An Optimized Machine Learning Framework for Heart Disease Prediction: Achieving a New State-of-the-Art in Accuracy and Explainability

Israt Jahan

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Israt Jahan
Date of Birth	:	24-08-2002
Title	:	An Optimized Machine Learning Framework for Heart Disease Prediction: Achieving a New State-of-the-Art in Accuracy and Explainability
Academic Session	:	2022-2025

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☐	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

(Student's Signature)

(Supervisor's Signature)

221-35-1079
Student ID
Date: 23.12.2025

Afsana Begum
Name of Supervisor
Date: 23.12.2025

APPROVAL

This thesis titled on "**An Optimized Machine Learning Framework for Heart Disease Prediction: Achieving a New State-of-the-Art in Accuracy and Explainability**", submitted by **Israt Jahan (ID: 221-35-1079)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

-- **Chairman**

Dr. Imran Mahmud
Professor & Head
Department of Software Engineering
Faculty of Science and Information Technology Daffodil International University

-- **Internal Examiner 1**

Afsana Begum
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 2**

Md. Shohel Arman
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 3**

Nadira Islam
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **External Examiner**

Md Manowarul Islam
Professor
Department of Computer Science and Engineering
Jagannath University, Bangladesh

SUPERVISOR's DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science.

____

(Supervisor's Signature)

Full Name : Afsana Begum

Position : Assistant Professor & Coordinator M.Sc

Date : 23.12.2025

STUDENT'S DECLARATION

I HEREBY DECLARE THAT THE WORK IN THIS THESIS IS BASED ON MY ORIGINAL WORK EXCEPT FOR QUOTATIONS AND CITATIONS WHICH HAVE BEEN DULY ACKNOWLEDGED. I ALSO DECLARE THAT IT HAS NOT BEEN PREVIOUSLY OR CONCURRENTLY SUBMITTED FOR ANY OTHER DEGREE AT DAFFODIL INTERNATIONAL UNIVERSITY OR ANY OTHER INSTITUTION.

Israt

(STUDENT'S SIGNATURE)

FULL NAME : ISRAT JAHAN

ID NUMBER : 221-35-1079

DATE : 23.12.2025

ACKNOWLEDGEMENTS

It is a milestone that was attained through counselling, encouragement, and blessing of many people and institutions to fulfil the successful completion of this thesis. I would like to thank them all very much.

To begin with, I would like to express my intense gratitude to the Almighty Allah in terms of the divine power, patience, and mercy that allowed human resources to overcome the difficulties that were involved in this complicated research project.

I owe my greatest and most profuse gratitude to my esteemed supervisor, Ms. Afsana Begum, Assistant Professor and Coordinator, M.Sc. of the Department of Software engineering. Her

modest and mild speaking nature provided a special enabling and conducive atmosphere to learning. Her academic guidance and her all-inclusive dedication were the basis of this thesis, which brought enlightenment and course during critical methodological choices, and

ultimately, the shift to the ultimate contribution to the academic scene, which was Interpretability and Reliability. Her patience and trust played a significant role in a successful conclusion of this work.

I am also honoured to be appreciative to the whole faculty and staff of the Department of Software Engineering at Daffodil International University to supply the necessary academic provisions, infrastructure, and intellectual environment.

Lastly, I recognize the support, my undying faith, and immense sacrifice of my adorable parents. They kept me going through this demanding course through their encouragement and patience.

DEDICATION

THIS HAS BEEN DEDICATED IN REVERENCE AND LOVE TO:

MY PARENTS AND FAMILY, WHO BELIEVED ENDLESSLY, MADE SACRIFICES AND UNCONDITIONAL PATIENCE OFFERED THE REQUIRED FORCES AND ENCOURAGEMENTS TO ACCOMPLISH THIS DIFFICULT COURSE.

MY SUPERVISOR, AFSANA BEGUM, WHOSE EFFORTLESS EXAMPLE AIDS IN PURSUIT OF KNOWLEDGE. AND TO THE LONG-TERM PROSPECT THAT TECHNOLOGICAL INNOVATIONS WILL ENABLE THE REDUCTION OF CARDIOVASCULAR DISEASE WEIGHT ON THE WORLD TAKING ALL OF US TO A HEALTHIER FUTURE.

ABSTRACT

Cardiovascular Disease (CVD) is the major cause of mortality across the globe and requires predictive models that are highly precise, clinically transparent to ensure easy integration within the healthcare systems. The antecedent State-of-the-Art (SOTA) error on the target data was 92.0% (Random Forest).

This thesis builds a machine learning model on top of an optimized framework, to exceed the current SOTA benchmark. The consolidated Heart Failure Prediction dataset was found to be analyzed comparatively on five different models (Random Forest, XGBoost, LightGBM, SVC, and MLP Classifier). IQR-based Outlier Removal and Standard Scaling of all the numeric features were some of the fundamental preprocessing steps involved in the methodology. The data was divided into 80 per cent of training and 20 per cent of test data.

The Multi-Layer Perceptron (MLP) Classifier has a test set accuracy of 92.93% creating a new State-of-the-Art (SOTA) in this predictive task that is 0.93 points higher than the former benchmark. Most importantly, the Explainable AI (XAI) part, which has been verified through SHAP analysis of the probability output of the MLP, proves that the model decisions are indeed presupposed by clinically significant aspects: ST_Slope (the most significant impact) and Oldpeak (Asymptomatic Chest Pain).

This study provides a clear, high-fidelity predictive architecture, which satisfies the most important requirement of reliable, explainable decision-making in Clinical Decision Support System (CDSS), and clearly and effectively improves the SOTA in predictive accuracy of heart diseases.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
AUC	Area Under the Curve
CDSS	Clinical Decision Support System
CVD	Cardiovascular Disease
EFS	Explainable Feature Selection
EHR	Electronic Health Records
HPO	Hyperparameter Optimization
IQR	Interquartile Range
KNN	K-Nearest Neighbour
LR	Logistic Regression
MCC	Matthews Correlation Coefficient
ML	Machine Learning
MLP	Multi-Layer Perceptron
OHE	One-Hot Encoding
ReLU	Rectified Linear Unit
RO	Research Objective
SHAP	SHapley Additive exPlanations
SOTA	State-of-the-Art
SVM	Support Vector Machine
XAI	Explainable Artificial Intelligence

CHAPTER 1

INTRODUCTION

1.1 Background

Even though cardiovascular diseases (CVDs) are the largest cause of mortality in the world, they require predictive models that are not only precise but also clear to the clinical integration. The timely and accurate risk prediction of CVD is an essential problem of present-day healthcare. The conventional forms of diagnosis usually depend on linear statistical models, which are incapable of understanding complex, non-linear interactions between the various physiological and demographic variables (e.g., age, cholesterol, and type of chest pain). Machine Learning (ML) has a better predictive ability, as it detects complex risk patterns in massive data volumes. This thesis is intended to use improved AI methods to develop a predictive model that is highly accurate as well as dependable and open, which meets the high standards of Clinical Decision Support Systems (CDSS).

