Article
JOURNAL OF UNIVERSITY OF BABYLON
For Pure and Applied Sciences (JUBPAS)
Vol. 34 ; No.2 | 2026

# Advanced Heart Disease Detection Using Stacked Machine Learning Classifiers

**Evan Madhi Hamzh Al Rubaie [1*], Mehdi Ebady Manaa[2]**

[1]College of Engineering, University of Babylon, Babylon, 51002, Iraq
[2] Department of Information Networks, College of Information Technology, University of Babylon, Hillah, 51002, Babylon, Iraq.
Mahdi.ebadi@uomus.edu.iq

# الكشف المتقدم عن أمراض القلب باستخدام مصنفات التعلم الآلي المكدسة

**ايفان ماضي حمزة الربيعي [1*]، مهدي عبادي مانع**

1 كلية الهندسة ، جامعة بابل _ بابل ، العراق ، 51002 ، العراق
eng.Evan.rubae@uobabylon.edu.iq
2 قسم شبكات المعلومات ، كلية تكنولوجيا المعلومات ، جامعة بابل _ الحلة ، 51002، بابل ، العراق
3 قسم الانظمة الطبية الذكية ، كلية العلوم ، جامعة المستقبل ، 51001 ، بابل ، العراق
mahdi.ebadi@uomus.edu.iq

| Accepted: | 16/4/2026 | Published: | 30/6/2026 |
|---|---|---|---|

## ABSTRACT

Background: Cardiovascular disease remains a primary global health concern, highlighting the urgent need for reliable and widely applicable diagnostic tools. This work aims to present an Advanced Heart Disease Detection system that integrates clinical data from four major international datasets: the Cleveland, Hungarian, Switzerland, Long Beach VA datasets, and Statlog Heart Diseases for CAD and No-CAD diseases. The system utilizes 14 clinically-relevant standardized features and employs an automated preprocessing pipeline that includes median imputation for numeric data, “most frequent” imputation for categorical attributes, and standard scaling for data integrity.

Methods: The proposed methodology is implemented in two phases. The first phase is learned with Deep Q-Network, which achieved high accuracy compared with others. The second phase is a Stacked Ensemble voting algorithm, which is synthesized by a Logistic Regression, Support Vector Machine (SVM), XGBoost (XGB), Multilayer Perceptron, and Random Forest (RF). The designed model has increased the performance accuracy by using extensive hyperparameter tuning with GridSearchCV. The results show high accuracy compared with other studies.

Results: The first model achieved 92.65%, while the second model achieved 92.52% % for DQN and ensemble methods, respectively. The proposed model demonstrated high and reliable performance, indicating its strong and reliable tool for early detection of heart diseases, partially for classification into CAD and No CAD and its applicability in real-world clinical decision support systems.

**Keywords:** Coronary Artery Disease (CAD), Data Preprocessing, Deep Reinforcement Learning (DQN), Machine Learning, Ensemble Learning, Clinical Decision Support



ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com
مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com



## 1. INTRODUCTION

Cardiovascular diseases (CVDs) remain the leading cause of global death in many countries. It represents a large percentage of total deaths and places a considerable economic and social burden on healthcare systems. Early and accurate detection of patients with CVDs is essential to improve patient outcomes and treatment options, enabling timely medical intervention and reducing long-term healthcare costs, especially for patients with developed CVDs. However, traditional methods for diagnosing CVDs rely heavily on the experience, manual interpretation and expert systems for patients. These traditional testing procedures and subjective interpretations of the patient's signs and symptoms lead to delays in diagnosis, which can be time-consuming for patient treatments[1,2,3].

In recent years, machine learning (ML) in computer science and engineering has become a powerful paradigm, assisting clinical decision-making by automating and analyzing the medical data. Machine learning plays an integral role in identifying patient patterns by learning from the complex, nonlinear relationships between clinical characteristics and disease progression. It supports clinicians in identifying complex patterns that are difficult to discover and prone to diagnostic errors. Various classification algorithms, including Support Vector Machines (SVM), Random Forest (RF), Logistic Regression (LR), Multilayer Perceptron (MLP) and Extreme Gradient Boosting (XGBoost), have shown promising applications in predicting cardiac diseases using structured clinical datasets [4,5,6].

Although substantial progress has been made in the machine learning field, developing predictive models with good generalization and accuracy is still a difficult challenge for healthcare applications and CAD disease due to several reasons, including 1) data heterogeneity, 2) class imbalance, 3) missing data, and 4) variable information collected from different clinical sources. Most earlier studies used only one dataset, which could limit the generalizability of developed predictive models. In addition to this limitation, other factors such as inadequate preprocessing of training data, insufficient hyperparameter tuning, or the possibility of data leakage when training predictive models could lead to an overestimation of the model's performance, which in turn will decrease the performance of the predictive models when they are applied to patients in real-world situations [7-11].

To address these problems, this study presents a rigorous and comprehensive machine learning methodology for predicting heart disease that utilizes a variety of clinical datasets from multiple hospitals worldwide. A comprehensive data preprocessing system, encompassing missing value imputation, feature scaling, and categorical encoding, will be employed to establish a unified pipeline, ensuring consistency and preventing leakage during model training. In addition to establishing data completeness through pre-processing, the proposed methodology will also apply extensive hyperparameter tuning with cross-validated grid search on several classification methods to determine the optimal hyperparameter configurations of each classifier and produce the maximum predictive performance.

