Heart Disease Diagnosis Using Decision Trees with Feature Selection Method

Alaa Sheta
Department of Computer Science
Southern Connecticut State University, USA
shetaa1@southernct.edu

Walaa El-Ashmawi
Department of Computer Science
Suez Canal University, Egypt
w.hashmawi@ci.suez.edu.eg

Abdelkarim Baareh
Department of Computer Science
Isra University, Jordan
baareh@bau.edu.jo

Abstract: *The advancement in treating medical data grows significantly daily. An accurate data classification model can help determine patient disease and diagnose disease severity in the medical domain, thus easing doctors' treatment burdens. Nonetheless, medical data analysis presents challenges due to uncertainty, the correlations between various measurements, and the high dimensionality of the data. These challenges burden statistical classification models. Machine Learning (ML) and data mining approaches have proven effective in recent years in gaining a deeper understanding of the importance of these aspects. This research adopts a well-known supervised learning classification model named a Decision Tree (DT). DT is a typical tree structure consisting of a central node, connected branches, and internal and terminal nodes. In each node, we have a decision to be made, such as in a rule-based system. This type of model helps researchers and physicians better diagnose a disease. To reduce the complexity of the proposed DT, we explored using the Feature Selection (FS) method to design a simpler diagnosis model with fewer factors. This concept will help reduce the data collection stage. A comparative analysis has been conducted between the developed DT and other various ML models, such as Logistic Regression (LR), Support Vector Machine (SVM), and Gaussian Naive Bayes (GNB), to demonstrate the effectiveness of the developed model. The results of the DT model establish a notable accuracy of 93.78% and an ROC value of 0.94, which beats other compared algorithms. The developed DT model provided promising results and can help diagnose heart disease.*

Keywords: *Machine learning, classification, decision tree, heart disease, diagnosis.*

Received December 24, 2023; accepted April 9, 2024
https://doi.org/10.34028/iajit/21/3/7

1. Introduction

One of the prominent contributors to morbidity and mortality on a worldwide basis is heart disease [25]. Heart disease involves various illnesses, such as coronary artery disease, stroke, and heart failure. Cardiovascular disease is a common term for identifying heart and other related heart and blood vessel disorders like peripheral arterial disease [25, 27, 28]. The European Public Health Alliance claimed in 2010 that circulatory diseases, such as heart attacks and strokes, account for 41% of all deaths. According to the Economic and Social Commission for Asia and the Pacific, diseases such as heart disease, cancer, diabetes, and lung diseases are the primary causes of death in one-fifth of Asian nations. Likewise, it was stated by the Australian Bureau of Statistics that heart diseases account for around 33.7% of the deaths in the country. In the USA, approximately 650,000 deaths occur annually due to heart diseases. Predicting the risk of heart disease is crucial for early diagnosis and timely intervention to reduce the associated health risks. Heart disease has become the number one killer for over a century, according to the 2024 Heart Disease and Stroke Statistics: A Report of U.S. and global data from the American Heart Association. The report also stated that the number of people dying from a heart attack in the United States each year has dropped from 1 in 2 in the 1950s to grow recently to 1 in 8.5.

The World Health Organization (WHO) forecasts indicate that Non-Communicable Diseases (NCDs) are responsible for approximately 70% of the world's annual deaths, equivalent to roughly 40 million individuals, and this proportion is anticipated to rise by an additional 10% by the year 2030 [27]. Cardiovascular disease is relatively expensive to diagnose and treat, making it often unaffordable for the whole community. Data mining methods facilitate early-stage risk assessment for heart disease, potentially reducing diagnosis and treatment costs [43].

Medical data mining for the diagnosis and treatment of heart disease is an emerging field that has sparked the interest of many researchers as Cios [12], Prather *et al*. [32], Purushottam *et al*. [34], and Yang *et al*. [50]. Extracting valuable insights from medical data poses a significant challenge for professionals. Employing machine learning techniques to condense prior studies on heart disease prediction, exploring a fusion of these methods to unveil the most appropriate and efficient approach was presented in [16]. Machine Learning (ML) methods are modern and promising technologies that employ advanced statistical techniques to uncover relationships and analyze information from large data sets. ML algorithms have arisen as robust methods,

mainly in the healthcare sector, playing an essential role in predicting and assessing the risks associated with heart diseases [22, 37]. Figure 1 shows some statistics about the causes of death in the USA in 2016 [3]. Coronary heart disease represents 43% of the cause of death [28].

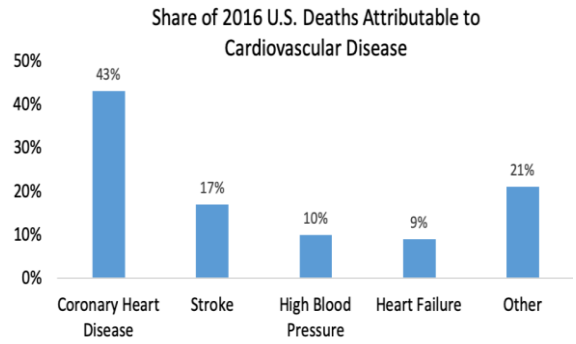

Figure 1. Source: american heart association: statistics on coronary heart disease in the USA [3].

ML is utilized to develop predictive models to diagnose heart disease by examining various patient-related features, such as medical history, lifestyle, and clinical measurements [5]. ML models are efficient for physicians in multiple domains, allowing them to make well-versed decisions and recommend appropriate treatments [27, 32]. Different models, such as Decision Tree classifier (DT), Random Forest (RF), XGBoost (XGB), Naive Bayes (NB) [23], and Multi-Layer Perceptron (MLP), were used to create a diagnosis model for predicting heart disease. ML is treasured for its transparency, providing insight into decision-making. However, care must be taken to prevent overfitting, and techniques like pruning can be applied to optimize the tree's structure.

1.1. Goals and Outlines

The research aims to improve heart disease detection accuracy by implementing ML methodologies. We propose an approach that leverages DTs for classification alongside a Feature Selection (FS) method. This work contributes by developing and evaluating a framework that uses DTs to identify the most relevant factors from patient data for accurate heart disease prediction. Consequently, this research aims to build a decision-tree model for detecting and diagnosing heart disease using a Public Health dataset obtained from [17]. The proposed approach involves some phases, including data scaling and FS. Furthermore, the developed Decision Tree (DT) model can be improved by restricting its depth for better performance. While various ML models exist for classification tasks, DTs are selected for their interpretability, capacity to handle categorical and numerical data, and efficiency in training. Additionally, DT is appropriate for real-world implementations in healthcare environments as it provides transparency, requires minimal computational power, and can handle categorical and numerical data. The decision rules extracted in the study offer an efficient model for clinical applications, excluding the need for further physician diagnosis. The proposed DT model can likewise serve healthcare professionals and patients, specifically with cost and time constraints in diagnosing heart disease.

This paper is organized as follows: An overview of the various studies conducted and exploring various ML models for heart disease prediction is presented in ssection 2. The core notion of designing and training a DT, accompanied by illustrative examples, is presented in section 3. The learning process for constructing a DT, employing two frequently used methods Gini and Entropy is also given in section 4. Section 5 offers an insight into the proposed classification method. The analysis of the results of the compared algorithms is discussed in section 6. Finally, section 7 encapsulates the essential findings and potential avenues for future research, providing a concluding perspective on the paper.