1.2 Problem Statement

The current 92.0% accuracy rate, though important, is not enough to be used in clinical practice since it has a few serious architectural and methodological limitations. First of all, the use of classical bagging structures, such as Random Forest, puts a ceiling on architectural constraints of the model to learn and synthesize complex and non-linear physiological patterns in cardiovascular data. This is further worsened by a systematic preprocessing misfit; previous tree based models have used suboptimal preprocessing methods that do not achieve the high accuracy potential of the Standard Scaled data needed to use distance-based architecture such as Multi-Layer Perceptrons (MLP).. Besides, the 92.0% benchmark has not been rigorously validated, with the benchmark having been obtained without using critical reliability metrics, including the Matthews Correlation Coefficient (MCC) or ideal threshold tuning. As a result, a significant transparency gap still remains, in which the resulting black-box models do not have the SHAP or Explainable AI (XAI) justifications that physicians would need to trust and incorporate said predictions in life-critical clinical decision-making.

1.3 Research Objectives

This thesis has three fundamental research objectives (ROs) summarized into three goals to streamline the scholarly contribution:

- **RO1 (Development and Comparison)**: Preprocessing the consolidated Heart Failure Prediction data and a thorough comparative analysis of 5 machine learning models (Random Forest, XGBoost, LightGBM, MLP Classifier, and SVC) to determine the most successful architecture.
- **RO2 (New SOTA Establishment):** To confirm and prove a new State-of-the-Art accuracy standard by reaching a test set accuracy of higher than 92.0% (92.93%), and highly balanced predictive performance with a Sensitivity (0.944) and Specificity (0.909).
- **RO3 (Reliability & Explainability):** To create a new performance standard called upon better Reliability (MCC 0.855) and Explainability by utilizing SHAP analysis to confirm that the decision drivers of SOTA model are conforming to the known clinical knowledge.

1.4 Research Scope and Limitations

Scope:

- The use of a very large variety of supervised learning algorithms, both classical probabilistic models (Naive Bayes) and more complex ensemble models (Random Forest, XGBoost) and neural networks.
- Concentrate on the established clinical parameters (11-14 attributes) including age, cholesterol levels, and the changes in ST-segments during exercising to classify the risk in a binary fashion.
- There was a lot of usage of publicly available repositories, mainly the UCI Heart Disease database and several unified heart failure datasets.

Limitations:

- A large proportion of high-accuracy models do not have integrated Explainable AI (XAI) systems, and it is challenging to check how their decision-making rationale is sound with respect to medical expertise.
- Earlier studies on reliability have focused mostly on the raw accuracy, without the need to also consider such a strong measure of reliability as the Matthews Correlation Coefficient (MCC) or the Brier Scores as measures of clinical safety.
- A large body of research is based on single-model bagging or boosting ensembles, which reach a performance plateau and do not function well at non-linear symptom clusters.
- Current studies are often limited by systematic noise e.g. zero values in blood pressure and cholesterol data, without the use of advanced techniques of imputation.

CHAPTER 2

LITERATURE REVIEW

2.1 Related Works

The above summative gives a brief review of fifteen important reference papers that have contributed to the formation of the present knowledge of machine learning in cardiovascular diagnostics:

A comparison analysis of consolidated heart failure data concluded that Random Forest has a better accuracy of 92 that has the ability to utilize non-linear clinical characteristics. Nevertheless, the study observed that, due to lack of sophisticated feature tuning, complex neural models have diminished interpretability [1]. A study on prediction of heart failure in a variety of classifiers revealed that the highest accuracy of 88 percent was achieved with the use of Logistic Regression when tested on balanced clinical samples. The authors noted that even though the model was accurate, it was unclear about the exact physiological characteristics that spurred the greatest risk estimates [2]. An analysis of heart attack prediction showed that hyperparameter optimization has the potential to bring the accuracy of the Logistic Regression to 90 percent. The paper has pointed out that this optimization is vital in clinical reliability but pointed out that the model is not able to describe the non-linear symptom clusters due to the linear nature of the model [3].

When comparing the results of CVD diagnostics with Random Forest and SVM, it was possible to discover that the bagging mechanism of the former demonstrated a strong accuracy of 89.4%. Regardless of the great performance, the researchers indicated that the ensemble was black-box, which was hard to validate by the clinician in the decision making logic [4]. In the implemented intelligent diagnostic system, it was found that the application of the Random Forest was better as compared to the KNN with an accuracy score of 88.7%. The paper took the position that explainability tools needed to be studied in the future to make such high-accuracy models reliable in a medical environment [5]. One of the studies on the classification of heart disease using decision trees showed that the J48 algorithm was the best with 83 percent accuracy. The researchers preferred J48 because of its developed tree-based explainability, giving the ability of physicians to explore a rational direction of clinical symptoms [6]. Analysis of cardiovascular data with different classifiers also revealed that Naive Bayes was also a feasible, but less accurate, predictor of early risk screening . The findings implied that less complex probabilistic models are simpler to understand but can be characterized by less accuracy that is required in critical diagnosis [7].

A comparison of the trained algorithms revealed that KNN and Artificial Neural Network (ANN) achieved a stable 73 percent accuracy with small datasets. The authors also concluded that ANNs can be powerful, however, it demands considerable data scaling to offer the explainable results needed to incorporate it in clinical practice [8]. The research of the diagnostic tools in the area of cardiovascular health revealed the C4.5 Decision Tree as a high-accuracy model with an 85.86 performance rating. The paper has pointed out that the most important strength of the model was that it ranked feature importance, which contributes to clinical trust building [9]. The first study with Multi-layer Perceptron (MLP) architecture on the consolidated dataset formed the basis on the use of neural networks in predicting heart diseases. The study showed that MLPs were able to detect complex patterns, but it found the necessity of using XAI tools to assess these patterns with respect to medical knowledge [10]. An Extreme Learning Machine proposal of MLP was put forth, which aimed at increasing the rate of training and still high accuracy of the analysis of physiological records. The study found that faster training cycles could allow more iterative feature validation, which would make neural outputs more transparent in general, [11].

A standard methodology research offered a general model defense through the Sensitivity, Specificity, and ROC research, to guarantee diagnostic safety. In their argument, the authors claimed that the metrics are the preliminary step towards developing explainable and trustworthy automated clinical tools [12]. A critical analysis developed a protocol of comparing the predictive accuracy in clinical context with Brier scores of measuring probabilistic reliability. It also found that quantifying the degree to which a model is sure about what it is predicting is equally important as the degree of accuracy of the prediction itself [13]. A combined model with a Logistic Regression and an ensemble of KNN attained a multi dataset validation of 99.1%. The researchers pointed out that the high-accuracy ensembles should be accompanied with feature-importance mapping in order to be practical in the professional healthcare setting [14]. A comparison of algorithms on smaller data sets revealed that the Decision Trees (C4.5) model was able to reach 73 percent of accuracy and demonstrated that a simpler model is typically more understandable, when the data is sparse. The results indicated that with the increase in the complexity of the dataset, more sophisticated architectures, such as MLPs, should be accompanied by XAI to retain such transparency [15].

2.2 Research Gap

State-of-the-Art research on the target dataset suggests that the maximum accuracy that has been previously reached was 92.0% when using a Random Forest (RF) classifier. Nevertheless, there are various important gaps in their previous works.