Moreover, this study evaluates and compares deep reinforcement learning using Q-Network algorithms and advanced machine learning classification models [12], [13]. Thus, providing a

ISSN: 2312-8135 | Print ISSN: 1992-0652 | www.journalofbabylon.com | jub@itnet.uobabylon.edu.iq | info@journalofbabylon.com



systematic and fair comparison for performance evaluation. The experimental framework will allow for reproducibility, robustness, and applicability to real-world clinical settings; training and tagging will be clearly separated, as will be the use of cross-validation for model tuning. The main contributions of this work can be summarized as follows:

1. Handling leakage-free preprocessing and training pipelines for CAD and No-CAD patients.
2. Integration of multiple clinical datasets from different medical institutions to improve model generalization.
3. Using cross-validated grid search for hyperparameter optimization of multiple machine learning models.
4. Using Deep Q-learning algorithms for multiple datasets to generate a knowledge experience for further classification.
5. Comparative evaluation of several state-of-the-art machine learning algorithms and identification of the most effective model for heart disease for patients with CAD and No-CAD binary classification.

The study provides evidence that well-tuned machine learning models can produce predictions with high accuracy and support decision making for diagnosing heart disease. This work demonstrates the importance of systematic preprocessing of data as well as optimization of model parameters for creating clinically relevant and robust predictive systems through the use of multiple datasets.

## 2. LITERATURE REVIEW

The study in machine Learning has been comprehensively researched as a powerful prediction tool for cardiovascular diseases, with many studies examining a variety of classification algorithms, preprocessing methods, and optimization approaches. This section presents a critical review of the most significant recent studies that are closely related to this research framework.

The work presented by [14] using machine-learning algorithms, including XGBoost, bags of trees or random forest machine-learning algorithms, achieved results for random forest with the highest accuracy of detection risk of heart disease with approximately 93%, while bagged trees had approximately 93% and lastly random forests had approximately 91%. Ensemble methods also ranked highest for ROC-AUC performance, with ensemble methods demonstrating better ability to capture the complex/non-linear relationship in medical data than all others tested (PMC). The study confirms that ensemble learning methods, such as XGBoost, are highly effective for solving medical classification problems. However, the authors do not show the preprocessing techniques or the train/test separation that will impact how well these results apply in real-world situations.

The study implemented by [15]. They used the Cleveland dataset to compare Logistic Regression and Random Forest. The obtained results show that the random forest was able to obtain approximately 91% accuracy, a much higher value than the 83% of Logistic Regression. Furthermore, Random Forest excelled in terms of delivery in Precision, Recall and ROC-AUC

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com



scores. The classification performance improved due to the Random Forest's effectiveness, which is based on a tree-based ensemble method in handling nonlinear feature interactions.

A comprehensive comparative study by [16]. They evaluated using multiple machine learning models SVM, Naïve bases, logistic regression, neural network, decision tree, and KNN. The study found that the SVM achieved the highest predictive accuracy with 85% superior other classifiers in precision and reliability. These results highlight the effectiveness SVM model as a baseline model and validates its effectiveness for heart disease classification tasks.

An optimized machine-learning method using the XGBoost developed by [17]. They presented a solution for heart disease using both traditional heart disease datasets with hyperparameter optimization and private datasets. The study used Naive Bayes, SVM, voting, XGBoost, AdaBoost, bagging, DT, KNN, RF, and LR for training and testing these models. The performance of this optimized classifier using dataset integration is 97.57% accurate. This study confirms that hyperparameter tuning and dataset combination significantly improve predictive performance, supporting the methodology used in the proposed system.

The study presented in [18] evaluated machine learning models using combined datasets from Cleveland, Hungary, Switzerland, and VA Medical Center. It is demonstrating that integrating multiple datasets improves predictive accuracy and model robustness. The results show improved generalization and reduced dataset bias. The dynamic kernel explored weighted linear and nonlinear terms, where the polynomial degree of 3 achieved the highest accuracy with 93.3%.

The study presented by [19] evaluated several machine learning algorithms using heart disease datasets. The results show the random forest 89% , SVM 82%, and XGBoost 79 %. The study finds the ensemble models outperformed single classifies and highlights the importance of ensemble and optimized models in improving classification accuracy.

The study by [20] presented a machine learning-based framework for heart disease prediction using the heart-disease dataset with hyperparameter tuning. Random Forest achieved approximately 91% accuracy and a high F1-score performance. The study finds that the hyperparameter optimization in improving classification play a critical role for model reliability and accuracy.

The accuracy of previous research on using machine learning for heart disease prediction has been shown; however, there are still several gaps that need to be addressed within this area of study:

1. Most studies were limited to only one data source; therefore, the results could not be generalized to other settings.
2. Many of the reviewed studies did not have any safeguards against potential data leakage during their respective preprocessing steps.
3. Moreover, a majority of the studies did not properly optimize their models' hyperparameters.
4. Extremely few studies use multiple data sources from many different clinical settings when evaluating model performance.