2. Related Work

Heart disease diagnosis has been a prominent area of research in ML, with various techniques demonstrating promising results. This section explores relevant studies that utilize DTs and FS methods for heart disease prediction. Several studies have successfully employed DTs for heart disease classification. In addition, Sharma and Kumar [41] examined different algorithms used in DT analysis in various domains. DTs have proven successful in diagnosing medical data. Bond and Sheta [7], utilized several datasets, including Pima Indians Diabetes, Heart Failure Clinical Records, and the Breast Cancer Coimbra datasets for the diagnosis using DT, Support Vector Machines (SVMs), and Artificial Neural Networks (ANN). Shouman *et al*. [44] investigate different DT algorithms, highlighting the potential of DTs for interpretable and accurate diagnosis. Krishnan and Geetha [21] utilized DT and NB algorithms to achieve better accuracy in a heart disease diagnosis. At the same time, Nichenametla *et al*. [29] have proven that DT has achieved promising results when compared with NB.

The heart disease dataset published at the UCI Machine Learning Repository was used to build a DT classifier [48]. In [1] a DT classifier with 5-fold cross-validation was explored. The findings implied that the DT classifier accomplished an accuracy rate of 81% in correctly identifying patients with heart disease and 82% in correctly identifying those without heart disease. These accuracy rates outperformed other ML algorithms utilized in prior studies. Tu *et al*. [47] used the bagging algorithm provided by the Weka software and explored its performance with the J4.8 DT for diagnosing heart disease. The developed results showed that the bagging algorithm accomplished an accuracy rate of 81.41%,

beating the DT, which achieved an accuracy rate of 78.91%.

Other classifications, such as SVMs, offer an alternative approach for heart disease diagnosis; however, Vijaya Saraswathi *et al*. [49] demonstrating their competitive results and proving the DT has achieved high accuracy. The paper [20] assesses many classifiers by employing data-mining techniques from Orange and Weka to make predictions about heart disease. The dataset consists of 297 records and 13 features. A hybridization technique has been proposed in Maji and Arora [24] to improve heart disease prediction performance. This technique combines DT and artificial neural network classifiers using Weka.

However, FS plays a crucial role in optimizing performance. Research by Dissanayake and Johar [13] highlights the benefits of selecting the most relevant features for DT models, leading to improved prediction accuracy. Techniques like information gain can be employed to identify these key features. Spencer *et al*. [45] empirically evaluate the effectiveness of various ML models by employing several feature-selection approaches to identify significant features that significantly affect heart disease prediction.

Table 1 presents a concise overview of prior research in this domain, including several studies utilizing various ML techniques for predicting heart disease. Most previous research has successfully predicted the occurrence of heart disease using multiple ML techniques.

Table 1. Summary of prior research in heart disease prediction.

Ref.	Year	Objectives	Techniques
[6]	2023	The study aims to develop a ML model for early heart disease prediction using various FS techniques.	Six various ML techniques, including LR, SVM, K-nearest neighbor, and DT, along with three FS techniques.
[42]	2023	This study utilizes various ML to predict heart disease by considering different features.	The three ML techniques utilized were DT, K-nearest neighbor, and NB.
[15]	2023	This study presents an adaptive version of the Hoeffding tree for early heart disease prediction.	Adaptive Hoeffding Tree (AHT) algorithm is utilized compared to standard ML techniques.
[8]	2022	This study introduces several machine learning methods for predicting heart diseases, utilizing patient data on essential health variables.	Four classifiers are utilized: MLP, SVM, RF, and NB.
[4]	2022	The study aims to identify risk factors for heart disease.	Several ML methods were used, including DT, SVM, LR, and NB.
[38]	2021	The study provides a hybrid system that assists doctors in the early stage of heart disease detection.	Various ML techniques, such as SVM, NB, and LR
[19]	2021	The goal is to develop a hybrid model for early detection of heart disease.	Three ML algorithms are used: RF, DT, and hybrid model.
[2]	2021	This study aims to determine the ML-based classification algorithms that provide the highest accuracy in diagnosing heart diseases.	Several supervised ML techniques, such as KNN, DT, and RF, were applied.
[40]	2020	This study aims to predict the patient's risk of acquiring heart disease	Four ML techniques are applied (NB, K-nearest neighbor, DT, and RF)
[39]	2019	The objective is to create a medical decision-support system that utilizes ML models to improve heart disease diagnosis accuracy.	NB classifier model

However, various improvements in classification techniques are still being made to increase prediction accuracy, particularly in the context of the DT model, which enables the extraction of decision rules that healthcare professionals can easily use.

3. Decision Tree-Basic Concept

One popular machine-learning approach for regression and classification tasks is the DT. It recursively splits the dataset into subsets based on each step's most significant attribute (i.e., feature).

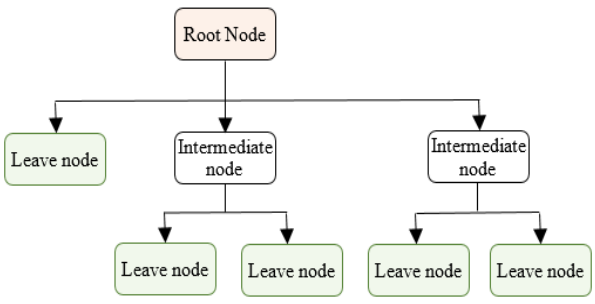

Figure 2. A simple explanation tree.

The objective is to construct a hierarchical arrangement in which every inner node signifies a choice or examination of a specific feature, and every terminal node signifies a predicted result or classification label. In Figure 2, we show the simple architecture of a DT and each node's definition.

The root serves as the initial point of consideration for the entire dataset, intermediate nodes represent decisions or conditions based on features, and leaf nodes represent the final predictions or outcomes of the DT. The DT's structure is built by recursively splitting the data into intermediate nodes until it reaches leaf nodes, where predictions are made. In the subsequent stages, we delineate constructing the DT.

1) Root node selection: the algorithm starts by selecting the feature that, when used to split the data, results in the best separation of classes or the highest reduction in impurity. Impurity measures disorder or uncertainty in the dataset.
2) Splitting data: the dataset is partitioned into subsets according to the values of the chosen feature. Each tree branch represents a specific value or range of values for the selected feature.
3) Recursive process: selecting the best feature and splitting the data is repeated recursively for each subset (child node). The algorithm continues to split the data until it satisfies one of the termination criteria: Reaching a specified maximum depth, attaining a minimum number of samples in a leaf node, or no longer improving the impurity measure.

4) Leaf node assignment: once a stopping condition is met, the algorithm assigns each leaf node a class label or a regression value. In classification tasks, the leaf node is assigned the most common class to determine its predicted class. In regression tasks, the predicted value for a leaf node is determined by assigning either the mean or median of the target values.
5) Prediction: the algorithm follows the path from the root node down to a leaf node by applying the same feature tests used during training to predict a new, unseen data point. The prediction at the leaf node is then assigned to the data point.
6) Handling missing data: DTs can handle missing data by assigning data points with missing values to the most common class or value at each split.
7) Model evaluation: whether the task at hand is regression or classification, measures like recall, accuracy, precision, F1-score, or mean squared error are often used to evaluate the model's performance.