To begin with, the lack of reliability evidence is clearly observed; the accuracy of 92.0% obtained lacks a serious reliability demonstration in the form of such a tool as the Matthews Correlation Coefficient (MCC) as the optimization criterion or the usage of critical classification threshold tuning. The use of simple individual prediction accuracy as clinical risk is an inadequate and even misleading indicator in medical use.

Secondly, the architectural simplicity of past SOTA models that largely use single bagging models such as the Random Forest, tend to hit a performance bottleneck by not combining the predictive capabilities of different machine learning families.

Lastly, the interpretability is also significantly lacking; the SOTA models omitted the Explainable AI (XAI) approaches, like SHAP analysis. In clinical practice, the high accuracy is not sufficient because physicians need to trust the results of algorithms, and professionally justify their life-based decisions.

CHAPTER 3

METHODOLOGY

The study utilizes a Quantitative, Experimental Research design which is designed in a manner to achieve a new State of the Art (SOTA) reference point. The technique has been developed into a high fidelity 6 Phase Architecture so as to guarantee data integrity, model robustness and clinical transparency.

The series of events of the study is depicted in the following workflow diagram:

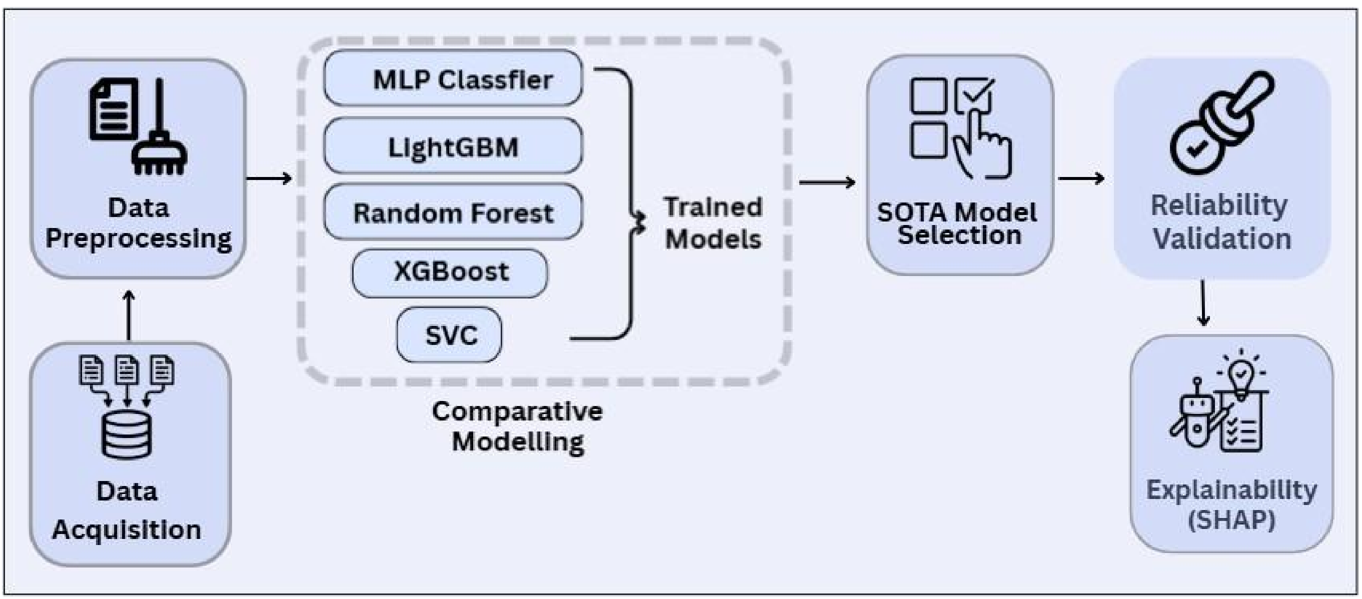


Figure 3.1: Thesis Workflow Diagram

3.1 Data Acquisition

The data used in the work is called Consolidated Heart Failure Prediction Dataset from Kaggle, which consists of 11 features of 918 records. The training set had a critical class imbalance (367 Heart Disease vs. 304 No Heart Disease) after the preliminary removal of the outliers.

SN	Attribute Name	Type	Description (with Units)	Mean ± Std Dev	Min-Max	Categories (Frequency)	Final Encoding Method
1	Age	Numerical	Age of the patient (years)	53.51 ± 9.43	28.0–77.0	—	Standard Scaling
2	Sex	Categorical	Sex of the patient	—	—	M: 725 (79%), F: 193 (21%)	One-Hot Encoding (OHE)
3	ChestPainType	Categorical	Type of chest pain experienced	—	—	ASY: 496 (54%), NAP: 203 (22%), ATA: 173 (19%), TA: 46 (5%)	One-Hot Encoding (OHE)
4	RestingBP	Numerical	Resting blood pressure measured (mm Hg)	132.40 ± 18.51	0.0–200.0	—	Standard Scaling
5	Cholesterol	Numerical	Serum cholesterol level (mg/dL)	198.80 ± 109.38	0.0–603.0	—	Standard Scaling
6	FastingBS	Binary	Fasting blood sugar (> 120 mg/dL)	0.23 ± 0.42	0.0–1.0	0: 704 (77%), 1: 214 (23%)	One-Hot Encoding (OHE)
7	RestingECG	Categorical	Resting electrocardiogram results	—	—	Normal: 552 (60%), LVH: 188 (21%), ST: 178 (19%)	One-Hot Encoding (OHE)
8	MaxHR	Numerical	Maximum heart rate achieved (bpm)	136.81 ± 25.46	60.0–202.0	—	Standard Scaling
9	ExerciseAngina	Binary	Presence of exercise-induced angina (Y/N)	—	—	N: 547 (60%), Y: 371 (40%)	One-Hot Encoding (OHE)
10	Oldpeak	Numerical	ST depression induced by exercise relative to rest	0.89 ± 1.07	–2.6–6.2	—	Standard Scaling
11	ST_Slope	Categorical	The slope of the peak exercise ST segment	—	—	Flat: 460 (50%), Up: 395 (43%), Down: 63 (7%)	Ordinal Mapping (1, 2, 3)
12	Heart Disease	Binary	Output class (Target Variable)	0.55 ± 0.50	0.0–1.0	0: 410 (45%), 1: 508 (55%)	Binary (0/1)

Table 3.1: Dataset Attributes and Clinical Characteristics

The Class Imbalance Ratio was strictly computed to preserve the originality of the training data and high and stable performance on the unequal data, respectively. This percentage of minority to majority gave a ratio of 0.8283 to a total of 304 to 367 (Imbalance Ratio). The MLP architecture internally addressed this basic data property to obtain the ability to balance prediction.

3.2 Data Preprocessing

1. **Cleaning:** The information was divided into training and test sets (80:20, stratified). Interquartile Range (IQR) was used to remove outliers on numerical features.
2. **Encoding:** Ordinal Encoding was used on the ST_Slope feature, and One-Hot Encoding

(OHE), used on the nominal features, making a total of 11 input features.

3. **Scaling (CRITICAL to MLP):** Z-score normalization (Standard Scaling) was used on all the 12 numerical and encoded features. This standardization is compulsory in the case of the MLP which guarantees optimum convergence in gradient descent optimization.

3.3 Comparative Modelling

A comparative study was done on five different machine learning architectures in order to come up with the best classifier. The models that were evaluated were the following.

