*Malaysian Journal of Computing*, 10 (2): 2159-2175, 2025
Copyright © UiTM Press
eISSN: 2600-8238

# IMPACT OF FEATURE STANDARDIZATION ON HEART DISEASE PREDICTION: A COMPARATIVE ANALYSIS OF LOGISTIC REGRESSION AND SUPPORT VECTOR MACHINE MODELS

**Norsyela Muhammad Noor Mathivanan[1*], Eric Foo Zhi Xian[2], Debbie Foo Yong Xi[3], Chua Hiang Kiat[4]**

*[1*,2,3,4]School of Computing and Creative Media, University of Wollongong Malaysia, 40150 Shah Alam, Malaysia*

*[1,4]UOW Malaysia KDU Penang University College Penang, 10400 George Town, Pulau Pinang, Malaysia*

[1*]norsyela.m@uow.edu.my, [2]0133702@student.uow.edu.my, [3]0136275@student.uow.edu.my, [4]hk.chua@uow.edu.my

## ABSTRACT

*Cardiovascular diseases are among the leading causes of global mortality. Heart disease, in particular, remains a major contributor to this burden, highlighting the need for effective predictive models to enable early detection. This study investigates the impact of feature standardization using StandardScaler on the performance of two prominent machine learning models involving Logistic Regression (LR) and Support Vector Machine (SVM) for predicting heart disease. The research utilizes a dataset comprising demographic and clinical attributes of patients, focusing on the role of feature standardization in enhancing model performance. The study compares models trained on raw data and standardized data, applying performance metrics such as accuracy, precision, recall, and F1-score. Results indicate that feature standardization significantly improves the performance of both models. LR showed a clear enhancement in macro F1-score on the testing set, rising from 0.82 without standardization to 0.87 with standardization. SVM was slightly superior in its raw form but still improved after standardization, with the macro F1-score increasing from 0.85 to 0.86. These findings highlight the importance of data pre-processing and demonstrate how feature scaling can optimize machine learning models for heart disease prediction. This research contributes to the growing field of predictive healthcare, offering valuable insights for clinicians seeking reliable early detection tools for cardiovascular conditions.*

**Keywords**: *Cardiovascular Diseases, Feature Standardization, Heart Disease, Logistic Regression, Machine Learning Model, Support Vector Machine*

Received for review: 06-06-2025; Accepted: 14-07-2025; Published: 01-10-2025
DOI: 10.24191/mjoc.v10i1.6835

## 1. Introduction

Cardiovascular diseases (CVDs), including heart disease, are among the leading causes of morbidity and mortality worldwide, responsible for an estimated 17.9 million deaths annually (World Health Organization, 2021). This staggering statistic underscores the immense global health burden of CVDs, which not only contribute to significant mortality but also result in widespread disability and reduced quality of life. Heart disease, as a major form of CVD,

This is an open access article under the CC BY-SA license (https://creativecommons.org/licenses/by-sa/3.0/).

encompasses a variety of conditions such as coronary artery disease, heart failure, and arrhythmias. These diseases are often the result of complex interactions between genetic, environmental, and lifestyle factors, including hypertension, diabetes, and hyperlipidemia (Rowden, 2024). Early detection and personalized treatment strategies are critical for managing the progression of these conditions, as Balaraju et al. (2024) provide the opportunity to mitigate the risk of complications and improve long-term health outcomes. As such, the prediction of heart diseases is an important task in medicine (Balaraju et al., 2024).

The importance of early detection in heart disease cannot be overstated. Timely diagnosis enables healthcare providers to initiate interventions that can significantly reduce the risk of severe complications, such as heart attacks, strokes, and chronic heart failure. For example, patients identified at high risk can benefit from targeted treatments, lifestyle modifications, and even preventive surgeries that may help delay or prevent the onset of full-blown heart disease.

In contrast, delayed diagnosis often leads to advanced stages of the disease that require more intensive, expensive treatments and often result in poorer health outcomes. Early detection also allows for the possibility of reversing or managing the disease before it becomes life-threatening, improving the quality of life for individuals at risk (Grgić et al., 2021). Traditional diagnostic methods for heart disease, though essential, are often limited by subjectivity and the clinician's judgment (Muhammad et al., 2020). These methods typically rely on physical examinations, medical history, and diagnostic tests such as electrocardiograms and blood pressure measurements. However, these approaches may miss subtle patterns or early symptoms, especially among patients presenting with atypical manifestations (Rowden, 2024).

Additionally, the reliance on clinical experience introduces variability and potential for human error. To overcome these limitations, machine learning (ML) techniques have gained traction as a promising tool in the prediction and diagnosis of heart diseases (Suhaimi et al., 2024). Machine learning (ML) models, such as Logistic Regression (LR) and Support Vector Machines (SVM), are capable of processing complex, high-dimensional data and identifying patterns that may not be easily detectable (Mojahid et al., 2025). These models offer a more objective and scalable approach to heart disease prediction, improving diagnostic accuracy and enabling earlier intervention. However, the performance of these models is highly dependent on the quality of the data used for training, where one crucial step in enhancing model accuracy is feature scaling.

Feature scaling, including standardization, is an essential pre-processing step in machine learning that ensures all features contribute equally to the model's performance. In datasets where features have different magnitudes or units, unscaled data can lead to biased results, particularly in algorithms like SVM, which are sensitive to the scale of input features (Guido et. al, 2024). Standardization transforms features into a common scale with a mean of zero and a standard deviation of one, allowing the model to learn more effectively by preventing any feature from dominating the learning process (Bhandari, 2025). Without proper scaling, models may fail to capture the true relationships between the features and the target variable, resulting in suboptimal predictions.

This study aims to examine the impact of feature standardization on the performance of LR and SVM models for heart disease prediction. The research utilizes a dataset that includes clinical attributes such as age, cholesterol levels, and blood pressure, comparing the predictive accuracy of the models when trained on raw and standardized data. The findings are expected to provide valuable insights into the importance of feature scaling in improving model performance, contributing to the development of more reliable machine learning-based diagnostic tools for heart disease prediction.

## 2. Related Works

Machine Learning has emerged as a powerful tool for heart disease prediction, leveraging large datasets of clinical and demographic features to identify patterns and predict outcomes. Among

the many ML algorithms explored, Logistic Regression and Support Vector are two of the most commonly used methods for binary classification tasks, particularly in medical diagnostics (Balaraju et al., 2024).