ISSN: 2312-8135 | Print ISSN: 1992-0652 | www.journalofbabylon.com | jub@itnet.uobabylon.edu.iq | info@journalofbabylon.com



The proposed framework resolves these limitations through:

A. The integration of four clinical datasets (Cleveland, Hungary, Switzerland, V.A.)
B. Implementing a preprocessing pipeline that does not require any leaks
C. The implementation of extensive hyperparameter tuning with GridSearchCV
D. Evaluating multiple optimized machine learning models and Deep Q-network under identical conditions
E. Ensuring reproducibility and generalizability

## 3. THE PROPOSED SYSTEM AND METHODOLOGY

The proposed system consists of two phases. The first phase proposes Deep Q-Network-based intelligent diagnostic system (DQIDS-CAD) for the five CAD datasets. The second phase uses the machine learning algorithms Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Multi-Layer Perceptron (MLP), and XGBoost (XGB). The datasets are loading using the Cleveland, Hungarian, Switzerland, Long Beach VA datasets, and Statlog Heart Diseases from UCI [21]. The datasets are preprocessed using missing value imputation, binary transformation, feature normalization, and 10-fold cross-validation. The deep Q-Network model includes many internal states and the final step for model evaluation is shown in Figure (1) to describe the architecture of the proposed Deep Q-Network for heart disease diagnosis.

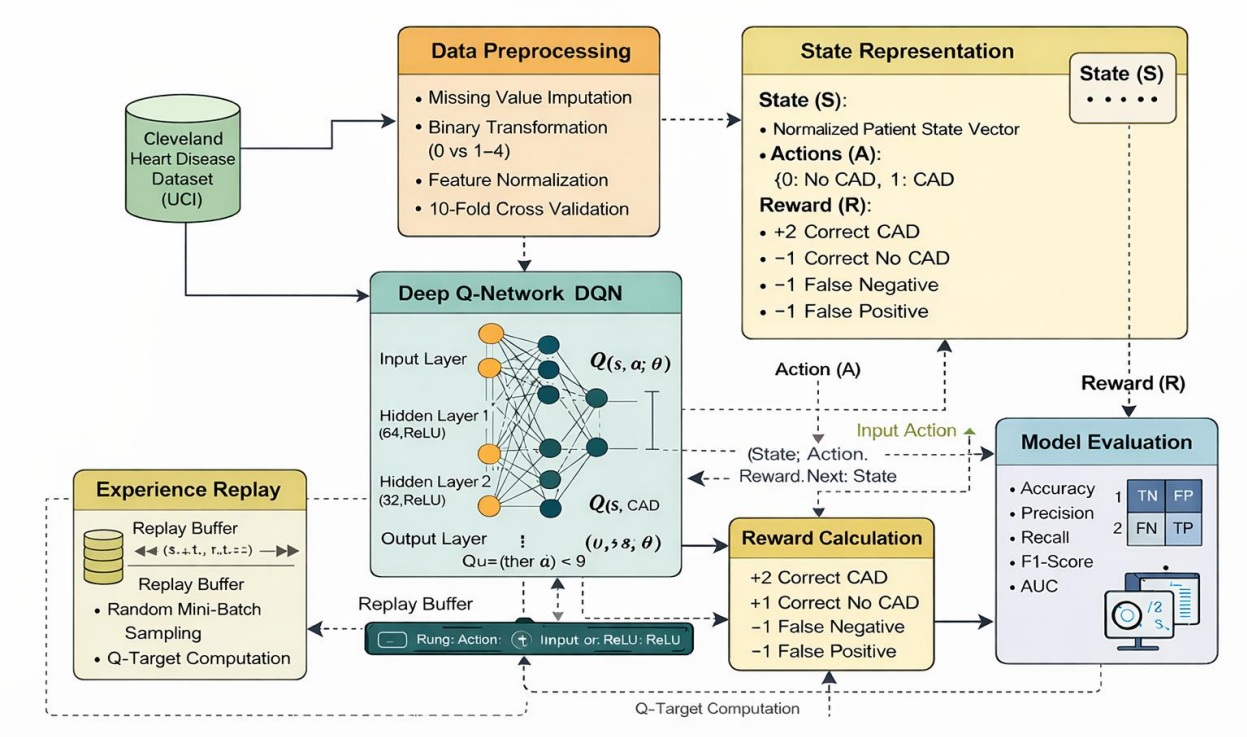


**Figure 1:** Architecture of the Proposed Deep Q-Network-based intelligent diagnostic system for coronary artery disease (DQIDS-CAD)

The second phase proposes the machine learning and ensemble voting algorithms to combine probabilistic outputs of all phases and find the final result. Preprocessing steps and model

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com



construction are shown in figure (2), which shows the architecture of the machine learning uses many machine classifiers and ensemble voting algorithms.