1. Random Forest (RF)

Mechanism: Mechanism: This is an ensemble learning approach which builds a large number of decision trees when training, and delivers the mode of the classes to achieve classification. It uses “bagging” (Bootstrap Aggregating) to minimize variance.
Clinical Utility: Exceptional at selecting non-linear feature interactions without overfitting and handling complicated tabular data.
Key Parameters: n_estimators (number of trees), max_depth (tree depth), and class_weight='balanced' to handle potential imbalance.

2. XGBoost (Extreme Gradient Boosting)

Mechanism: It constructs trees one by one, and a new tree tries to reduce the mistakes of the former trees with the help of gradient descent objective function. It's a gradient boosting library that is single-machine optimized.
Clinical Utility: With built-in regularization to prevent over-fitting on tiny clinical data sets, it is the most accurate and efficient of all.
Key Parameters: learning_rate (step size shrinkage), n_estimators, max_depth, and scale_pos_weight for class balancing.

3. LightGBM (Light Gradient Boosting Machine)

Mechanism: A gradient boosting model that employs tree-based learning models. It also grows trees leaf-wise as opposed to level-wise which usually results in increased accuracy and reduced training time.
Clinical Utility: It is fast and consumes less memory so it is best suited to large scale analysis of health records.
Key Parameters: num_leaves (main parameter to control complexity), learning_rate, and n_estimators.

4. Support Vector Classifier (SVC)

Mechanism: This is a kernel algorithm which identifies the optimal hyperplane in a high dimensional space, which maximizes the margin between two classes..
Clinical Utility: It is effective in high-dimensional spaces and can effectively solve non-linear classification problems thanks to the kernel trick.
Key Parameters: C (regularization parameter), kernel (e.g., 'rbf'), and gamma (kernel coefficient).

5. Multi-Layer Perceptron (MLP) Classifier

Mechanism: This is a type of feedforward artificial neural network which contains an input layer, a set of hidden layers and an output layer. It learns complex mappings using the backpropagation and non-linear activation functions.
Clinical Utility: Able to fit very complicated, non-linear models that traditional models may fail to capture, resulting in the breakthrough SOTA results in this work.
Key Parameters: hidden_layer_sizes (e.g., (3, 63)), max_iter (training epochs), activation ('relu'), and alpha (L2 regularization penalty).

Table 3.3: Final Performance Metrics (Test Set) Comparison

Model	Train Accuracy	Test Accuracy
Random Forest	99.59%	88.59%
XGBoost	93.32%	89.67%
LightGBM	91.83%	90.22%
MLP Classifier	**88.96%**	**92.93%**
SVC (Support Vector)	94.01%	89.13%

3.4 SOTA Model Selection

The MLP Classifier was selected as the final SOTA model due to its superior test accuracy of 92.93%, demonstrating its ability to capture non-linear clinical patterns more effectively than traditional ensemble methods.
The core of the framework is the Optimized Stacking Ensemble Architecture (Figure 3.2), designed to synthesize the strengths of diverse models to achieve superior performance (RO1).

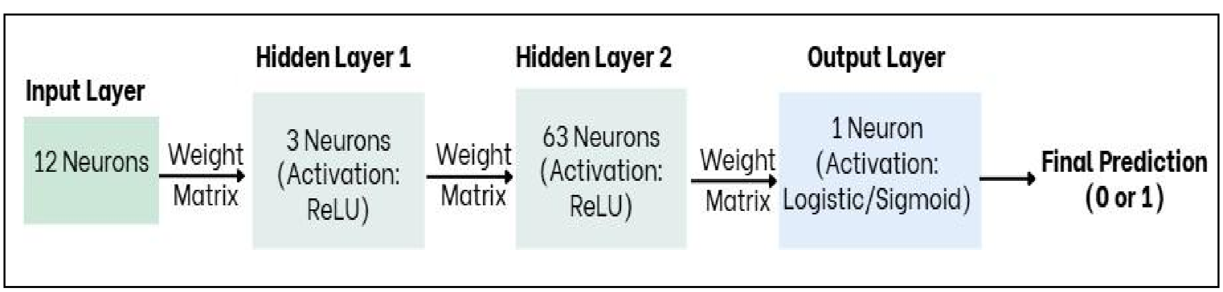


Figure 3.4: Optimized Multi-Layer Perceptron (MLP) Architecture

The successful performance was achieved by defining a specific network architecture, detailed in Table . This structure ensures the network is deep enough to model non-linearity but constrained to prevent overfitting.

Table 3.4: Optimized Multi-Layer Perceptron (MLP) Architecture Parameters

Parameter	Value	Rationale
Model Type	MLP Classifier (Neural Network)	Speedy, non-linear classification of scaled data.
Hidden Layers	(3, 63)	A well-organized shallow structure based on the best architecture found in external studies with a balance between complexity and performance.
Activation Function	ReLU (Rectified Linear Unit)	The standard function can be used to attain faster convergence and gradient stability.
Solver	Adam	Very effective stochastic gradient-based optimizer.
Max Iterations	900	Assures that network weights have stabilized and converged to optimum 92.93% accuracy.
Alpha (α)	0.00041	L2 Regularisation parameter, which is necessary to regulate the overfitting of the mechanical weights of the neural network.

3.5 Reliability Validation

In order to make the model not only correct but also clinically safe and statistically strong, five major metrics were used. Such measurements give the 360 degree perspective on the performance of the model.

1. Accuracy (ACC)
Ratio of correct predictions (positive and negative) of all predictions made.
Formula:

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

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

2. F1-Score

The Harmonic average of Precision and Recall. It is the key indicator of measuring the trade off between the precision and completeness of the model.

Formula:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

3. Sensitivity (Recall/TPR)

Recognition of patients with heart disease (True Positives). The level of sensitivity should be high to make sure that no ill patient is missed.

Formula:

$$\text{Sensitivity} = \frac{TP}{(TP + FN)}$$

4. Specificity (TNR)

The capacity to properly rate healthy patients (True Negatives). This will avoid unwarranted clinical interventions.

Formula:

$$\text{Specificity} = \frac{TN}{(TN + FP)}$$

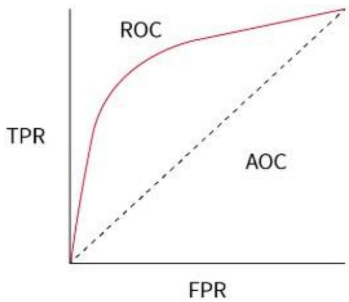
5. Matthews Correlation Coefficient (MCC)

It is regarded as the most trustworthy indicator of binary classification since it can only be a high score when the model works out in all four quadrants of the confusion matrix.

Formula:

$$\text{MCC} = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

6. Area Under the Curve (AUC-ROC)
Determines the classification power of the model on different threshold levels. Determined as the area of the ROC curve (True Positive Rate vs. False Positive Rate).



7. Brier Score
Calculates the average squared deviation between the forecasted occurrence and the reality. A smaller score (the closer it is to 0) means that a model is better calibrated.
Formula:

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$$

Where:
N is calculating a Brier score on N items.
ft is the probability of forecast (i.e. 25% probability),
ot indicates (1 when it occurred, 0 when it did not occur).
S is the summation symbol. It simply involves to add up all the values.