DTs have fundamental benefits, such as interpretability, simplicity of visualization, and capacity to capture nonlinear relationships in the data [11, 31]. However, they can be prone to overfitting, in which the tree grows too intricately and adjusts to the noise in the training data. Methods such as pruning and establishing the maximum tree depth can help alleviate this issue.

4. Learning in Decision Trees

Within the framework of DTs, impurity serves as a metric for evaluating the degree of disorder or uncertainty inherent in a dataset or its constituent subsets. It aids in assessing how effectively a particular feature can partition the data into more consistent groups. The selection of an impurity measure depends on whether the task involves classification or regression.

- In classification tasks, impurity measures, such as the Gini index or entropy, quantify the degree of mixing of different class labels within a set of data points. The goal is to select features that minimize impurity, leading to well-defined and distinct groups.
- For regression tasks, impurity measures like variance assess the spread or dispersion of the target variable values within subsets. The objective is to identify features that result in subsets with reduced variance, contributing to more accurate and cohesive predictions.

4.1. Gini Impurity

Gini impurity [9] is a way to determine how likely it is to label an item from the dataset incorrectly based on how the classes are spread out in that subset. While constructing a DT, the algorithm assesses each potential feature's capacity to reduce impurity at each node and chooses the feature that maximizes impurity reduction as the splitting criterion. This recursive process continues to build the tree. It is calculated as follows for a node with multiple classes:

$$Gini(p) = 1 - \sum_{i=1}^{c} p_i^2 \quad (1)$$

Where the probability of a class i being present in the node is p_i and c is the number of classes.

Gini impurity ranges from 0 (perfectly pure, all samples belong to the same class) to 0.5 (maximum impurity, samples are evenly distributed across classes).

4.3. Entropy

Entropy [36] measures the disorder or randomness in a data set. It is calculated as follows:

$$Entropy(p) = -\sum_{i=1}^{c} p_i \log_2 p_i \quad (2)$$

The terms are the same as those for Gini impurity. Entropy ranges from 0 (perfectly pure) to 1 (maximum impurity).

5. Classification Process

Using a DT for classification is a standard and easy-to-understand ML method. It works like a tree, where each level represents a different characteristic, and the branches show various possibilities for that characteristic. The idea is to group data into specific categories or groups by making decisions at each tree level. It starts at the top and goes down through the tree until it reaches a final category at the bottom. Figure 3 shows a diagram that helps explain how this classification process works, especially when diagnosing heart disease.

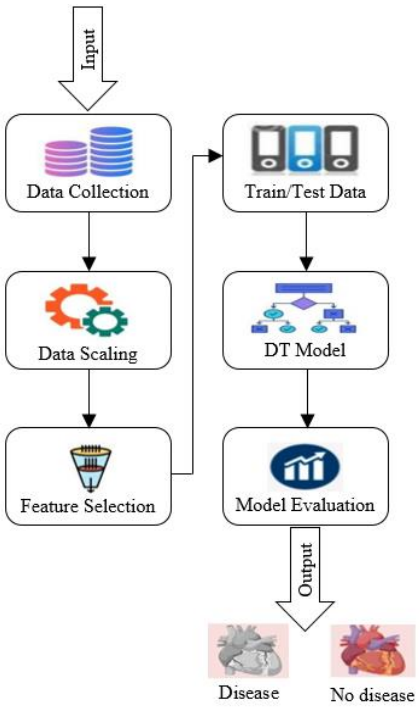

Figure 3. Proposed classification process.

5.1. Data Set Description

The dataset combines data from multiple sources, including hospitals in Cleveland, Hungary, Switzerland, and Long Beach, VA [18]. This dataset boasts 76 distinct attributes encompassing various aspects of a patient's health, which can be broadly categorized into three groups [14]:

- Demographic features: these include age, gender, and height, which provide a foundational understanding of the patient.
- Physiological features: these could encompass blood pressure (systolic and diastolic), cholesterol, and blood sugar levels, which represent the patient's cardiovascular health.
- Lifestyle features, including smoking habits, alcohol consumption, and physical activity level, which reveal potential risk factors associated with lifestyle choices.
- One of these attributes is the target variable, which typically indicates the presence or absence of heart disease (often coded as 0 or 1). Notably, all published experiments predominantly utilize a subset of 14 attributes listed in Table 2 that provide a more comprehensive depiction of the data. The heart disease dataset comprises 526 patients diagnosed with heart disease (51.32%) and 499 patients without (48.68%), as shown in Figure 4.

Table 2. Feature description.

Id	Features			
	Name	**Types**	**Description**	**Values Range**
id_1	Age	Numeric	Age in years	from 28 to 77; Mean: 51.9
id_2	Sex	Nominal	Gender	1 for male; 0 for female (206:1; 97:0)
id_3	Cp	Nominal	Chest pain type	1 for typical (23) 2 for atypical (50) 3 for nonanginal (86) 4 for asymptomatic (144)
id_4	Trestbps	Numeric	Resting blood pressure (mmHg)	from 94 to 200; Mean: 131.6
id_5	Chol	Numeric	Serum cholesterol (mg/dl)	from 126 to 564; Mean: 246.6
id_6	Fbs	Nominal	Fasting blood sugar	>120 mg/dl (1 for true; 0 for false), (45:1; 258:0)
id_7	Restecg	Nominal	Resting electrocardiographic results	0 for normal (151) 1 for STTWaveAbnormality (4) 2 for showingProbable (148)
id_8	Thalach	Numeric	Maximum heart rate achieved	from 71 to 202; Mean: 149.6
id_9	Exang	Nominal	Exercise-induced angina	(1 for yes; 0 for no) (99:1 204:0)
id_{10}	Oldpeak	Numeric	ST depression induced by exercise	from 0 to 6.2; Mean: 1.03
id_{11}	Slope	Nominal	Slope of peak exercise ST segment	1 for upsloping (142) 2 for flat (140) 3 for downsloping (21)
id_{12}	Ca	Nominal	Number of major vessels	0-3 (24:3; 38:2; 65:1; 176:0)
id_{13}	Thal	Nominal	Heart status	3 for Normal (168) 6 for Fixed defect (18) 7 for Reversible defect (117)
id_{14}	Target	Nominal	Output class	(1 for presence; 0 for absence) (139:1; 164:0)

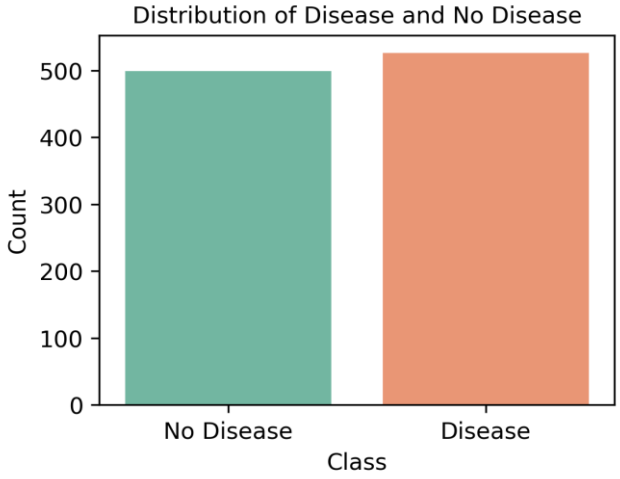

Figure 4. Classes (1: Disease and 0: No disease).