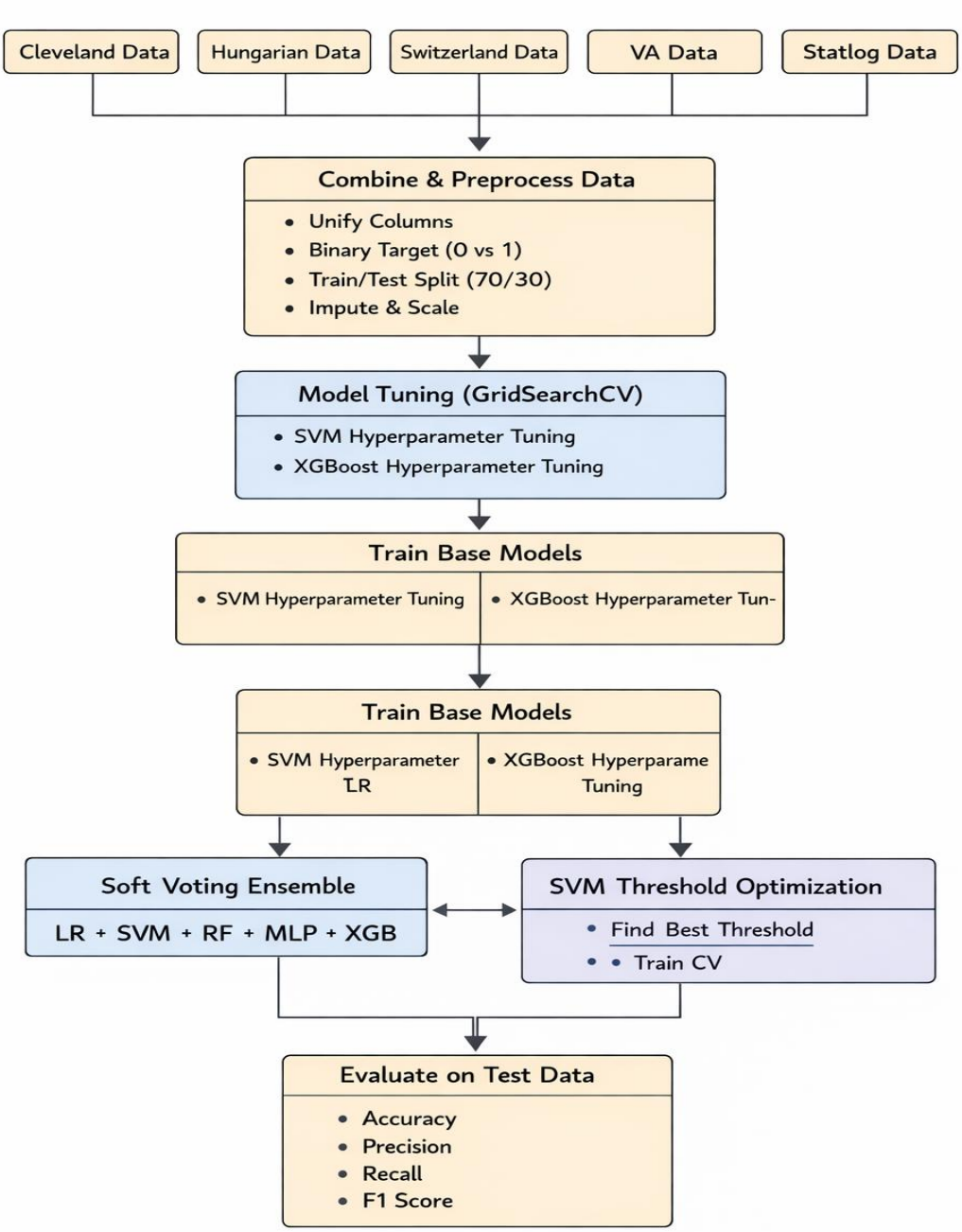


**Figure 2:** Architecture of the Proposed Heart Disease Ensemble System for coronary artery disease (DQIDS-CAD)

The following subsection describes the main mathematical model for the proposed Deep Q-Network-based intelligent diagnostic systems for coronary artery disease (DQIDS-CAD) and machine learning classifiers.

ISSN: 2312-8135 | Print ISSN: 1992-0652 | www.journalofbabylon.com | jub@itnet.uobabylon.edu.iq | info@journalofbabylon.com



### 3.1. Data Acquisition and Preprocessing

The system uses the Cleveland heart disease dataset from the UCI repository. The feature vector consists of corresponding patient described instances in equation (1)

$$X_i = [x_{i1}, x_{i2}, x_{i3}, \dots, x_{id}] \in \mathrm{R}^{\mathrm{d}} \dots (1)$$

Where,

- ***d*** is the number of the patient attributes
- ***i=1,2,….,N*** indexed for each patient row
- ***N*** is the total number of samples.

The target values include five values {0,1,2,3,4), indicating severity levels of CAD disease, where the 0 value indicates NOCAD.

The missing value $\boldsymbol{M_j}$ handling using the median in equation (2) for the numerical value.

$$x_{i,j} = median(x.j), \forall\, i \in \boldsymbol{M_j} \dots (2)$$

### 3.2. Data Binary Transformation

The target value consists of severity levels {0,1,2,3,4}, then, we used the following threshold to transform it into binary values in equation (3).

$$y_i = \begin{cases} 0 \quad if\ original\ taget\ value = 0 \quad (NO\ CAD) \\ 1\ if\ original\ taget\ value \in \{1,2,3,4\}\ (CAD) \end{cases} \quad \dots (3)$$

The equation in (3) converts the target value $y_i = \{0,1\}$, each numerical value is normalized using equation (4)

$$\tilde{x}_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{.j}) - \min(x.j)} \quad \dots (4)$$

Where,

$\boldsymbol{max(x_j)}$ represents the maximum value of column feature j across all samples used in this column.

$\boldsymbol{min(x.j)}$ represents the minimum value of column feature j across all samples used in this column.

The training/ testing using 10-fold validation partitions the dataset into 10 disjoint subsets in equation (5)

$$D = \bigcup_{k=1}^{10} D_k \quad for\ each\ fold\ k \dots (5)$$

Then, training/ testing and final metrics is defined as below:

$$D_{train} = \frac{D}{D_k} \quad \dots (6)$$

$$D_{test} = D_k \quad \dots (7)$$

Metric final is averaged for all folds in equation (7)

$$Metric_{final} = \frac{1}{10} \sum_{k=1}^{10} Metric_k \quad \dots (8)$$

ISSN: 2312-8135 | Print ISSN: 1992-0652 | www.journalofbabylon.com | jub@itnet.uobabylon.edu.iq | info@journalofbabylon.com



### 3.3. Reinforcement learning Algorithm

The DQIDS-CAD reinforcement learning is modeled in algorithm (1). Algorithm (1) shows the main steps for the proposed Deep Q-Network for the coronary artery disease diagnostic system.

