseph, P., Rangarajan, S., Islam, S., Mente, A., Hystad, P. et al. (2020). Modifiable risk factors, cardiovascular disease, and mortality in 155,722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. *The Lancet*, 395(10226), pp. 795-808.