The Heart Disease Dataset is a valuable baseline for ML techniques in heart disease prediction. While its size of around 1025 instances offers a starting point, its true strength lies in its diverse composition. Therefore, the Heart Disease Dataset remains a crucial resource for researchers, paving the way for advancements in heart disease prediction through ML.

5.2. Data Scaling

Data scaling is a critical preparatory step in tackling classification problems, imparting improvements in prediction accuracy and model stability. In classification, the objective is to predict categorical labels based on input features. These features frequently display varying scales or units, impeding the efficient learning of the model. Our study utilized the "StandardScaler" utility from the scikit-learn library [30] to execute data standardization. This class operates by adjusting the input data to have a mean of 0 and a standard deviation 1. This normalization is accomplished by subtracting the feature mean from each data point and dividing the result by the feature's standard deviation. Mathematically, this traditional scaling process can be stated using Equation (3).

$$x_{scaled} = \frac{x - \bar{x}}{\sigma} \tag{3}$$

Here, x represents the initial feature value, $\bar{x}$ denotes the mean of the feature values, σ signifies the standard deviation of the feature values, and x_{scaled} corresponds to the scaled feature value achieved through the standard scaling process.

5.3. Feature Selection

Using dimensionality reduction, FS removes irrelevant characteristics that do not affect classifier performance from an extensive data collection [26, 35, 46]. Many FS methods have been successfully used to help in reducing the problem of dimensionality. FS aims to improve the effectiveness of data mining and analysis by identifying

relevant features while eliminating unrepresentative ones. Researchers have investigated several methods for selecting features and classifiers by utilizing a range of heart disease datasets available in the UCI Machine Learning Repository [48].

Dealing with diverse and diverse data becomes extremely important when using computer-based methods to diagnose various diseases. Model overfitting and increased training time are potential outcomes of high-dimensional data. Existing FS algorithms vary in criteria, and combining filter, wrapper, and embedded techniques has been underexplored [10, 33].

5.4. SelectKBest Feature Selection

SelectKBest is a FS technique commonly used in ML to choose the top K most informative features from a dataset [30]. It operates based on statistical tests designed explicitly for classification or regression tasks. The mathematics behind SelectKBest primarily involves statistical tests to evaluate the relevance of each feature concerning the target variable.

The most common statistical test used for classification tasks is the F-test, while for regression tasks, it's typically the ANOVA F-value. Here's a brief overview of how SelectKBest works mathematically:

1) Compute scores: for each feature in the dataset, SelectKBest computes a statistical score that quantifies the relationship between that feature and the target variable. It often uses the F-statistic for classification tasks, which measures the difference in means between classes relative to the variance within each class. Regression tasks use a similar ANOVA F-value. The formula for the F-statistic used in the F-test for classification tasks is as follows:

$$F = \frac{Between - class\ variance}{Within - class\ variance} \quad (4)$$

Where:

- "Between-class variance" measures the variance between different classes (groups) in the target variable.
- "Within-class variance" measures the variance within each class.

The F-statistic quantifies how different the means of various classes are compared to the variation within each class. Features that contribute significantly to class separability will have higher F-values.

2) Rank features: after calculating the scores for each feature, SelectKBest ranks the features based on these scores in descending order. Features with higher scores are considered more informative or relevant to the target variable.
3) Select top-k features: the final step is to select the top K features with the highest scores. SelectKBest considers these features the most relevant to the ML task. You specify the value of K when using SelectKBest.

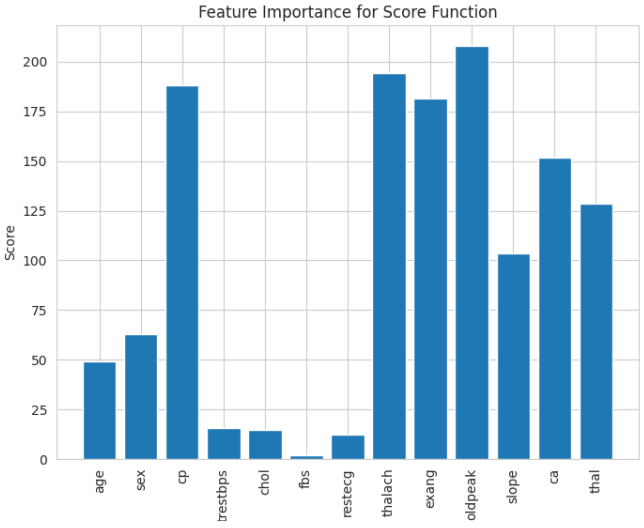

Figure 5. Feature importance.

Figure 5 shows the importance of each feature based on the score function. The 'oldpeak' feature has the highest importance value of 208.0028 in the data set. This implies that the ST Depression Induced by Exercise will likely significantly impact the decision model. Conversely, the feature of resting 'fbs' is the least significant among the features employed in the dataset, indicating that it is likely to have a reduced impact on the model.

According to our study, the best-selected features are 'Chest Pain Type (cp)', 'Maximum Heart Rate (thalach)', 'Exercise-Induced Angina (exang)', 'ST Depression Induced by Exercise (oldpeak)', and 'Number of Major Vessels Colored by Fluoroscopy (ca)'. In Figure 6, we show the correlation between the target class and the best-selected features in our case.

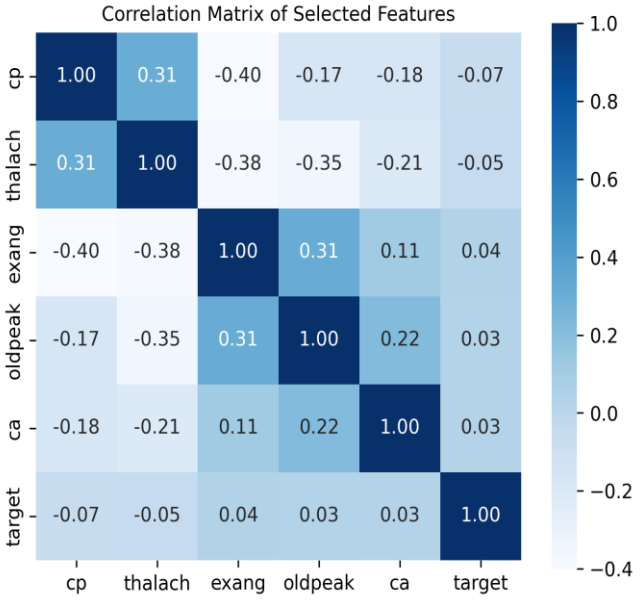

Figure 6. Correlation matrix for the selected features.

The proposed approach requires using FS to enhance the accuracy and efficiency of the DT model. We incorporate a FS method that identifies the most relevant patient characteristics for heart disease prediction. With more details, by focusing on the most informative features, the model can learn more effectively, achieve higher diagnostic accuracy, and reduce the DT's complexity, leading to a faster training process.