LR has been widely utilized for predicting heart disease, primarily due to its simplicity and interpretability. It allows for the identification of relationships between clinical variables, such as age, cholesterol levels, and blood pressure, and the likelihood of heart disease. Studies, such as those by Zhang et al. (2021), have demonstrated the effectiveness of LR in heart disease prediction, finding that while it performs well in linearly separable cases, it may struggle when faced with more complex, non-linear relationships in the data. As heart disease involves complex interactions among multiple risk factors, models that capture these non-linear patterns may offer enhanced performance.

On the other hand, SVM is particularly effective in scenarios where the data is not linearly separable. SVM works by identifying an optimal hyperplane that maximizes the margin between different classes, particularly useful when the data exhibits complex, non-linear relationships. Studies such as those by Owusu et al. (2021) and Krishna et al. (2023) have shown that SVM consistently outperforms LR in heart disease prediction, delivering higher accuracy and sensitivity. SVM's ability to handle non-linearities through kernel functions, such as the Radial Basis Function (RBF) kernel, makes it a more flexible model for medical datasets that may contain intricate patterns and interactions between clinical features.

A critical aspect of improving machine learning models for heart disease prediction is the application of feature scaling, particularly standardization. Standardization is essential in ensuring that all input features contribute equally to the model's learning process. When features have different scales, models may be biased towards those with larger magnitudes, leading to inaccurate predictions (Leino et al., 2018). This issue is especially pronounced in SVM, which is sensitive to the scale of input features. Bhandari (2024) highlights that without standardization, models may fail to capture the true relationships between variables, leading to suboptimal performance. Previous research has demonstrated the significant benefits of feature standardization in improving the accuracy of SVM models. Ozsahin et. al (2022) found that when SVM models were trained on standardized data, their performance improved significantly, with higher F1-scores and more accurate predictions. Sarra et. al (2022) further support this by showing that standardized data helps in avoiding biases, ensuring that features with smaller magnitudes are not underrepresented in the model's decision-making process.

Given the strengths and limitations of LR and SVM, this study seeks to compare the two models by training them on both the raw and standardized datasets. By doing so, the study aims to evaluate how feature standardization impacts the predictive accuracy of these models and to provide insights into the best practices for preparing data in heart disease prediction. The findings of this research are expected to contribute to the ongoing efforts to enhance predictive models for heart disease, helping to ensure that machine learning tools are more reliable and accurate in medical diagnostics, ultimately leading to more timely and effective healthcare interventions.

## 3. Methodology

This study employs a structured approach to develop and evaluate the performance of two supervised classification machine-learning models. The following sections outline the study workflow, delineating the procedural steps undertaken in each section. Figure 1 illustrates the overall sequence of processes from the initiation to completion of the study. This systematic approach ensures that the models are robust, generalize well to new data, and provide reliable predictions for heart disease diagnosis. The methodology adopted in this study aims to not only compare the performance of different models but also to assess the impact of pre-processing steps, such as feature standardization on the overall effectiveness of the prediction process.