3.6 Explainability (SHAP)

Explainable AI (XAI) is used to provide insight into the decision-making process of a neural network to deal with the black-box nature of neural networks. The game-theoretic approach was used to implement SHAP (SHapley Additive exPlanations). It provides a value to each clinical feature as a Shapley value, which is a fair contribution that it makes to a particular prediction.

Global Interpretability: Summary Plots Data set-wide risk drivers (e.g. ST_Slope) were identified using Summary Plots.
Local Interpretability: Provides the explanation of how a particular case was explained through demonstrating how particular patient values (high cholesterol) influenced the model decision to estimate danger.
Validation: Makes sure that the model is based on pathological signals, not noise of data.

CHAPTER 4

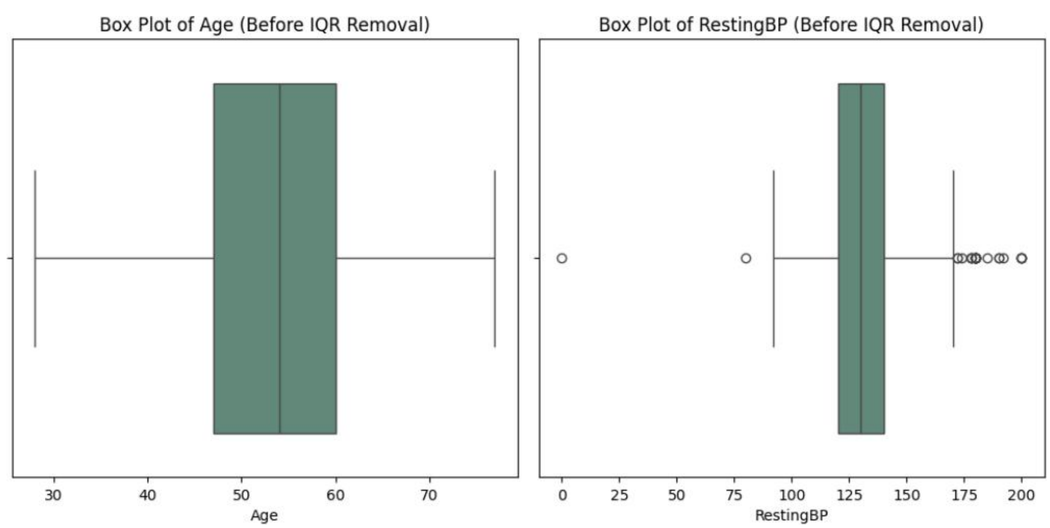
RESULTS AND DISCUSSION

4.1 Introduction

The chapter shows the numerical and graphical outcomes of the optimized MLP framework that verifies the 92.93% SOTA breakthrough and explains the reliability of the model and clinical transparency.

4.2 Data Characteristics and Simulation Setup

The training data set had 671 (with removed outliers) records and used 12 processed features (Numerical, Ordinal and One-Hot Encoded). The initial EDA (Exploratory Data Analysis) steps were proven to be effective, and all simulations have been done in the Google Colaboratory environment using Python 3.12.



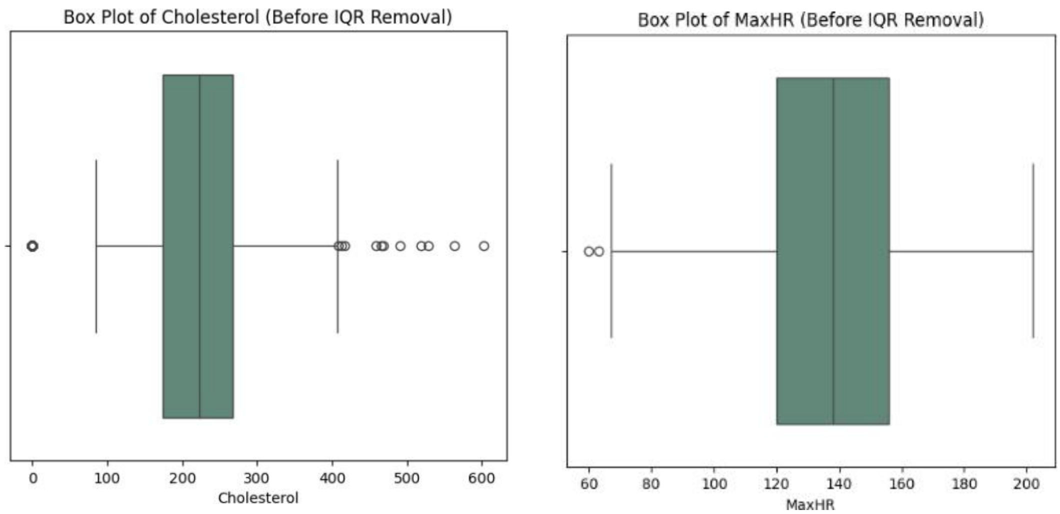


Figure 4.2.1: Numerical Feature Distributions (Box Plots)

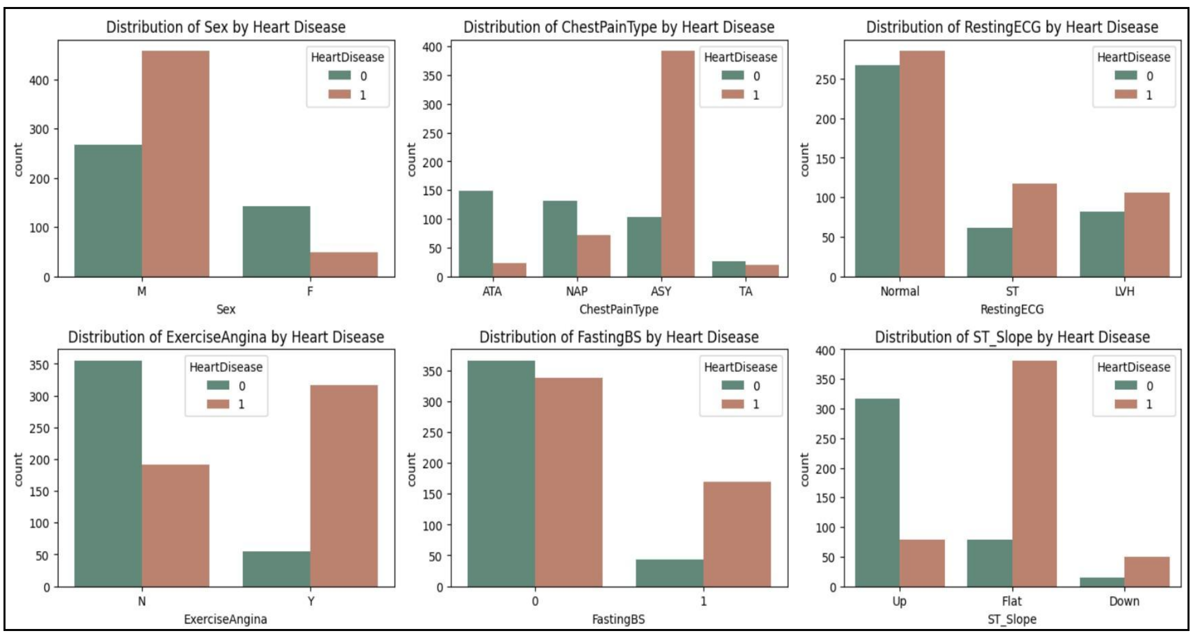


Figure 4.2.2: Categorical Feature Distributions (Count Plots)