5.5. Train/Test

The Heart Disease dataset adopted in this study is known as a benchmark in ML for tasks associated with predicting heart disease [48]. It is a group of medical data for diagnosing individuals' heart disease. This work employed an 80/20 train-test split on the Heart Disease dataset to assess the proposed model's generalizability. The model was trained on 80% of the data (i.e., 820 instances) and evaluated on the remaining 20% (i.e., 205 instances). This split aimed to balance the need for sufficient training data with ensuring an unbiased evaluation set.

5.6. DT Model

In this study, the heart disease diagnosis model was developed using a pre-labeled dataset. The process had three main steps: collecting data, pre-processing it, and training the model. The "sklearn" library and the Python programming language were utilized for this comparison study. The analytical model was constructed within the Google Colab environment, which offers advantages for dataset exploration and facilitates the effective identification of patterns.

Figure 7 shows a sample model of the DT developed with a depth of 3 levels. The optimal tree and the calculated results were obtained with a DT with a depth of 7.

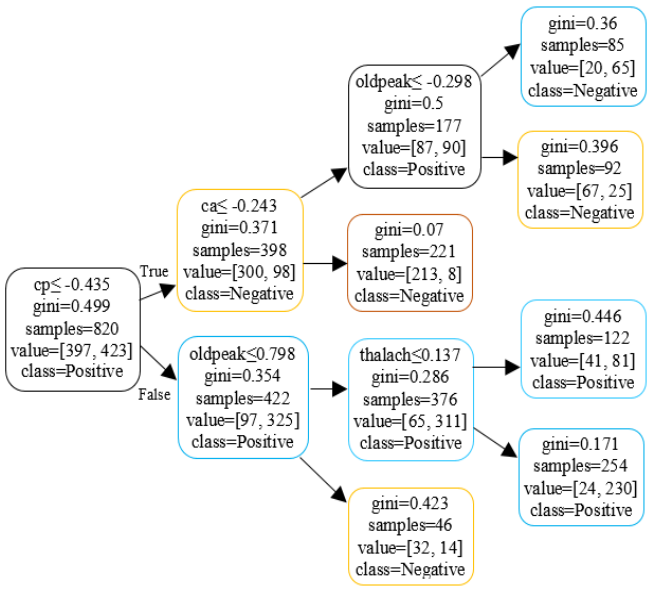

Figure 7. The DT model with depth 3.

5.7. Model Evaluation

Checking the performance of a classifier is an essential step in evaluating the effectiveness of a machine-learning model. The developed classification model exhibits strong performance on both the training and testing data sets based on the values of the following parameters:

1) True Positive (TP): the predicted and actual values are positive.
2) True Negative (TN): the predicted and actual values are both negative.
3) False Positive (FP): as opposed to the positive predicted value, the actual value is negative.
4) False Negative (FN): as opposed to the negative prediction, the actual value is positive.

Several metrics may be calculated using the following formulas to assess the DT classification model's performance:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (8)$$

Table 3. Confusion matrix of heart disease prediction.

Actual/Predicted	Heart disease	Normal
Heart disease	TP: The number of patients diagnosed with heart disease and are accurately predicted to have heart disease.	FN: The number of patients diagnosed as normal but with heart disease.
Normal	FP: The number of patients diagnosed with heart disease but normal in reality.	TN: The number of patients who are currently normal and predicted to be normal.

A confusion matrix, a common method for evaluating classification models, is shown in Table 3. The accuracy measure reflects the percentage of correct predictions made by the model. However, precision is a reliable assessment measure, especially when the proposed classification model has to evaluate its performance by comparing predicted and actual results. It computes the proportion of predicted outcomes that are indeed positive. Consequently, it depends on the values of TP and FPs. Recall is a valuable assessment statistic that helps identify the number of positives that can be accurately predicted. It measures the proportion of positives that are correctly classified. Recall is quantified by calculating the TP and FN values. The F1 score ensures a harmonious equilibrium between the accuracy and recall of a classifier.

6. Analysis of the Results

To demonstrate the DT classifier's effectiveness, a comparative analysis was conducted using various well-known ML techniques, such as Logistic Regression (LR), SVM, and Gaussian Naive Bayes (GNB).

Table 4 compares evaluation metrics across training and testing sets of various ML models. The DT model achieved the highest training accuracy (0.938) and F1-score (0.942) on the training data, indicating strong learning capabilities. It also achieved the highest training recall (0.986), suggesting it captured most actual heart disease cases during training. On the testing data, although there is a decrease in these metrics, the model still demonstrates strong performance with 89.32% (recall).

Table 4. Train/Test results of comparative ML models.

ML Model	Train				Test			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
LR	0.837805	0.820796	0.877069	0.848	0.785366	0.739837	0.883495	0.80531
SVM	0.87561	0.855876	0.91253	0.883295	**0.8**	**0.762712**	0.873786	**0.81448**
GNB	0.807317	0.807425	0.822695	0.814988	0.75122	0.720339	0.825243	0.769231
DT	**0.937805**	**0.902597**	**0.985816**	**0.942373**	0.795122	0.747967	**0.893204**	0.814159

Although the SVM has achieved the highest testing accuracy (0.80) and F1-score (0.81448) among the models, the DT performs well in testing with a minimal difference in testing accuracy and F1-score approximately less than 1%.

LR shows decent training performance in accuracy (0.838) and F1-score (0.848) but is lower than SVM and DT. In testing performance, LR has also achieved lower results than both SVM and DT, suggesting limitations in generalizability.

In contrast, GNB exhibits the lowest overall performance across training and testing data. Generally, the DT model generalizes reasonably to unseen data but may benefit from further optimization to enhance its performance in real-world applications.

Furthermore, Figure 8 illustrates the confusion matrix obtained from training and testing the DT.

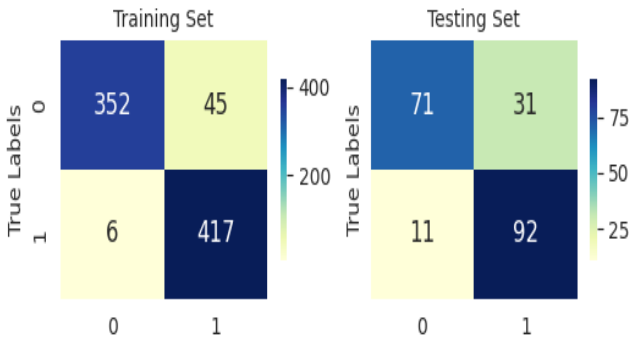

Figure 8. The DT confusion matrix.

In training, the model identified TP and TN cases well, correctly classifying 769 out of 820 instances. The model incorrectly predicted that 45 cases (or around 5.5% of the total) are patients with heart disease, while in fact, none of those patients have the disease. Excessive FNs in medical tests pose significant risks since they fail to indicate the presence of heart disease in those who have the disease. The DT model has recognized 6 cases out of 820 FN (i.e., approximately 0.73%), indicating the model is doing well at identifying most of the actual heart disease cases in the data. It's catching a good portion of the positive instances.

In testing, the model correctly predicted 71 instances where the patient did not have heart disease, resulting in TNs. Conversely, the model correctly predicted the presence of the disease 92 times, indicating a TP. The model incorrectly predicted 11 times that the patient did not have heart disease, indicating a *FN*, while predicting 31 instances of disease in the patient during the absence of the disease, indicating a *FP*. Overall, out of 205 instances, 82 stated "No' and 123 stated "Yes" to the prediction.