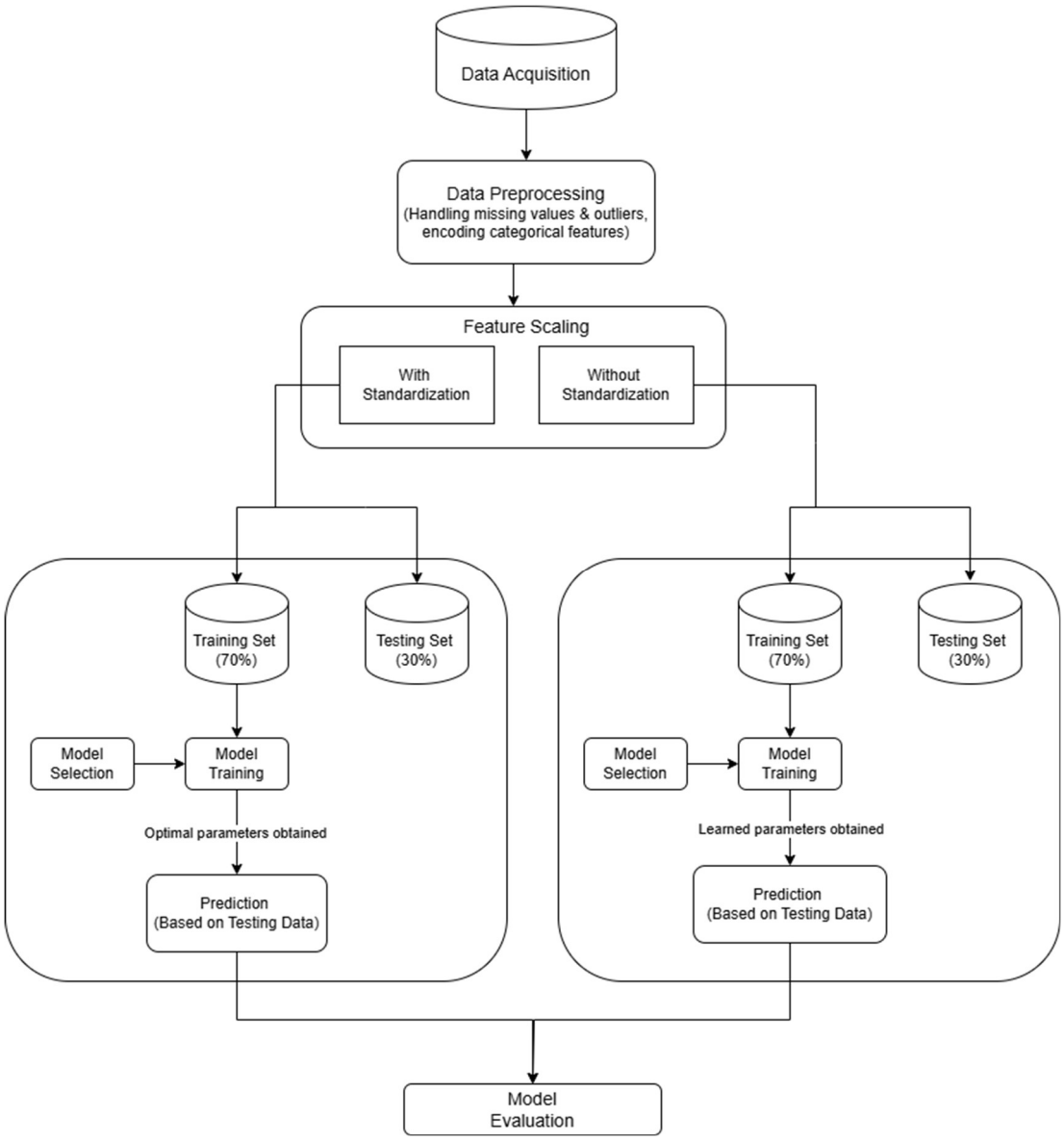


Figure 1. Workflow of the Study

### 3.1 Data Acquisition and Description

The dataset utilized by this research is obtained from Kaggle (Fedesoriano, 2021). It consists of patient entries merged from five independently available datasets, where eleven common features are included as clinical parameters. The total observations or entries present within the final dataset is 918. The dataset is slightly imbalanced, with 410 normal observations and 508 observations with heart disease. According to the source data card, the incorporated datasets that were utilized include the popular Cleveland and Statlog heart datasets, alongside the Hungarian, Switzerland, and Long Beach VA datasets. From these datasets, the observations were aggregated based on 11 identified common features, forming a more comprehensive and the largest heart disease dataset currently accessible for research endeavours. The variables used in this research comprise 11 predictors and a target variable, which the latter indicates the occurrence of heart disease using a binary value. The specific definition and relevance of each variable, along with the corresponding units of measurement, are presented in Table 1.

Table 1. Data Description

| Variable | Description (Units/Values) |
|---|---|
| Age | The age of the patient recorded. (Years) |
| Sex | The sex of the patient. (M: Male, F: Female) |
| ChestPainType | The type of chest pain experienced by the patient. (TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic) |
| RestingBP | The patient's resting blood pressure. (mm Hg) |
| Cholesterol | The patient's serum cholesterol level. (mg/dL) |
| FastingBS | The patient's fasting blood sugar. (1: FastingBS > 120 mg/dL, 0: Otherwise) |
| RestingECG | The resting electrocardiogram results recorded for the patient. (Normal: Normal, ST: ST-T wave abnormality, LVH: Left ventricular hypertrophy) |
| MaxHR | The maximum heart rate achieved by the patient. (Numeric value between 60 and 202, bpm) |
| ExerciseAngina | Exercise-induced angina experienced by the patient. (Y: Yes, N: No) |
| Oldpeak | The ST depression caused by activity in comparison to rest. (Numeric value measured in depression) |
| ST_Slope | The slope of the peak exercise ST segment. (Up: Upsloping, Flat: Flat, Down: Down sloping) |
| HeartDisease | The target variable; whether the patient has heart disease. (1: Heart disease, 0: Normal) |

In terms of variable types, there are five numerical variables e.g., Age, RestingBP, Cholesterol, MaxHR, and OldPeak. Additionally, there are seven categorical variables e.g., Sex, ChestPainType, FastingBS, RestingECG, ExerciseAngina, and ST_Slope. Notably, the data description in the data card on Kaggle contains a typographical error on the "Cholesterol" attribute. The unit of measurement is incorrectly labelled as "mm/dl" instead of the standard "mg/dL" (milligrams per deciliter). This mistake is rectified in this study for accurate analysis and interpretation.

### 3.2 Data Pre-processing

Before model training, the dataset undergoes a pre-processing phase to ensure its suitability for analysis. Data pre-processing is a crucial step in the data classification process because it directly affects the task success rate (Hon et al., 2023). This phase includes tasks such as handling missing values and outliers, encoding categorical variables, and applying feature scaling. The goal of data pre-processing is to improve the quality, consistency, and compatibility of the dataset, making it ready for accurate model training (Wanyonyi & Masinde, 2025).

#### 3.2.1 Handling Missing Values and Outliers

The dataset is free from null values and duplicate rows, simplifying the data cleansing process, as no additional steps are needed to address missing or duplicated data. However, the absence of null values does not necessarily imply the absence of outliers, which can distort model performance. As shown in Figure 2, a box plot is constructed for all numerical variables to visualize and analyze potential anomalies for detecting outliers. Outliers in medical datasets require careful consideration, as they could represent genuine but rare health conditions, rather than erroneous data. Removing or altering these outliers without due diligence could lead to the loss of valuable insights and diagnostic information. For example, cholesterol levels above 200 mg/dL are considered high, indicating an increased risk of heart disease (Cleaveland Clinic, 2024). Similarly, resting blood pressure readings above or below the normal range of 120/80 mmHg indicate potential cardiovascular issues (World Health Organization, 2021). Maximum

heart rate varies with age, typically calculated using a standard formula (Lach et al., 2021). Given the age range in the dataset (28 to 77 years), it would be inappropriate to treat the variation in max heart rates as outliers. Additionally, the "Oldpeak" variable refers to ST depression, where a baseline value of 0 indicates a healthy heart. Deviations from this baseline could suggest heart-related conditions, such as heart failure (Rowden, 2024).

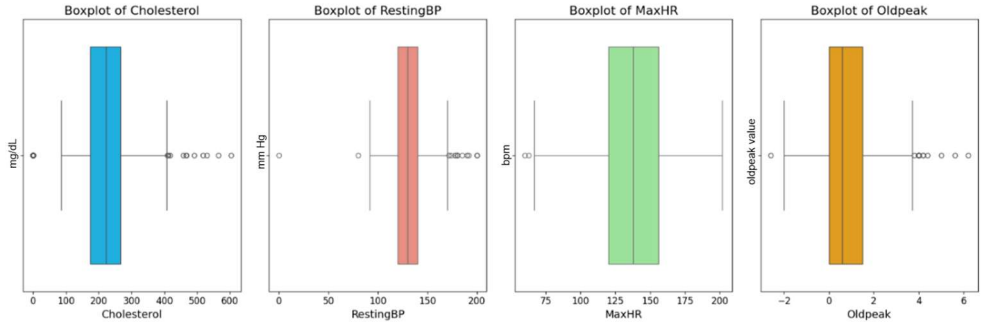


Figure 2. Boxplot for Numerical Variables

In this study, the presence of 0 mg/dL in the Cholesterol and 0 mm Hg in RestingBP columns is considered unusual and likely a result of data collection errors, where 0 values were mistakenly used to indicate missing data. To address this, mean imputation is applied, replacing the 0 values with the mean of the respective columns. After imputation, the dataset contains only valid values in these columns, ensuring that the data is accurate and ready for analysis.