**Algorithm (1):** Deep Learning Pseudo-code

| | |
|---|---|
| **Input** | S,A,P,R, γ<br>S: state Space<br>A: Action Space<br>P: Transition Probability<br>R: Reward function<br>γ ∈ [0,1]: *which represent discount factor* |
| **Output** | A = {0,1}, where a=0: No CAD; a=1: CAD |
| 1. | Define the reward function as :<br>$R(s,a) \leftarrow \begin{cases} +2 & Correct\ CAD \\ +1 & Correct\ NO\ CAD \\ -2 & False\ Negative \\ -1 & False\ Positive \end{cases}$ |
| 2. | The optimal action value function is calculated using :<br>$Q^{*(s,a)} = \max_{\pi} E \leftarrow \left[ \sum_{k=0}^{\infty} \gamma^k r_{t+k} \mid s_t = s, a_t = a \right]$ |
| 3. | The input-hidden-output layers are calculated by:-<br>**Input layer**<br>$Hidden\ layer\ h_1 \leftarrow \text{RelU}(W_1 s + b_1)$<br>$Hidden\ layer\ h_2 \leftarrow \text{RelU}(W_{21} s + b_2)$<br>**output layer**<br>output layer $Q(s,a,\theta) \leftarrow W_3 h_2 + b_3)$, $then\ producing\ [Q(s,0), Q(s,1)]$ |
| 4. | Q-Target Computation<br>$y_t = r_t$ with a specific loss function |

### 3.4. Machine Learning Algorithms

The proposed system uses many machine learning algorithms including Support Vector Machine (SVM) developed by Vapnik and his colleagues in 1995 [22]. The description of this algorithm can be found in [23]. Figure (3) shows the hyperplane that separates the training samples.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com



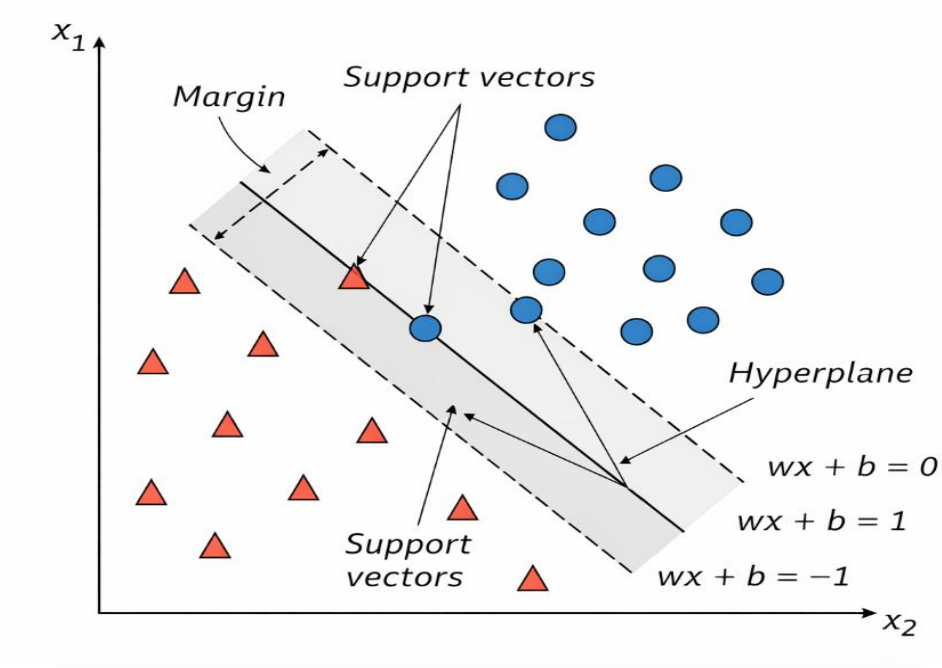


**Figure 3:** Hyperplane that separate training samples [23]

### 3.4.1. Random Forest

The random forest was first proposed by [24]. It is used for classification by using of many independent decision trees (base classifiers). Then, the vote from each of the classifications gives the classification for that example. The complete random forest classification method is shown in figure (4) below.