Figures 9, 10, and 11 show the relative confusion matrices of the comparison algorithms LR, SVM, and GNB, respectively.

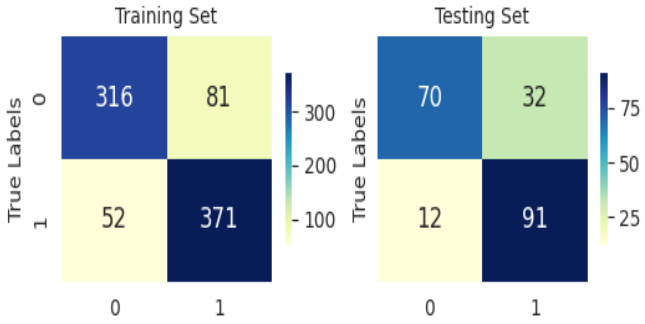

Figure 9. The LR Confusion matrix.

The LR model correctly classified 687 instances in training and 161 instances in testing. Although the model recognized 12 cases out of 205 FN (i.e., approximately 5.85%) in testing data, it recognized a larger number of FNs in training (i.e., 52 out of 820). Consequently, substantial risk is involved since it does not detect heart disease in patients with it.

According to Figure 10, the SVM model seems to be performing well. It correctly classified a significant number of cases equal to 386 as TP and 332 as TN in the training set. At the same time, 65 patients without heart disease were incorrectly classified as having it (i.e., inaccurate diagnosis of heart disease). In the case of the testing set, the model correctly classified 164 out of 205 patients with positive and negative disease and misclassified 41 patients (i.e., approximately 20% of test cases).

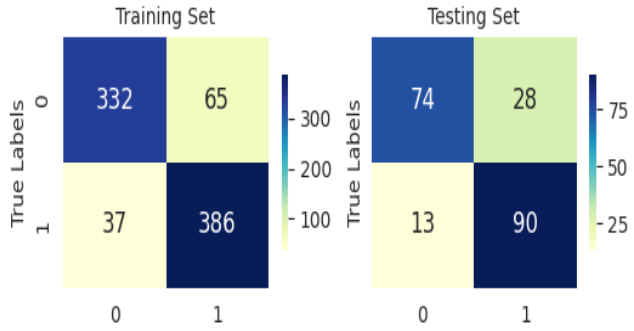

Figure 10. The SVM confusion matrix.

Based on the confusion matrix of the GNB classifier (as shown in Figure 11), it recognized that 75 patients with heart disease were misclassified as healthy in training and 18 in testing. Conversely, it counted 83 and 33 healthy people being flagged for heart disease in training and testing data, respectively.

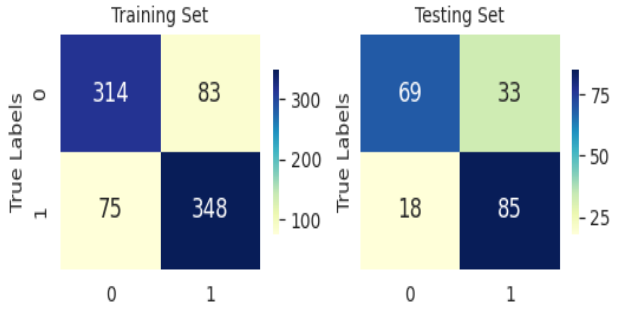

Figure 11. The GNB confusion matrix.

Additionally, Figure 12 visually evaluates the performance of the developed DT model using Receiver Operating Characteristic (ROC) curves against other comparative algorithms. One way to build it analytically is to plot the True Positive Rate (TPR) versus the False Positive Rate (FPR) at various threshold levels. The positioning of these curves relative to each other reflects the algorithms' efficiency levels in making predictions.

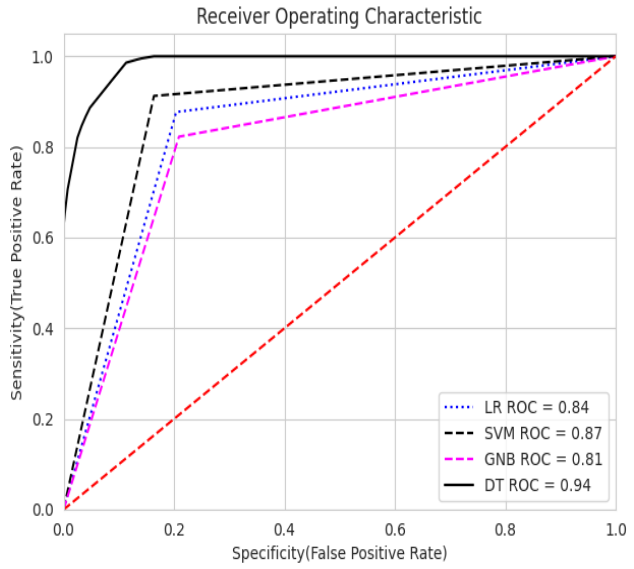

Figure 12. ROC curves of the comparative models.

The ROC curve shows that the developed DT model outperforms other comparative models, with an ROC equal to 0.94. DT's steeper initial rise suggests it can effectively differentiate between positive and negative classes at lower classification thresholds. This can be beneficial if you prioritize catching the most positive cases early, even if it leads to some FPs.

7. Conclusions

This study investigates the application of DTs as a supervised learning technique to diagnose heart disease. While DT achieved promising results, comparing it with other classifiers such as SVM, LR, and GNB is valuable. The developed diagnosis process involves data pre-processing, FS, model development, and model evaluation. We have demonstrated that a DT model with a FS process can yield an accurate diagnosis. It was found that accurately selecting the most appropriate features can improve our ability to diagnose and understand heart diseases, reducing the risk and enhancing patient care. The DT model achieved a superior performance accuracy of 93.78% compared with other models. These results show the advantages of utilizing it for effective diagnosis of heart disease. We discovered trees with various depths to harvest the finest diagnosis results.

References

[1] Agbemade E., "Predicting Heart Disease Using Tree-based Model," *Data Science and Data Mining*, University of Central Florida, 2023. https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1000&context=data-science-mining

[2] Ali M., Paul B., Ahmed K., Bui F., Quinn J., and Moni M., "Heart Disease Prediction Using Supervised Machine Learning Algorithms: Performance Analysis and Comparison," *Computers in Biology and Medicine*, vol. 136, pp. 104672, 2021. DOI:10.1016/j.compbiomed.2021.104672

[3] American Heart Association, Heart Disease and Stroke Statistics Update Fact Sheet at-a-Glance-2024, https://www.heart.org/-/media/PHD-Files-2/Science-News/2/2024-Heart-and-Stroke-Stat-Update/2024-Statistics-At-A-Glance-final_2024.pdf, Last Visited, 2024.