### 3.2.2 Encoding Categorical Variables

The next step is encoding variables, crucial for preparing the dataset for analysis. Categorical variables, which represent qualitative attributes, cannot be directly utilized by most machine learning algorithms. Therefore, encoding is necessary to convert categorical data into a numerical format that can be processed by the selected classification models (Anitha, Savarimuthu & Bhanu, 2025). In this study, label encoding, also known as ordinal encoding, is applied to map categorical variables into numerical representations. This method assigns a unique integer to each category within a variable, thereby preserving the nominal quality of the categories. This step ensures that the categorical data is standardized and compatible with classification models, preparing the data for model training and the application of supervised classification techniques.

## 3.3 Feature Scaling

Feature scaling is a critical pre-processing step that involves transforming the range of numeric features to a standard scale. In this implementation, standardization is utilized via StandardScaler. This ensures that features with larger magnitudes do not disproportionately influence the model training process, and that all input features contribute equally to the model training process (Bhandari, 2024). Due to the difference in orders of magnitude present in the dataset, feature scaling becomes necessary. Hence, StandardScaler is applied to map the feature value of the dataset to the same range. The equation is presented in Equation (1).

$$X' = \frac{X - \mu}{\sigma} \tag{1}$$

Where $X$ is the data value, μ is the mean and σ is the standard deviation. The scaler is fitted on the training data before it is used to transform both the training and testing sets. This

approach is employed to prevent any leakage of data from the testing set to the training set, which could lead to inaccuracies in the estimation of performance on unseen data during the model evaluation process. By fitting the scaler on the full dataset before data splitting, instead, information about the testing set is passed downstream, where the distribution of the data may influence the way the models are parameterized during the training process. To ascertain the necessity of standardization for the classification task, the models were trained with and without feature scaling. They were fitted to both raw and standardized data, respectively, to compare their performance and determine the optimal approach for achieving the best results. In both scenarios, grid search was utilized for hyperparameter tuning.

Grid search is a technique that systematically searches through a specified hyperparameter space to determine the optimal combination of hyperparameters for a machine learning algorithm. It accomplishes this by exhaustively evaluating each combination of hyperparameters using cross-validation and selecting the combination that yields the highest performance metric. Due to the computational load of grid search, only two hyperparameters were selected for each of the two classification models. For LR, the key hyperparameters are the regularization parameter (alpha) and the penalty type (L1 or L2). The alpha parameter controls the strength of regularization, with higher values leading to stronger regularization. L1 regularization promotes sparsity by penalizing the absolute values of the coefficients, while L2 regularization penalizes the squared values of the coefficients. Together, these parameters help reduce model complexity and the risk of overfitting by treating high variance (Wu et. al, 2025).

For SVM, the important hyperparameters are the C regularization parameter and gamma ($\gamma$). The C parameter controls the trade-off between minimizing training error and the complexity of the model. A larger C value results in a smaller margin but higher accuracy on the training data, while a smaller C value allows a larger margin but may accept more classification errors, leading to better generalization. The gamma parameter controls the influence of a single training point, with smaller gamma values creating a smoother decision boundary, and larger gamma values resulting in more complex boundaries. Grid search is used to find the optimal combination of these hyperparameters to maximize model performance.

### 3.4 Data Splitting

In order to assess the generalization capabilities of the classification models, the dataset is partitioned into separate subsets for training and testing. The commonly used technique of random sampling is utilized to divide the dataset into two distinct sets e.g., the training set and the testing set, for model training and evaluation, respectively. In this study, the dataset was split using a 70:30 ratio, with 70% allocated for training the model and the remaining 30% reserved for testing its performance, following a common practice adopted in previous research (Ibrahim et al., 2024). Meanwhile, a same random state value is set for consistency in data splitting across multiple executions, guaranteeing identical results when re-running the code. This is helpful for reproducibility, as maintaining the same split facilitates documentation and comparison of results across iterations of the study.

### 3.5 Model Selection, Training and Testing

Two distinct supervised classification machine learning models are selected for this study. The choice of models is determined by their suitability for the task at hand, as well as their widespread usage and established performance in similar contexts. In this case, the task is predicting the presence of heart failure based on an input of predictors in the context of cardiovascular diseases. In this study, LR and SVM are chosen for model training and evaluation.

Following the model training phase, the models with learned parameters are evaluated using the dataset reserved for testing. This stage involves feeding unseen observations into the trained models and comparing their predictions against the actual target values in the testing

set. The study intends to ascertain the models' ability to generalise to new instances beyond the training set by assessing their performance on unseen data.

### 3.6 Model Assessment

The final stage of the study involves evaluation of the trained models' performance. Various metrics such as precision, recall, F1 score, and accuracy are assessed to gauge the effectiveness of the models in correctly classifying instances across different classes. In the context of the dataset used, there exist only two classes for the output variable: Normal or heart disease. This is known as binary classification, where the model predicts between the two classes given an input of predictors. Various performance metrics, including accuracy, precision, recall, and F1-score, are taken into consideration to assess the models' effectiveness in correctly classifying instances of observations.

The four-performance metrics i.e., accuracy, precision, recall, and F1-score are essential for evaluating classification models, as they are directly derived from the confusion matrix. The confusion matrix provides a tabular representation of the model's predictions against the actual class labels. It groups instances into four categories: true positives, true negatives, false positives (type 1 error), and false negatives (type two error) (Rainio, Teuho, & Klén, 2024). The confusion matrix for binary classification is illustrated in Figure 3.

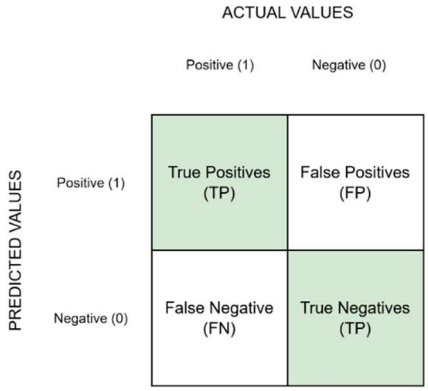


Figure 3. Confusion Matrix for Binary Classification

#### 3.6.1 Precision and Recall

Precision and recall are key evaluation metrics that together form the foundation for calculating the F1-Score (Bohani et al, 2024). Precision quantifies the proportion of correctly predicted positive instances among all instances predicted as positive by the model. In other words, it is the degree to which the model correctly classifies an instance, or how many of the predicted positive instances are positive. The formula for the precision score is presented in Equation (2).