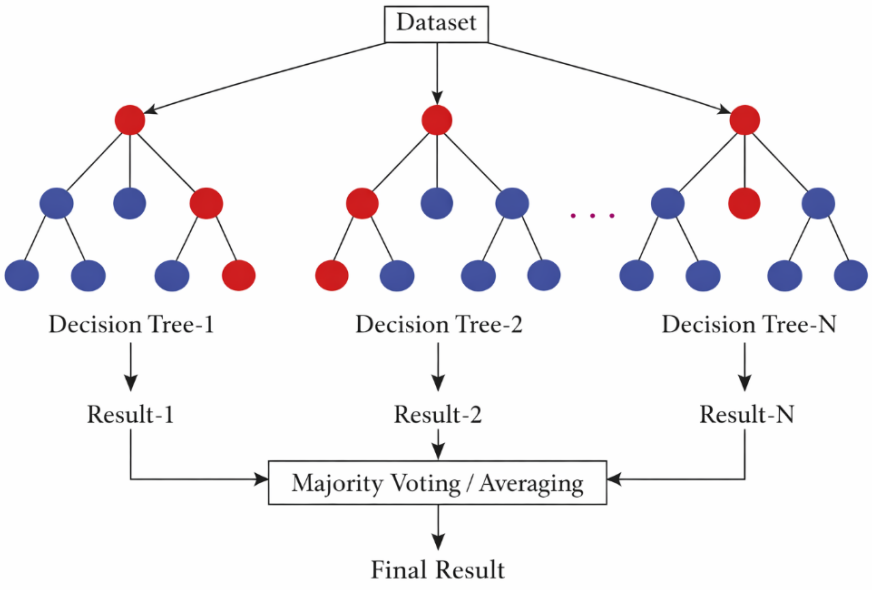


**Figure 4:** The Random Forest and Majority Voting / Averaging [24]

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com



### 3.4.2. Decision Tree

The main scheme o decision tree is hierarchical within an object (decision tree nodes). Each path relate to possible outcomes for each attribute. It consists of three levels: root node, decision node, and leaf node. Decision Tree learning methods are extensively used within the fields of statistics, data mining and machine learning, and healthcare as their predictive abilities provide excellent accuracy results. Figure (5) shows an example of the structure of a decision tree (DT).

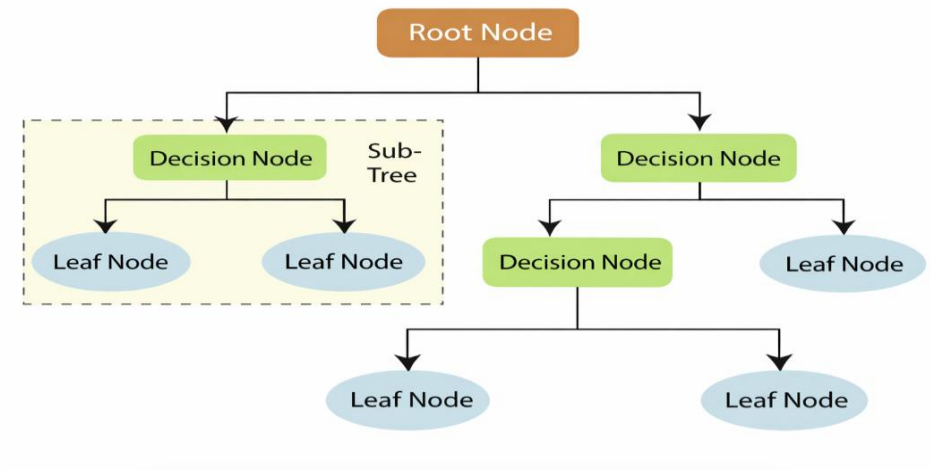


**Figure 5:** The scheme of Decision Trees [25]

Multilayer perceptron (MLP) is a type of artificial neural network made of an input layer, one or more hidden layers, and output layer. In addition, logistic regression used in this work is a linear classification used mainly in binary classification such as CAD/ NCAD diseases. These stacked classifiers can be programmed using the python programming language for many healthcare applications [26].

### 3.4.3. Data Set Description

The proposed work in this study uses five datasets. Class distribution for each dataset is shown in table (1). The preprocessing removed records with missing target labels and the dataset remains modernly balanced.

**Table (1):** Datasets Class distributions for five heart diseases

| Dataset | Class 0 (Non-disease) | Class 1 (Disease) | Total Samples |
|---|---|---|---|
| Cleveland | 164 | 139 | 303 |
| Hungarian | 188 | 106 | 294 |
| Statlog | 150 | 120 | 270 |
| Switzerland | 8 | 115 | 123 |
| VA Long Beach | 51 | 149 | 200 |
| Overall Class Distributions | | | |
| Class 0 (Non-disease) | 561 | | |
| Class 1 (Disease) | 629 | | |
| **Total Dataset Size** | **1190** | | |

ISSN: 2312-8135 | Print ISSN: 1992-0652 | www.journalofbabylon.com | jub@itnet.uobabylon.edu.iq | info@journalofbabylon.com



Table (2) shows the main attributes for the heart diseases used in DQIDS-CAD

**Table 2:** The Cleveland Heart Disease attributes used in DQIDS-CAD

| No. | Feature Name | Type | Description |
|---|---|---|---|
| 1 | **age** | **Integer** | The Age of the patient in (years) |
| 2 | **sex** | **Categorical** | Only two values for Sex (1 = male, 0 = female) |
| 3 | **cp** | **Categorical** | This value represents Chest pain severity type (0–3) |
| 4 | **trestbps** | **Integer** | The Resting blood pressure (mm Hg) |
| 5 | **chol** | **Integer** | The Serum cholesterol (mg/dl) |
| 6 | **fbs** | **Categorical** | Fasting blood sugar > 120 mg/dl (true=1, false=0) |
| 7 | **restecg** | **Categorical** | Resting electrocardiographic results with values (0..2) |
| 8 | **thalach** | **Integer** | To measure the Maximum heart rate achieved |
| 9 | **exang** | **Categorical** | To measure the Exercise-induced angina (yes=1, no=0) |
| 10 | **oldpeak** | **Integer** | ST depression induced by exercise |
| 11 | **slope** | **Categorical** | Slope of peak exercise ST segment (0–2) |
| 12 | **ca** | **Integer** | This value refers to Number of major vessels colored by fluoroscopy (0–3) |
| 13 | **thal** | **Categorical** | It refere to Thalassemia value (3 = normal, 6 = fixed defect, 7 = reversible defect) |
| 14 | **target** | **Integer** | Diagnosis of heart disease (0–4 → later transformed to 0/1) |

ISSN: 2312-8135 | Print ISSN: 1992-0652 | www.journalofbabylon.com | jub@itnet.uobabylon.edu.iq | info@journalofbabylon.com



## 4. RESULTS AND DISCUSSION

The proposed work was implemented on five dataset including the Cleveland, Hungarian, Switzerland, Long Beach VA datasets, and Statlog Heart Diseases with a total of 1190 patient cases. Table (3) shows the main hyperparameters used in machine learning classifiers using the python programming language.