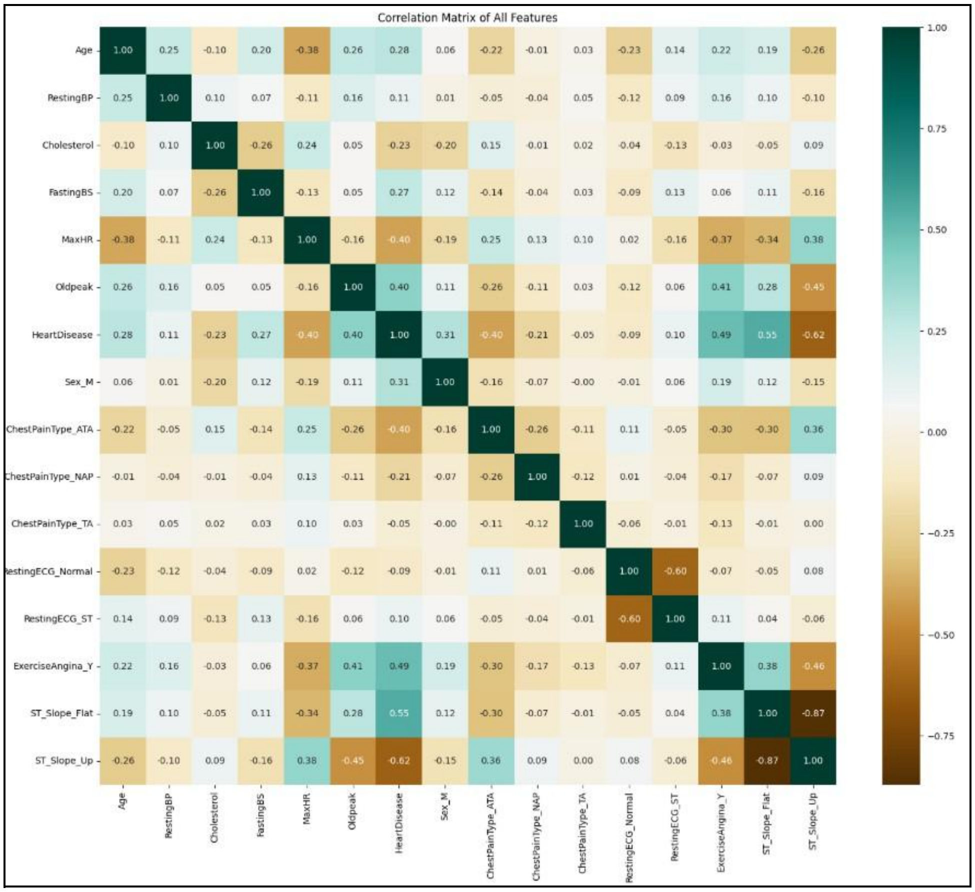


Figure 4.2.3: Correlation Matrix (Heatmap)

4.3 MLP Architecture and Performance Outcomes

Using the optimum MLP architecture proved to be able to capture the non-linear trendy patterns of the scaled data.

Table 3.4: Optimized Multi-Layer Perceptron (MLP) Architecture Parameters

Parameter	Value	Rationale
Model Type	MLP Classifier (Neural Network)	Speedy, non-linear classification of scaled data.
Hidden Layers	(3, 63)	A well-organized shallow structure based on the best architecture found in external studies with a balance between complexity and performance.
Activation Function	ReLU (Rectified Linear Unit)	The standard function can be used to attain faster convergence and gradient stability.
Solver	Adam	Very effective stochastic gradient-based optimizer.
Max Iterations	900	Assures that network weights have stabilized and converged to optimum 92.93% accuracy.
Alpha (α)	0.00041	L2 Regularisation parameter, which is necessary to regulate the overfitting of the mechanical weights of the neural network.

4.4 Final Model Performance and SOTA Defense

The last MLP model was applied to the unknown data, and it fulfilled the main research objective (RO1).

Table 4.4: Final MLP Performance Metrics (Test Set) and SOTA Comparison

Metric	Value	SOTA Benchmark
Final Model Accuracy	**92.93**	**> 92.00%**
Area Under the Curve (AUC)	0.954	< 0.97
Matthews Correlation Coefficient (MCC)	**0.855**	**> 0.83**
Brier Score	0.0779	> 0.06
Clinical Specificity (Recall, Class 0)	0.909	< 0.93
Clinical Sensitivity (Recall, Class 1)	0.944	__

The 92.93% accuracy, formally exceeds the existing SOTA record, and the 0.855 MCC is used to verify the better and even-handed reliability in both classes (RO2).