$$\text{Precision} = \frac{tp}{tp + fp} \tag{2}$$

Recall refers to the proportion of correctly predicted positive instances among all positive instances. Put simply, it measures how often the model is correct in its prediction, given that the actual value is positive. It can also be thought of as the measurement of how many of the actual positive instances are correctly classified by the trained model. The formula for the recall score is presented in Equation (3).

$$\text{Recall} = \frac{tp}{tp + fn} \tag{3}$$

### 3.6.2 F1-Score

F1-score is defined as the harmonic mean between precision and recall, which combines the two metrics into a single score. This metric is helpful when a balance between high precision and high recall is required, since it penalizes extreme negative values of either component. The formula for the F1-Score is presented in Equation (4).

$$F_1 = \left( \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} \right) = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

The F1 score balances both precision and recall by treating false positives and false negatives equally. In some tasks, however, the importance of precision and recall may differ, depending on the consequences of each type of error. In medical domains like heart disease diagnosis, false negatives failing to detect the condition are far more harmful than false positives (Maxim, Niebo & Utell, 2014). Thus, recall may be prioritized, as having some false positives is more tolerable than missing a positive case. However, false positives can lead to unnecessary anxiety and invasive procedures (Saito & Rehmsmeier, 2015). While high recall indicates the model's ability to correctly identify positive cases, the F1 score offers a better balance by combining precision and recall, ensuring the model detects positives without excessive false alarms. This study will use the macro-average F1 score as the primary performance metric, especially given the minor class imbalance observed, where heart disease cases outnumber non-heart disease cases.

## 3.7 Simulation Procedure for Model Performance Evaluation

To strengthen the validity of the results, a simulation-based performance evaluation was conducted using Stratified Shuffle Split cross-validation. This method maintains the original class distribution across splits and allows for assessing model robustness under different train-test configurations.

A total of 50 independent simulation runs were carried out. In each run, the data was split into 70% training and 30% testing subsets using stratified sampling. Four models were evaluated: (1) Logistic Regression (unscaled) using SGDClassifier with log_loss, alpha=0.01, penalty='l1'; (2) Logistic Regression (scaled) using a pipeline with StandardScaler, alpha=0.001, penalty='l1'; (3) Support Vector Machine (unscaled) with RBF kernel, C=1000, gamma=0.001; and (4) Support Vector Machine (scaled) using a pipeline with StandardScaler, C=100000, gamma=0.000001.

Each model was trained on the training subset and evaluated on the testing subset. The F1 macro score was used as the evaluation metric to capture performance across all classes, especially in the presence of class imbalance. The simulation scores were collected and analyzed across all 50 runs to measure performance consistency, with mean and standard deviation computed for each model. The results were also visualized to compare the stability and generalization of each model under repeated trials.

## 4. Results

This section presents the results of the model evaluations, beginning with an analysis of the output class distribution to understand how common heart disease is in the dataset. Then, the performance of the two classification models is assessed using selected metrics. Graphs and

tables are used to clearly show how the models performed. Based on these results, conclusions are drawn to identify which model performs better in predicting heart disease.

### 4.1 Output Class Distribution

Figure 4 depicts the patient count for each output class within the dataset. Upon closer inspection of the distribution, it becomes evident that there is a slight imbalance between the two classes. Specifically, there are 410 instances classified as "Normal" and 508 instances classified as "Heart Disease". This observation indicates a minor class imbalance existing within the dataset, warranting the need for certain measures to be taken when interpreting the performance metrics of the trained models.

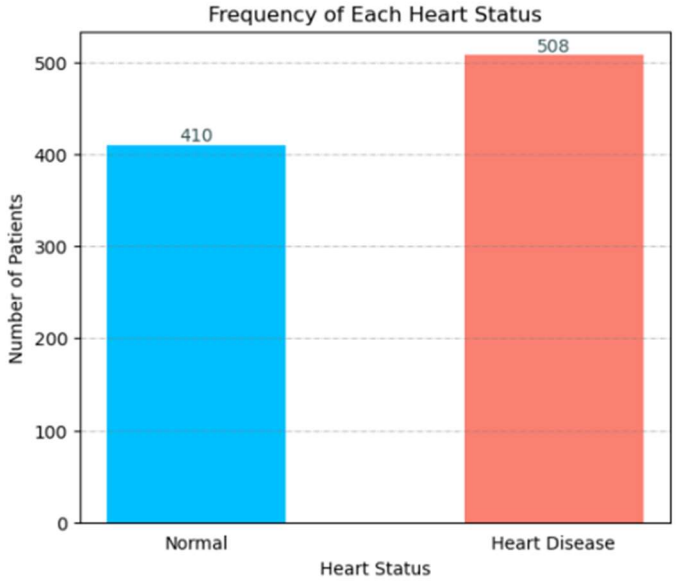


Figure 4. Heart Status Distribution

### 4.2 Classification Models Performances

Table 2 presents the macro F1 scores obtained from the evaluation of LR and SVM models on both the training and testing sets. The scores are provided for models trained with and without standardization, allowing for a comparative analysis of their performance across the standardized dataset and the raw dataset. The macro F1 scores for both LR and SVM models, with and without standardization, provide valuable insights into their performance on both the training and testing datasets. Without standardization, LR achieved a macro F1 score of 0.78 on the training set and 0.82 on the testing set.

Table 2. Macro F1-Scores for Logistic Regression (LR) and Support Vector Machine (SVM)

| | **LR Macro F1 Score** | | **SVM Macro F1 Score** | |
|---|---|---|---|---|
| | Train | Test | Train | Test |
| **Without Standardization** | 0.78 | 0.82 | 0.85 | 0.85 |
| **With Standardization** | 0.78 | 0.87 | 0.87 | 0.86 |

Conversely, SVM demonstrated higher performance, with a macro F1 score of 0.85 on both the training and testing sets. With standardization, LR's performance notably improved, achieving a macro F1 score of 0.86 on the training set and 0.87 on the testing set. Similarly, the

SVM's performance saw a slight improvement, with macro F1 scores of 0.87 on the training set and 0.86 on the testing set. Comparing the LR and SVM models, it is evident that SVM consistently outperformed LR in terms of macro F1 scores, irrespective of standardization. SVM exhibited higher accuracy rates on both training and testing sets, showcasing its superior predictive capabilities compared to LR. These results underscore the effectiveness of SVM in capturing complex patterns within the data and highlight its potential as a robust classification model for the given task.