**Table 3:** Machine Learning Hyperparameters for Predicting CAD/ NCAD diseases

| Classifier | Hyperparameters Used | Description |
|---|---|---|
| **Logistic Regression (LR)** | max_iter=5000, class_weight="balanced", random_state=42 | Linear classifier with class balancing to handle dataset imbalance. |
| **Support Vector Machine (SVM)** | kernel="rbf", C=2, gamma=0.2, probability=True, class_weight="balanced", random_state=42 | Non-linear SVM with RBF kernel. Parameters C and gamma selected using GridSearchCV. |
| **Random Forest (RF)** | n_estimators=800, class_weight="balanced_subsample", random_state=42 | Ensemble of decision trees designed to capture nonlinear relationships in the data. |
| **Multi-Layer Perceptron (MLP)** | hidden_layer_sizes=(64,32), activation="relu", solver="adam", alpha=1e-4, learning_rate_init=0.001, max_iter=2000, random_state=42 | Feed-forward neural network with two hidden layers. |
| **XGBoost (XGB)** | n_estimators=900, max_depth=3, learning_rate=0.08, subsample=0.9, colsample_bytree=0.9, scale_pos_weight≈0.94, eval_metric="logloss", random_state=42, n_jobs=-1 | Gradient boosting model optimized using GridSearchCV. |
| **Soft Voting Ensemble** | voting="soft" with models {LR, SVM, RF, MLP, XGB} | Combines probabilistic outputs of all classifiers. |
| **SVM Threshold Optimization** | Same SVM parameters + threshold = 0.53 | Decision threshold optimized using cross-validation on the training set. |

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

The main results for accuracy, precision, recall, and F1-score across deep reinforcement Q-Network, five machine learning models (LR, SVM, RF, MLP, XGB), and voting ensemble is shown in figure (6), figure (7), figure (8) and figure (9) respectively.