[4] Anderies A., Tchin J., Putro P., Darmawan Y., and Gunawan A., "Prediction of Heart Disease UCI Dataset Using Machine Learning Algorithms," *Engineering, Mathematics and Computer Science Journal*, vol. 4, no. 3, pp. 87-93, 2022. https://doi.org/10.21512/emacsjournal.v4i3.8683

[5] Bhatt C., Patel P., Ghetia T., and Mazzeo P., "Effective Heart Disease Prediction Using Machine Learning Techniques," *Algorithms*, vol. 16, no. 2, pp. 1-14, 2023. https://doi.org/10.3390/a16020088

[6] Biswas N., Ali M., Abdur Rahman M., Islam M., Mia M., Azam S., Ahmed K., Bui F., Al-Zahrani F., and Moni M., "Machine Learning-based Model to Predict Heart Disease in early Stage Employing Different Feature Selection Techniques," *BioMed Research International*, vol. 2023, pp. 1-15, 2023. https://doi.org/10.1155/2023/6864343

[7] Bond K. and Sheta A., "Medical Data Classification Using Machine Learning Techniques," *International Journal of Computer Applications*, vol. 183, no. 6, 2021. DOI:10.5120/ijca2021921339

[8] Boukhatem C., Youssef H., and Nassif A., "Heart Disease Prediction Using Machine Learning," *in Proceedings of the Advances in Science and Engineering Technology International Conferences*, Dubai, pp. 1-6, 2022. https://ieeexplore.ieee.org/document/9734880

[9] Breiman L., Friedman J., Olshen R., and Stone C., *Classification and Regression Trees*, Chapman and Hall/CRC, 1984. https://doi.org/10.1201/9781315139470

[10] Chen C., Tsai Y., Chang F., and Lin W., "Ensemble Feature Selection in Medical Datasets: Combining Filter, Wrapper, and Embedded Feature Selection Results," *Expert Systems*, vol. 37, no. 5, pp. e12553, 2020. https://doi.org/10.1111/exsy.12553

[11] Chrimes D., "Using Decision Trees as an Expert System for Clinical Decision Support for Covid-19," *Interactive Journal of Medical Research*, vol. 12, no. 1, pp. 1-12, 2023. https://pubmed.ncbi.nlm.nih.gov/36645840/

[12] Cios K., *Medical Data Mining and Knowledge Discovery*, Springer, 2001. https://link.springer.com/book/9783790813401

[13] Dissanayake K. and Johar M., "Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms," *Applied Computational Intelligence and Soft Computing*, vol. 2021, pp. 1-17, 2021. https://doi.org/10.1155/2021/5581806

[14] Dua D. and Graff C., UCI Machine Learning Repository-2017, http://archive.ics.uci.edu/ml, Last Visited, 2024.

[15] Elbasi E. and Zreikat A., "Heart Disease Classification for Early Diagnosis Based on Adaptive Hoeffding Tree Algorithm in IoMT Data," *The International Arab Journal of Information Technology*, vol. 20, no. 1, pp. 38-48, 2023. https://doi.org/10.34028/iajit/20/1/5

[16] Franklin R. and Muthukumar B., "Survey of Heart Disease Prediction and Identification Using Machine Learning Approaches," *in Proceedings of the 3rd International Conference on Intelligent Sustainable Systems*, Thoothukudi, pp. 553-557, 2020. DOI: 10.1109/ICISS49785.2020.9316119

[17] Heart Disease Dataset on Kaggle, https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset, Last Visited, 2024.

[18] Janosi A., Steinbrunn W., Pfisterer M., Detrano R., UCI Machine Learning, Heart Disease Repository-1988, https://doi.org/10.24432/C52P4X, Last Visited, 2024.

[19] Kavitha M., Gnaneswar G., Dinesh R., Sai Y., and Suraj R., "Heart Disease Prediction Using Hybrid Machine Learning Model," *in Proceedings of the 6th International Conference on Inventive Computation Technologies*, Coimbatore, pp. 1329-1333, 2021. https://ieeexplore.ieee.org/document/9358597

[20] Kodati S. and Vivekanandam R., "Analysis of Heart Disease Using in Data Mining Tools Orange and Weka," *Global Journal of Computer Science and Technology*, vol. 18, no. C1, pp. 17-21, 2018. https://globaljournals.org/GJCST_Volume18/4-Analysis-of-Heart-Disease.pdf

[21] Krishnan S. and Geetha S., "Prediction of Heart Disease Using Machine Learning Algorithms," *in Proceedings of the 1st International Conference on Innovations in Information and Communication Technology*, Chennai, pp. 1-5, 2019. DOI: 10.1109/ICIICT1.2019.8741465

[22] Li J., Ul Haq A., Ud Din S., Khan J., Khan A., and Saboor A., "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," *IEEE Access*, vol. 8, pp. 107562-107582, 2020. DOI:10.1109/ACCESS.2020.3001149

[23] Maheswari S. and Pitchai R., "Heart Disease Prediction System Using Decision Tree and Naive Bayes Algorithm," *Current Medical Imaging Reviews*, vol. 15, no. 8, pp. 712-717, 2019. DOI:10.2174/1573405614666180322141259

[24] Maji S. and Arora S., *Decision tree Algorithms for Prediction of Heart Disease*, Springer, 2019. https://link.springer.com/chapter/10.1007/978-981-13-0586-3_45

[25] Mc Namara K., Alzubaidi H., and Jackson J., "Cardiovascular Disease as a Leading Cause of Death: How are Pharmacists Getting Involved?," *Integrated Pharmacy Research and Practice*, vol. 8, pp. 1-11, 2019. DOI:10.2147/IPRP.S133088/

[26] Miao J. and Niu L., "A Survey on Feature Selection," *in Proceedings of the 4th International Conference on Information Technology and Quantitative Management, Promoting Business Analytics and Quantitative Management of Technology*, Seoul, pp. 919-926, 2016. https://doi.org/10.1016/j.procs.2016.07.111

[27] Münzel T., Hahad O., Sørensen M., Lelieveld J., Duerr G., Nieuwenhuijsen M., and Daiber A., "Environmental Risk Factors and Cardiovascular Diseases: A Comprehensive Expert Review," *Cardiovascular Research*, vol. 118, no. 14, pp.

2880-2902, 2022. https://doi.org/10.1093/cvr/cvab316

[28] Nawsherwan., Wang B., Zhang L., Sumaira M., Guo F., and Yan W., "Prediction of Cardiovascular Diseases Mortality and Disability-Adjusted Life-Years Attributed to Modifiable Dietary Risk Factors from 1990 to 2030 among East Asian Countries and the World," *Frontiers in Nutrition*, vol. 9, pp. 1-12, 2022. DOI:10.3389/fnut.2022.898978

[29] Nichenametla R., Maneesha T., Hafeez S., and Krishna H., "Prediction of Heart Disease Using Machine Learning Algorithms," *International Journal of Engineering and Technology*, vol. 7, no. 5, pp. 363-366, 2018. DOI:10.14419/ijet.v7i2.32.15714

[30] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., and Grisel O., "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011. http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

[31] Podgorelec V., Kokol P., Stiglic B., and Rozman I., "Decision Trees: An Overview and their Use in Medicine," *Journal of Medical Systems*, vol. 26, no. 5, pp. 445-463, 2002. DOI:10.1023/a:1016409317640

[32] Prather J., Lobach D., Goodwin L., Hales J., Hage M., and Hammond W., "Medical Data Mining: Knowledge Discovery in a Clinical Data Warehouse," *in Proceedings of the AMIA Annual Fall Symposium*, Nashville, pp. 101-105, 1997. https://www.ncbi.nlm.nih.gov/pmc/issues/160771/