The optimal hyperparameter values for both LR and SVM models, considering both scenarios with and without standardization, are presented in Table 3. Without standardization, LR employed an alpha of 0.001 and penalty type L1, whereas with standardization, LR's alpha increased to 0.01 while maintaining the L1 penalty type. Similarly, SVM utilized a C of 100,000 and gamma of 0.000001 without standardization, whereas with standardization, SVM's C decreased to 1,000 and gamma increased to 0.001. These variations highlight the impact of standardization in feature scaling on the selection of hyperparameter values for both LR and SVM models, with SVM showing a more substantial improvement in hyperparameter stability and thus benefiting more from standardization.

Table 3. Optimal Hyperparameters

| | **LR** | | **SVM** | |
|---|---|---|---|---|
| | Alpha | Penalty Type | C | Gamma |
| **Without Standardization** | 0.001 | L1 | 100 000 | 0.000001 |
| **With Standardization** | 0.01 | L1 | 1 000 | 0.001 |

Additionally, figure 5 provides a visual representation of the precision and recall scores. Results revealed that standardization substantially improved both precision and recall, particularly for LR, which showed low precision (0.73) but very high recall (0.98) without scaling, suggesting a tendency to over-predict positive cases. Upon applying standardization, the precision and recall for LR became more balanced, improving to 0.92 and 0.85 respectively on the test set. Similarly, for the Support Vector Machine (SVM) model, standardization also yielded improvements. Without scaling, SVM achieved a training precision of 0.85 and a testing precision of 0.88. After standardization, SVM's training precision saw a slight increase to 0.87, and its testing precision remained strong at 0.90.

JKK UHEK

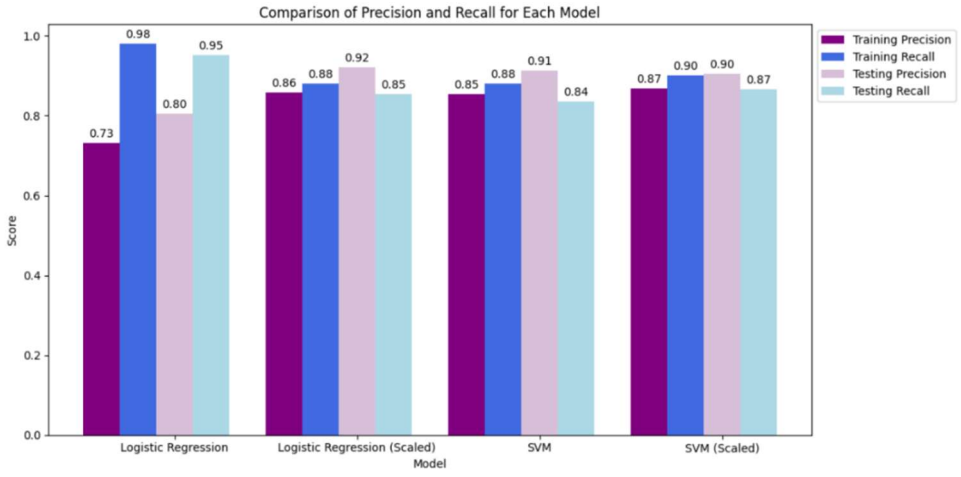


Figure 5. Precision and Recall for Each Classification Model

Based on these performance metrics, it is evident that standardization is highly impactful for both Logistic Regression and Support Vector Machine models. It consistently led to more balanced and robust results across precision and recall, demonstrating its crucial role in optimizing model performance in this context.

Figure 6 illustrates the F1 macro scores across 50 simulation runs for both scaled and unscaled versions of LR and SVM models. It is evident that models trained on unscaled data (blue and green lines) exhibit high variability and inconsistent performance across runs. In contrast, models using standardized data (orange and red lines) show much greater consistency, with lower standard deviation in scores. Among all models, the scaled SVM consistently achieves the highest F1 macro scores, indicating superior and stable performance when feature standardization is applied.

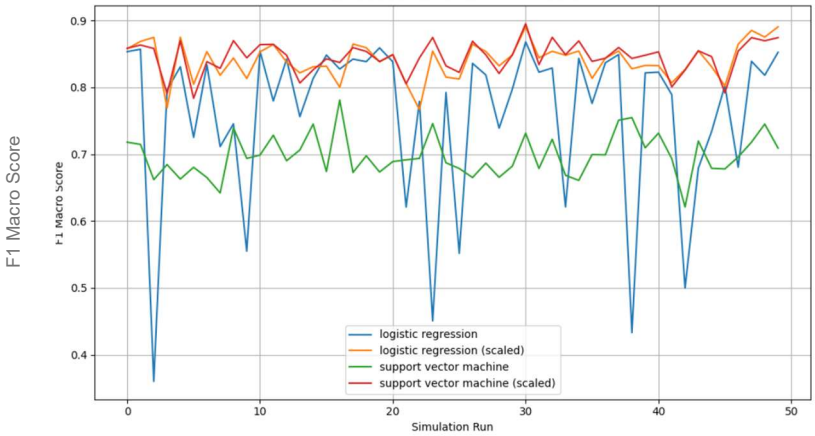


Figure 6. F1 Macro Scores Across 50 Simulations

## 5. Discussion

In the previous section, it was observed that Support Vector Machine performed slightly better compared to Logistic Regression in predicting heart disease, both in cases with raw data and standardized data. Moreover, the application of feature scaling had also improved the macro F1 score for both models, with LR receiving a greater positive impact from standardization.

### 5.1 The Impact of Feature Scaling

Without feature scaling, the original feature space consisted of variables with significantly different scales or ranges. When features are on different scales, the optimization algorithm can be dominated by the feature with a larger order of magnitude. This uneven influence on the optimization process may lead to the algorithm prioritizing the minimization of the loss function concerning features that have larger magnitudes, potentially overshadowing the importance of features with smaller magnitudes (Bhandari, 2024).

In contrast, when feature scaling is applied, the input features are transformed to a common scale and range. This standardization process ensures that each feature contributes proportionally to the optimization process, preventing any undue influence based on feature magnitude (Islam, 2024). As a result, the final model fit would not depend on the scale on which the predictors are measured (James et al., 2014), and can better capture the relationships between the features and the target variable, leading to a more balanced performance on precision and recall across both output classes, thereby improving the macro-average of the F1 score. From the increase in macro F1 scores in both LR and SVM, it can be seen that standardizing the input features enhanced the classifier's performance.