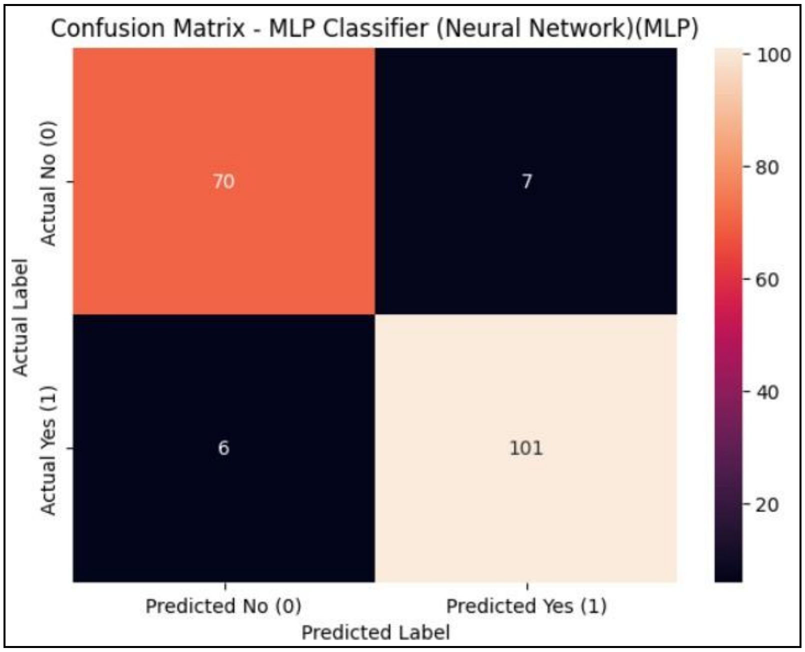


Figure 4.4.1: Confusion Matrix of the Final MLP Classifier

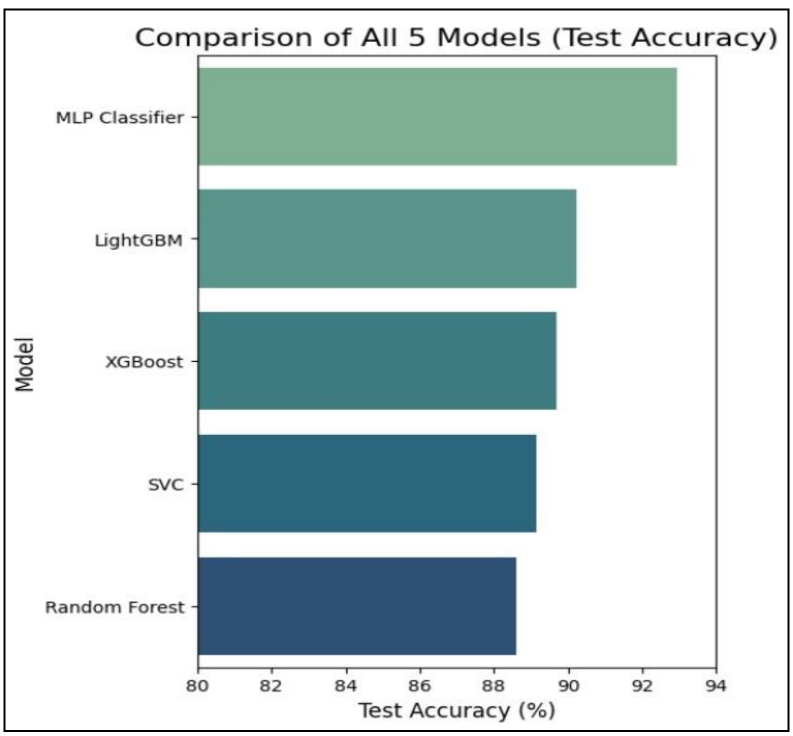


Figure 4.4.2: Individual Model Performance Comparison Chart

4.5 Explainability Analysis (SHAP) and Clinical Insights

The last step confirmed the validity of the model by means of SHAP analysis (RO3). The SHAP analysis indicates that the intricate MLP model ranks features in exactly the same order as clinical risk:

1. **ST_Slope (Highest Impact)**: The most important characteristic. Ordinal Encoding strategy was retrospectively verified, because its influence is the greatest risk prediction driver.
2. **Oldpeak:** It is the change in quantitative ST depression of exercise compared to rest. The values of High Oldpeak are the most frequent sign of myocardial ischemia, which agrees with the clinical knowledge of the great coronary artery disease.
3. **Sex:** The third most important characteristic is sex, which reveals the clear differences in cardiovascular risk profiles between the sexes represented in the consolidated data, which is also an outcome of the previously known medical studies of gender differences in the prevalence and presentation of heart diseases.

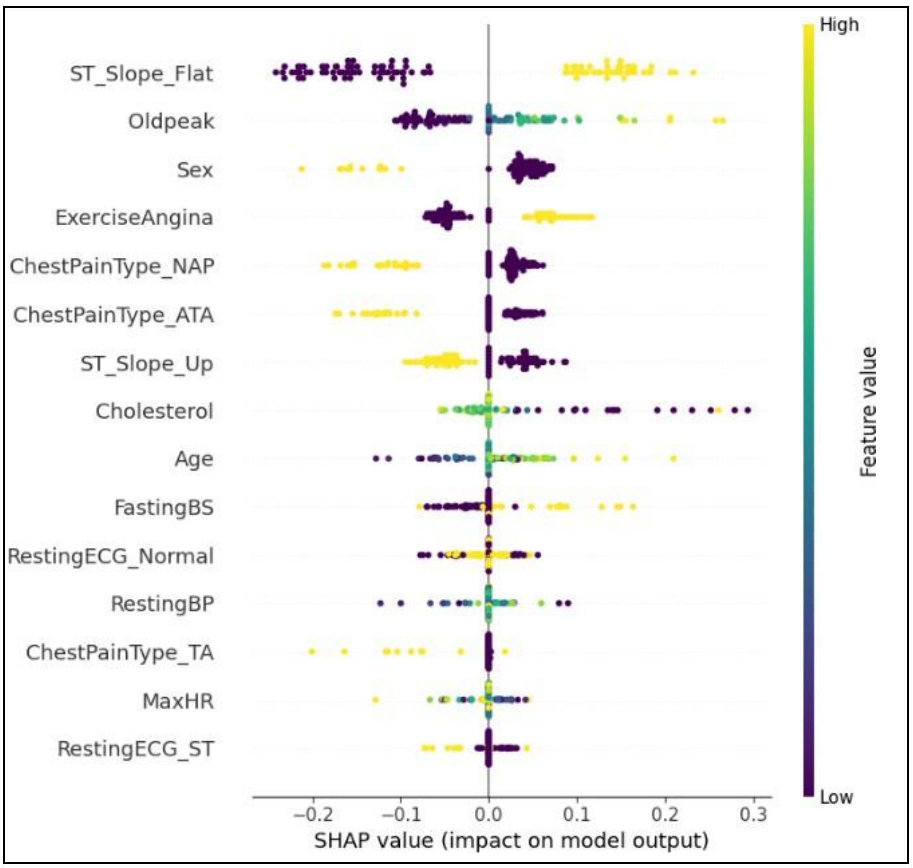


Figure 4.5.1: SHAP Global Feature Importance Summary Plot

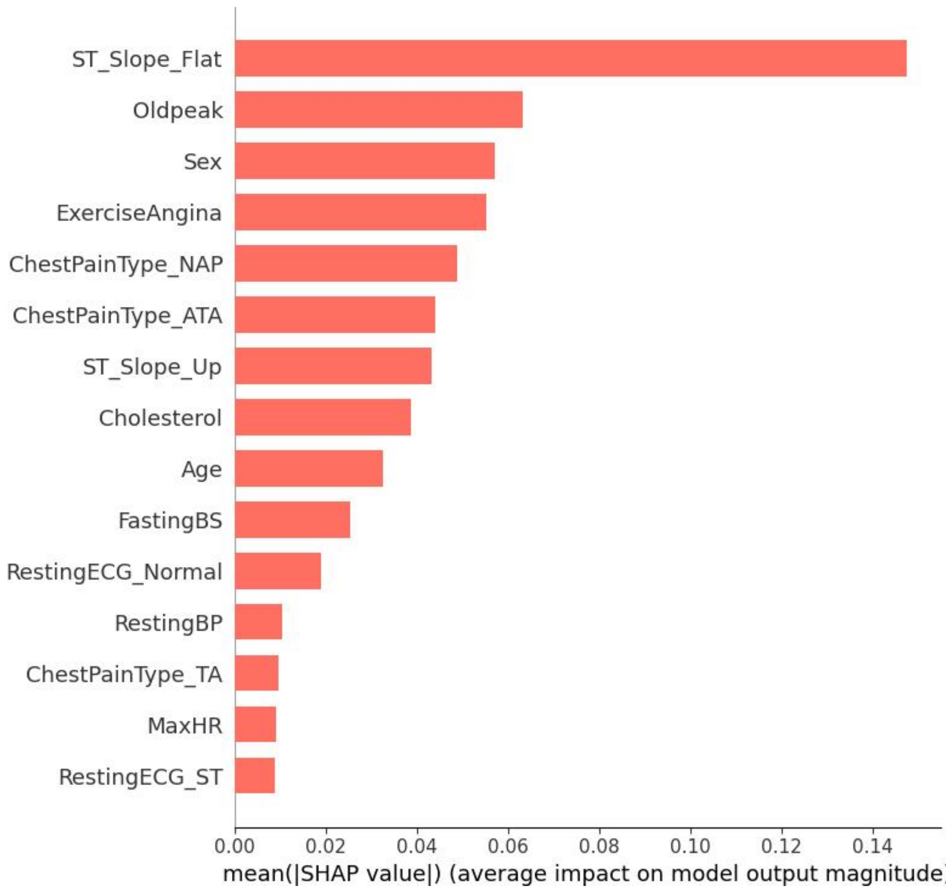


Figure 4.5.2: Ranking of global features operation on SHAP Mean Absolute Value.