[33] Pudjihartono N., Fadason T., Kempa-Liehr A., and O'Sullivan J., "A Review of Feature Selection Methods for Machine Learning-based Disease Risk Prediction," *Frontiers in Bioinformatics*, vol. 2, pp. 1-17, 2022. https://doi.org/10.3389/fbinf.2022.927312

[34] Purushottam., Saxena K., and Sharma R., "Efficient Heart Disease Prediction System Using Decision Tree," *in Proceedings of the International Conference on Computing, Communication Automation*, Greater Noida, pp. 72-77, 2015. https://ieeexplore.ieee.org/document/7148346

[35] Qiao Q., Yunusa-Kaltungo A., and Edwards R., "Feature Selection Strategy for Machine Learning Methods in Building Energy Consumption Prediction," *Energy Reports*, vol. 8, pp. 13621-13654, 2022. https://doi.org/10.1016/j.egyr.2022.10.125

[36] Quinlan J., "Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81-106, 1986. https://link.springer.com/article/10.1007/BF00116251

[37] Radhika R. and George S., "Heart Disease Classification Using Machine Learning Techniques," *in Proceedings of the International Conference on Novel Approaches and Developments in Biomedical Engineering*, Coimbatore, pp. 1-9, 2021. DOI: 10.1088/1742-6596/1937/1/012047

[38] Rani P., Gujral R., Sid Ahmed N., and Jain A., "A Decision Support System for Heart Disease Prediction Based upon Machine Learning," *Journal of Reliable Intelligent Environments*, vol. 7, no. 3, pp. 263-275, 2021. https://link.springer.com/article/10.1007/s40860-021-00133-6

[39] Repaka A., Ravikanti S., and Franklin R., "Design and Implementing Heart Disease Prediction Using Naives Bayesian," *in Proceedings of the 3rd International Conference on Trends in Electronics and Informatics*, Tirunelveli, pp. 292-297, 2019. DOI:10.1109/ICOEI.2019.8862604

[40] Shah D., Patel S., and Bharti S., "Heart Disease Prediction Using Machine Learning Techniques," *SN Computer Science*, vol. 1, no. 6, pp. 345, 2020. https://doi.org/10.1007/s42979-020-00365-y

[41] Sharma H. and Kumar S., "A Survey on Decision Tree Algorithms of Classification in Data Mining," *International Journal of Science and Research*, vol. 5, no. 4, pp. 2094-2097, 2016. file:///C:/Users/user/Downloads/1221663df46568d5e1edf3e0476d1d2422cc.pdf

[42] Shilpa K. and Adilakshmi T., "An Enhanced Machine Learning Technique to Predict Heart Disease," *in Proceedings of the 2nd International Conference on Cognitive and Intelligent Computing*, Hyderabad, pp. 177-183, 2023. https://link.springer.com/chapter/10.1007/978-981-99-2742-5_19

[43] Shouman M., Turner T., and Stocker R., "Using Data Mining Techniques in Heart Disease Diagnosis and Treatment," *in Proceedings of the Japan-Egypt Conference on Electronics, Communications and Computers*, Alexandria, pp. 173-177, 2012. DOI:10.1109/JEC-ECC.2012.6186978

[44] Shouman M., Turner T., and Stocker R., "Using Decision Tree for Diagnosing Heart Disease Patients," *in Proceedings of the 9th Australasian Data Mining Conference*: *Australian Computer Society*, Ballara, pp. 23-30, 2011. https://dl.acm.org/doi/pdf/10.5555/2483628.2483633

[45] Spencer R., Thabtah F., Abdelhamid N., and Thompson M., "Exploring Feature Selection and Classification Methods for Predicting Heart Disease," *Digital Health*, vol. 6, no, 2, pp. 1-10, 2020. DOI:10.1177/2055207620914777

[46] Suresh S., Newton D., Everett T., Lin G., and Duerstock B., "Feature Selection Techniques for a Machine Learning Model to Detect Autonomic Dysreflexia," *Frontiers in Neuroinformatics*, vol.

16, pp. 1-10, 2022. DOI:10.3389/fninf.2022.901428

[47] Tu M., Shin D., and Shin D., "Effective Diagnosis of Heart Disease through Bagging Approach," *in Proceedings of the 2nd International Conference on Biomedical Engineering and Informatics*, Tianjin, pp. 1-4, 2009. DOI:10.1109/BMEI.2009.5301650

[48] UCI Machine Learning Repository, https://archive.ics.uci.edu/ml/index.php, Last Visited, 2024.

[49] Vijaya Saraswathi R., Gajavelly K., Kousar Nikath A., Vasavi R., and Reddy Anumasula R., "Heart Disease Prediction Using Decision Tree and SVM," *in Proceedings of the 2nd International Conference on Advances in Computer Engineering and Communication Systems*, Hyderabad, pp. 69-78, 2022. https://link.springer.com/chapter/10.1007/978-981-16-7389-4_7#citeas

[50] Yang J., Li Y., Liu Q., Li L., Feng A., Wang T., Zheng S., Xu A., and Lyu J., "Brief Introduction of Medical Database and Data Mining Technology in the Big Data Era," *Journal of Evidence-based Medicine*, vol. 13, no. 1, pp. 57-69, 2020. https://doi.org/10.1111/jebm.12373

Alaa Sheta obtained his Ph.D. in Information Technology from George Mason University, Virginia, USA, in 1997. Before that, he earned his B.E. and M.Sc. degrees in Electronics and Communication Engineering from Cairo University, Egypt in 1988 and 1994, respectively. He holds a tenured professorship at the Computer Science Department of Southern Connecticut State University in New Haven, Connecticut, USA. Dr. Sheta's research interests span various areas, including machine and deep learning, meta-heuristics, data science, image processing, and robotics. He has significantly contributed to these fields, publishing more than 170 papers in international journals and conferences. In addition to his scholarly achievements, he has actively participated in various capacities, such as serving as a chair, guest editor, and program committee member for numerous international events.

Walaa El-Ashmawi is an Associate Professor at the Faculty of Computer Science, Misr International University, on leave from Suez Canal University, Egypt. She received her B.Sc. in 2001 and M.Sc. in 2008 from Egypt. She got her Ph.D. She earned her degree in 2013 from the Computer Science Department, College of Information Science and Engineering, Hunan University, China. Her research interests Span Various Domains in Computer Science including Artificial and Computational Intelligence, Meta-heuristic algorithms Optimization problems, and Machine Learning techniques. She authorizes over 40 published papers in various international journals and conferences. She led different administrative and academic positions, both full-time and part-time.

AbdelKarim Baareh holds a Doctor of Informatics/Artificial Intelligence degree from Damascus University, Syria, which he earned in 2009. He also completed his B.Sc. and M.C.A (Computer Application) degrees from Mysore and Bangalore University, India, in 1992 and 1999, respectively. Professor Baara is on sabbatical leave and affiliated with the Data Science and Artificial Intelligence Department at the Information Technology College, Isra University in Jordan. Additionally, he holds a permanent academic position within the Computer Science Department at Al-Balqa Applied University, Jordan. Dr. Baara's extensive research portfolio spans various domains, including meta-heuristics, global optimization, machine learning, data mining, bioinformatics, graph theory, and parallel programming. Over the years, he has contributed significantly to his field, with a publication record that boasts over 35 papers in international journals and conferences.