### 5.2 Model Hyperparameter Selection Behaviour

The choice of the hyperparameters was determined through grid search, where different values for said hyperparameters are evaluated and the combination that provides the best performance on a validation set is selected. The difference in the regularization parameter alpha between the LR models trained with and without feature scaling can be attributed to the scaling effect on the input features. The regularization term penalizes large coefficients in the model, enforcing simpler models and reducing the impact of individual features (James et al., 2014). Without the help of feature scaling, the contributions to the cost function and the regularization term can occur among features of different ranges and units. The value of 0.001 may have been sufficient to prevent overfitting, as the optimization algorithm struggles to balance the impact of features with varying magnitudes without underfitting. In this case, the regularization parameter needs to be smaller to compensate for the potentially large coefficients that may arise due to the significant differences in scale of the features. This was likely to prevent overly aggressive regularization. Otherwise, large coefficients would be heavily penalized by a larger term, diminishing the contribution of larger-scale features.

On the other hand, when feature scaling is applied, all features are scaled to have a similar range (with a standard deviation of 1 through standardization). This ensures that no features dominate the cost function or the regularization term due to their scale. As a result, a larger regularization parameter can be used to achieve the desired level of regularization, as the coefficients can be penalized evenly without over-penalizing any particular feature due to their similar scale. The alpha value of 0.01 may have been necessary to achieve comparable regularization effects, as the scaled features allow for more effective optimization and model generalization.

This reasoning applies to L1 regularization (lasso regression) as well, where the regularization term adds the sum of absolute values of coefficients multiplied by the alpha value (typically depicted as $\sigma$ instead). In the case of L1 regularization without feature scaling, features with larger scales will tend to have larger coefficient magnitudes. Therefore, if a large alpha value is used, the regularization terms, or the sum of absolute coefficients multiplied by alpha, will heavily penalize these larger coefficients, effectively shrinking or eliminating the contribution of larger-scale features. Due to their smaller scales, smaller features with smaller coefficients will not be penalized as much as the regularization term. Thus, a smaller alpha is used to allow the larger-scale features to have a more substantial contribution to the model. With feature scaling, a large alpha can be used in L1 regularization without penalizing any particular feature too harshly due to their similar scale and subsequently, coefficient ranges.

For SVM, grid search found that higher C and lower gamma values were optimal when features were not standardized. A higher C value reduces misclassifications by forcing the model to focus more on accurately classifying data points, resulting in a stricter, hard-margin classifier. This encourages a more complex decision boundary to improve training accuracy, especially when features have different scales (Kumar, 2020). A lower gamma value, on the other hand, widens the region of influence for each support vector, making points further away more similar (Al-Mejibli, Alwan & Abd, 2020). This helps prevent overfitting by making the model more general, especially with the larger C value. After standardization, the C value decreased because the features were on the same scale, so the model needed a lower C to manage misclassifications effectively. The reduced scale of data points likely helped lower misclassification rates. Additionally, a higher gamma value was more suitable post-standardization to shrink the similarity region and avoid overfitting (Kumar, 2020).

### 5.3 Better Overall Performance by SVM

In comparing the performance metrics of LR and SVM, it was found that SVM generally achieved slightly better F1-scores than LR. This result aligns with the inherent characteristics of both models. LR is a statistical model used to predict the probability of an instance belonging to a particular class based on a linear combination of features (Zhang et. al, 2021). In contrast,

SVM is a discriminative model that seeks the optimal hyperplane to separate two classes by maximizing the margin between the support vectors. SVMs excel at handling non-linear decision boundaries by utilizing kernel functions that map the data into higher-dimensional spaces, providing them with the flexibility to outperform LR when data is not linearly separable (Guido et. al, 2024). This suggests that the dataset used in this study may contain non-linear relationships that SVM can better capture, as LR struggles with non-linearly separable data.

If the results obtained from standardized data are analyzed in isolation, LR and SVM were found to be comparable in performance, where the difference in F1-score is marginal. The choice between them may depend on additional considerations, such as interpretability, computational efficiency, or the ability to handle non-linear decision boundaries. In addition, the resilience against non-standardized data may also become a factor, where SVM demonstrated a greater resistance against unscaled input features, provided that the hyperparameters are properly tuned. In hindsight, LR may have the potential to perform better on unscaled data if a more comprehensive hyperparameter tuning process were employed.

## 6. Conclusion and Recommendation

This study demonstrated the potential of machine learning models, specifically LR and SVM, in predicting the likelihood of heart disease using demographic and clinical data. The findings highlight that both models can provide valuable insights for early detection and personalized healthcare, with SVM showing a slightly better performance, especially when feature standardization was applied. The use of precision, recall, and F1 score allowed for a more balanced evaluation, revealing that the models were effective in identifying true cases of heart disease while minimizing false positives and negatives. For future improvement, applying feature selection methods such as SelectKBest can help identify the most important predictors, enhance model accuracy, and improve interpretability, particularly in clinical applications.

## Acknowledgement

The authors would like to express their sincere gratitude to the School of Computing and Creative Media at the University of Wollongong Malaysia for their unwavering support and resources throughout this research.

## Funding

The author(s) received no specific funding for this work.

## Author Contribution

Author 1 designed the experimental framework, oversaw the research process, rearranged the content for the manuscript, and supervised the study. Author 2 prepared the literature review, developed the research methodology, wrote the first draft of the manuscript, and performed the statistical analysis. Author 3 handled the data pre-processing, prepared the literature review, and ensured compliance with journal guidelines. Author 4 provided domain expertise, reviewed the manuscript for intellectual content, and finalized the manuscript for submission.

## Conflict of Interest

The authors have no conflicts of interest to declare.