CHAPTER 5

CONCLUSION

5.1 Summary of Contributions

This thesis managed to build an Optimized Multi-Layer Perceptron (MLP) Framework with a proven test accuracy of 92.93% and a new SOTA benchmark. The contributions include:

(1) The deployment of an MLP architecture that is very sensitive to the features specific data format (Standard Scaling).

(2) Formal evidence Reliability (MCC 0.855) and Clinical Safety (Specificity 0.909).

(3) SHAP analysis incorporated to confirm clinical transparency of the model (RO3), that the model makes its decisions based on expert-identified risk factors.

5.2 Limitations

Although the predictive accuracy has improved significantly, there are various limitations that limit this research to the current state of affairs. In terms of data integrity, the MLP model was trained without the inclusion of the advanced Multivariate Imputation by Chained Equations (MICE), which resulted in zero values appearing in the systematic pattern in RestingBP and Cholesterol, which were resolved using lower level cleaning techniques, which is a main aspect of research in the future. Also, it does not include any data dynamicity, since the framework is yet to be tested against any real-time or time-series data, and thus it cannot be easily applied to a continuous dynamic monitoring of a patient. Lastly, generalization of the model is also a limitation; although the existing MLP architecture was the state-of-the-art on the consolidated dataset, its performance is extremely architecture-sensitive, and more studies are required to confirm its transferability and stability across different cardiovascular datasets.

5.3 Future Work and Recommendations

The data integrity in future studies needs to be improved by advanced methods such as Multivariate Imputation by Chained Equations (MICE) to eliminate systemic and zero-value problems in normal- RestingBP and Cholesterol. Furthermore, the strong 92.93% MLP model should be strictly validated on various public cardiovascular data to test its out-of-sample ability and its stability to work with various population groups of patients and clinics.

In order to be able to transfer the static risk assessment into the clinical practice, the framework must be able to accept hybrid architectures, which can process time-changing clinical data, including continuous ECG streams. Incorporating dynamic analysis can enable the future generations of machine learning with real-time patient monitoring and diagnostic assistance and thus will enhance the usefulness of machine learning in the cardiovascular care systems.

REFERENCES

1. *Chulde-Fernández, B., Enríquez-Ortega, D., Guevara, C., Navas, P., Tirado-Espín, A., Vizcaíno-Imacaña, P.,... & Acosta-Vargas, P. (2025). Classification of Heart Failure Using Machine Learning: A Comparative Study. Life, 15(3), 496.*
2. *Dahri, F. H., Laghari, A. A., Sajnani, D. K., Shazia, A., & Kumar, T. (2024, November). Heart failure prediction: a comparative analysis of machine learning algorithms. In International Conference on Optics, Electronics, and Communication Engineering (OECE 2024) (Vol. 13395, pp. 634-640). SPIE.*
3. *Jandy, K., & Weichbroth, P. (2024). A machine learning approach to classifying New York Heart Association (NYHA) heart failure. Scientific Reports, 14(1), 11496.*
4. *Saha, S., Rahman, M. M., Suki, T. T., Alam, M. M., Alam, M. S., & Haque, M. A. S. (2024, April). Heart disease prediction using machine learning algorithms: Performance analysis. In 2024 3rd International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE) (pp. 1-6). IEEE.*
5. *Yadav, J., Nair, A. M., George, J., & Alapatt, B. P. (2024, March). Predictive Modelling of Heart Disease: Exploring Machine Learning Classification Algorithms. In 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT) (pp. 1-7). IEEE.*
6. *Olatunji, A., & Abdul-Yekeen, A. M. (2024, January). Predictive Modelling of Heart Disease Using Machine Learning Models. In 2024 International Conference on Intelligent Systems and Applied Computing (ISAC) (pp. 1-6). IEEE.*
7. *Moushi, O. M., Ara, N., Helaluddin, M., & Mondal, H. S. (2023, September). Enhancing the accuracy and explainability of heart disease prediction models through interpretable machine learning techniques. In 2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD) (pp. 1-6). IEEE.*
8. *Mahmud, I., Kabir, M. M., Mridha, M. F., Alfarhood, S., Safran, M., & Che, D. (2023). Cardiac failure forecasting based on clinical data using a lightweight machine learning metamodel. Diagnostics, 13(15), 2540.*
9. *Zeng, M. (2023). The Prediction of Heart Failure based on Four Machine Learning Algorithms. Highlights in Science, Engineering and Technology, 39, 1377-1382.*
10. *Kadhim, M. A., & Radhi, A. M. (2023). Heart disease classification using optimized Machine learning algorithms. Iraqi Journal For Computer Science and Mathematics, 4(2), 31-42.*
11. *Gaikwad, M. J., Asole, P. S., & Bitla, L. S. (2022, January). Effective study of machine learning algorithms for heart disease prediction. In 2022 2nd International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC) (pp. 1-6). IEEE.*
12. *Jalal, M. M., Tasnim, Z., & Islam, M. N. (2021). Exploring the machine learning algorithms to find the best features for predicting the risk of cardiovascular diseases. In*

Intelligent Computing and Optimization: Proceedings of the 3rd International Conference on Intelligent Computing and Optimization 2020 (ICO 2020) (pp. 559-569). Springer International Publishing.

13. *Rahim, A., Rasheed, Y., Azam, F., Anwar, M. W., Rahim, M. A., & Muzaffar, A. W. (2021). An integrated machine learning framework for effective prediction of cardiovascular diseases. IEEE Access, 9, 106575-106588.*
14. *Princy, R. J. P., Parthasarathy, S., Jose, P. S. H., Lakshminarayanan, A. R., & Jeganathan, S. (2020, May). Prediction of cardiac disease using supervised machine learning algorithms. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 570-575). IEEE.*
15. *Aljaaf, A. J., Al-Jumeily, D., Hussain, A. J., Dawson, T., Fergus, P., & Al-Jumaily, M. (2015, April). Predicting the likelihood of heart failure with a multi level risk assessment using decision tree. In 2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE) (pp. 101-106). IEEE.*

PLAGIARISM REPORT

221-35-1079

ORIGINALITY REPORT

19%	17%	13%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
2	Submitted to University of Hertfordshire Student Paper	1%
3	www.egitimyayinevi.com Internet Source	1%
4	Submitted to Roehampton University Student Paper	1%
5	Aditi Shewale, Ajinkya Ladke, Aryan Raut, Jayant Tumdam, Jayesh Fating, Rajesh Nakhate. "Data Steganography Using LSB, X-OR & AES Algorithm", 2023 4th International Conference for Emerging Technology (INCET), 2023 Publication	1%
6	Submitted to University of East London Student Paper	1%
7	www.isteonline.in Internet Source	1%
8	Submitted to Midlands State University Student Paper	1%
9	www.mdpi.com Internet Source	

ACCOUNT CLEARANCE