## References

Al-Mejibli, I. S., Alwan, J. K., & Abd, D. H. (2020). The effect of gamma value on support vector machine performance with different kernels. *International Journal of Electrical and Computer Engineering, 10*(5), 5497–5506. https://doi.org/10.11591/ijece.v10i5.pp5 497-5506

Anitha, M., Savarimuthu, N., & Bhanu, S. M. S. (2025). Chi-Square Target Encoding for Categorical Data Representation: A Real-World Sensor Data Case Study. *SN Computer Science*, *6*(3). https://doi.org/10.1007/s42979-025-03766-z

Balaraju, G., Reddy, M. D. S., Manjunath, S. R., Hemalatha, M., & Veena, N. (2024). Heart Disease Prediction Using Classification Techniques of Supervised Machine Learning. *2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON)*, 1–5. https://doi.org/10.1109/nmitcon62075.2024.10699057

Bhandari, A. (2025, April 23). *What is Feature Scaling and Why is it Important?* Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning -normalization-standardization/

Bohani, F. A., Rashid, F. S. M., Mahmud, Y., & Yahya, S. R. (2024). Analyzing The Impact of Feature Selection Using Information Gain for Airlines' Customer Satisfaction. *Malaysian Journal of Computing (MJOC)*, *9*(1), 1673–1689. https://doi.org/10.24191 /mjoc.v9i1.24163

Cleveland Clinic. (2022, August 4). *Hyperlipidemia*. Cleveland Clinic. https://my.cleveland clinic.org/health/diseases/21656-hyperlipidemia

Fedesoriano. (2021, September 10). *Heart Failure Prediction Dataset*. Kaggle. https://www. kaggle.com /datasets/fedesoriano/heart-failure-prediction

Grgić, V., Mušić, D., & Babović, E. (2021). Model for predicting heart failure using Random Forest and Logistic Regression algorithms. *IOP Conference Series Materials Science and Engineering*, *1208*(1), 012039. https://doi.org/10.1088/1757-899x/1208/1/012039

Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information*, *15*(4), 235. https://doi.org/10.3390/info15040235

Hon, H., Wah Khaw, K., Chew, X., & Wong, W. (2023). Prediction of Customer Churn for ABC Multistate Bank Using Machine Learning Algorithms. *Malaysian Journal of Computing*, *8*(2), 1602–1619. https://doi.org/10.24191/mjoc.v8i2.21393

Ibrahim, N., Ishak, U. M., Ali, N. N. A., & Shaadan, N. (2024). Machine Learning-Based Approaches for Credit Card Debt Prediction. *Malaysian Journal of Computing (MJOC)*, *9*(1), 1722–1733. https://doi.org/10.24191/mjoc .v9i1.25656

Islam, N. (2024). *DTization: A New Method for Supervised Feature Scaling*. ArXiv.org. https://arxiv.org/abs/2404.17937

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). *An introduction to statistical learning: with Applications in R.* Springer.

Krishna, T. B., Vimala, N., Vinay, P., Siddhardha, N., & Manohar, P. M. (2023). Early heart disease prediction using support Vector machine. In *Lecture notes in networks and systems* (pp. 471–479). https://doi.org/10.1007/978-981-99-3758-5_43

Kumar, A. (2023, April 15). *SVM RBF Kernel Parameters: Python Examples - Analytics Yogi*. Analytics Yogi. https://vitalflux.com/svm-rbf-kernel-parameters-code-sample/

Lach, J., Wiecha, S., Śliż, D., Price, S., Zaborski, M., Cieśliński, I., Postuła, M., Knechtle, B., & Mamcarz, A. (2021). HR Max Prediction Based on Age, Body Composition, Fitness Level, Testing Modality and Sex in Physically Active Population. *Frontiers in Physiology*, *12*. https://doi.org/10.3389/fphys.2021. 695950

Leino, K., Black, E., Fredrikson, M., Sen, S., & Datta, A. (2018). Feature-Wise Bias Amplification. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1812.08999

Maxim, L. D., Niebo, R., & Utell, M. J. (2014). Screening tests: a review with examples. Inhalation Toxicology, 26(13), 811–828. https://doi.org/10.3109/08958378.2014.955932

Mojahid, H. Z., Zain, J. M., Yusoff, M., Basit, A., Jumaat, A. K., & Ali, M. (2025). Examining The Impact of Feature Selection Techniques on Machine and Deep Learning Models for The Prediction of Covid-19. *Malaysian Journal of Computing*, *10*(1), 2135–2158. https://doi.org/10.24191/mjoc.v8i2.21393

Muhammad, Y., Tahir, M., Hayat, M., & Chong, K. T. (2020). Early and accurate detection and diagnosis of heart disease using intelligent computational model. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-76635-9

Ozsahin, D. U., Taiwo Mustapha, M., Mubarak, A. S., Said Ameen, Z., & Uzun, B. (2022, August). Impact of feature scaling on machine learning models for the diagnosis of diabetes. *2022 International Conference on Artificial Intelligence in Everything (AIE)*. https://doi.org/10.1109/aie57029.2022.00024

Owusu, E., Boakye-Sekyerehene, P., Appati, J. K., & Ludu, J. Y. (2021). Computer-Aided diagnostics of heart disease risk prediction using boosting Support Vector machine. *Computational Intelligence and Neuroscience*, *2021*(1). https://doi.org/10.1155/2021 /3152618

Rainio, O., Teuho, J. and Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, [online] 14(1), pp.1–14. doi:https://doi.org/10.1038/s41598-024-56706-x.

Rowden, A. (2024, April 25). *What does ST depression on an ECG result mean?* https://www.medicalnewstoday.com/articles/st-depression-on-ecg

Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Sarra, R. R., Dinar, A. M., Mohammed, M. A., & Abdulkareem, K. H. (2022). Enhanced Heart Disease Prediction Based on Machine Learning and $\chi 2$ Statistical Optimal Feature Selection Model. *Designs*, *6*(5), 87. https://doi.org/10.3390/designs6050087

Suhaimi, M. S. A., Ramli, N. A., & Muhammad, N. (2024). Heart disease prediction using ensemble of k-nearest neighbour, random forest and logistic regression method. *AIP Conference Proceedings*, *3080*, 040009. https://doi.org/10.1063/5.0192203

Wanyonyi, E. N., & Masinde, N. W. (2025). The Impact of Data Preprocessing on Machine Learning Model Performance : A Comprehensive Examination. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *11*(2), 3814–3827.

World Health Organization: WHO. (2021, June 11). *Cardiovascular diseases (CVDs)*. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

Wu, W., Wang, J., Lin, J. and Liu, X. (2025). Comparative Study of Adaptive l1-Regularization for the Application of Structural Damage Diagnosis Under Seismic Excitation. *Buildings*, [online] 15(10), pp.1628–1628. doi:https://doi.org/10.3390 /buildings15101628.

Zhang, Y., Diao, L., & Ma, L. (2021). Logistic regression models in predicting heart disease. *Journal of Physics Conference Series*, *1769*(1), 012024. https://doi.org/10.1088/1742-6596/1769/1/012024