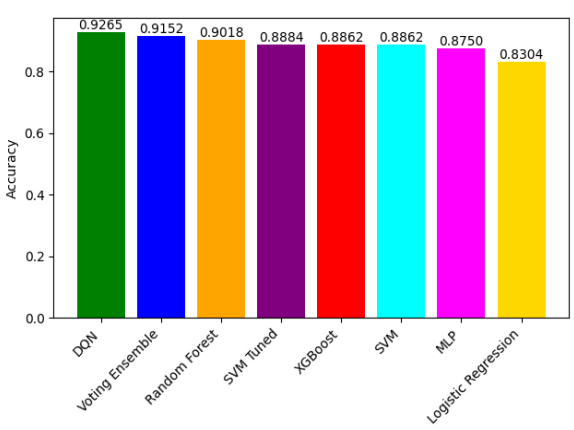


**Figure (6):** Accuracy Comparison Across Models

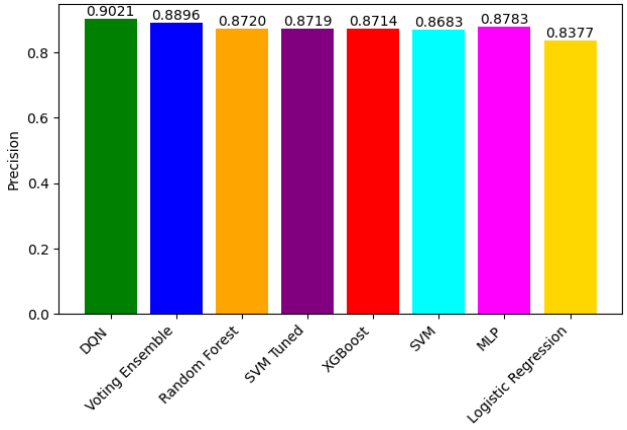


**Figure (7):** Precision Comparison Across Models

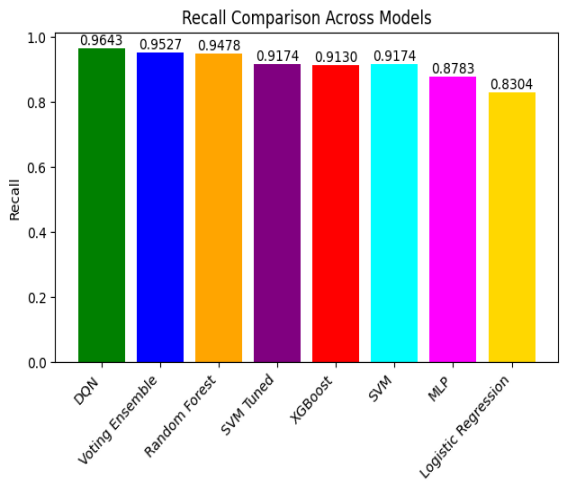


**Figure (8):** Accuracy Comparison Across Models

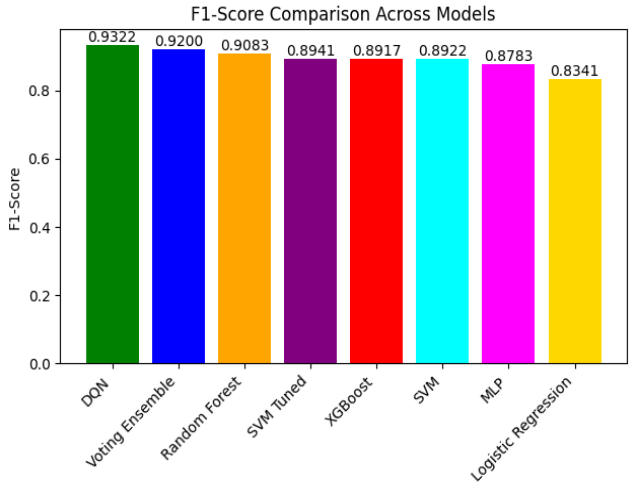


**Figure (9):** Precision Comparison Across Models

The results show the DQN and voting ensemble outperforms in CAD/NCAD detection due to their ability to model complex nonlinear relationships within clinical data. Although the superiority of the DQN model and the group voting mechanism is clearly demonstrated by various performance indicators such as accuracy, fine-tuning, recall, and F1 score, this superiority cannot be considered a random result or a direct reflection of the nature of the data only; rather, it reflects a fundamental

ISSN: 2312-8135 | Print ISSN: 1992-0652 | info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com



difference in the cognitive structure of these models, and in how they deal with structural complexity, uncertainty, and representation learning mechanisms within the five datasets.

The results show that DQN leverages deep reinforcement learning to capture the dynamic patterns, and ensemble voting improves robustness by combining multiple predictive models. Table (4) shows value of all machine learning, ensemble voting, and DQN techniques.

**Table (4):** Shows the main results of all learning models

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| **DQN (Deep Q-Network)** | **0.9265** | **0.9021** | **0.9643** | **0.9322** |
| **Voting Ensemble** | **0.9152** | **0.8896** | **0.9527** | **0.9200** |
| Random Forest | 0.9018 | 0.8720 | 0.9478 | 0.9083 |
| SVM Tuned Threshold (0.53) | 0.8884 | 0.8719 | 0.9174 | 0.8941 |
| XGBoost | 0.8862 | 0.8714 | 0.9130 | 0.8917 |
| SVM | 0.8862 | 0.8683 | 0.9174 | 0.8922 |
| MLP | 0.8750 | 0.8783 | 0.8783 | 0.8783 |
| Logistic Regression | 0.8304 | 0.8377 | 0.8304 | 0.8341 |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| **DQN (Deep Q-Network)** | **0.9265** | **0.9021** | **0.9643** | **0.9322** |
| **Voting Ensemble** | **0.9152** | **0.8896** | **0.9527** | **0.9200** |
| Random Forest | 0.9018 | 0.8720 | 0.9478 | 0.9083 |
| SVM Tuned Threshold (0.53) | 0.8884 | 0.8719 | 0.9174 | 0.8941 |
| XGBoost | 0.8862 | 0.8714 | 0.9130 | 0.8917 |
| SVM | 0.8862 | 0.8683 | 0.9174 | 0.8922 |
| MLP | 0.8750 | 0.8783 | 0.8783 | 0.8783 |
| Logistic Regression | 0.8304 | 0.8377 | 0.8304 | 0.8341 |

## 5. CONCLUSIONS AND RECOMMENDATIONS

This study implemented a combination of five different datasets: the Cleveland, Hungarian, Switzerland, Long Beach VA datasets, and Statlog Heart Diseases and multiple machine learning techniques. The results demonstrate that the advanced approaches, particularly DQN and ensemble learning, outperform other methods Random Forest, SVM Tune, XGBoost, SVM, MLP, and Logistic Regression by effectively capturing complex patterns within CAD/ NCAD clinical data. DQN achieved accuracy with 92.65%, precision 90.21%, highest recall 96.43%, and F1-score 93.22%. The voting ensemble achieved accuracy with 91.52%, precision 88.96%, recall 95.27%, and F1-score 92%. These methods achieved higher accuracy and robustness, indicating their suitability for reliable heart disease classification and supporting their potential application in intelligent medical systems to detect CAD/ NCAD. The future framework should, therefore, be recommended for extension with larger and more diverse clinical datasets on other CAD diseases

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com



plus a real-world clinical environment and resource-constrained device validation of the model to enforce practical applications of the proposed framework.

## Acknowledgment

The authors would like to express our deepest gratitude to the College of Engineering and Information Technology at the University of Babylon and the College of Science at Al-Mustaqbal University for their invaluable support in the publication of this paper.

## Conflict of interests.

There are non-conflicts of interest.

## References

[1] G. A. Roth *et al.*, "Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019," *J. Am. Coll. Cardiol.*, vol. 76, no. 25, pp. 2982–3021, Dec. 2020, doi: 10.1016/j.jacc.2020.11.010.

[2] O. Gaidai, Y. Cao, and S. Loginov, "Global Cardiovascular Diseases Death Rate Prediction," *Curr. Probl. Cardiol.*, vol. 48, no. 5, p. 101622, May 2023, doi: 10.1016/j.cpcardiol.2023.101622.

[3] R. Jagannathan, S. A. Patel, M. K. Ali, and K. M. V. Narayan, "Global Updates on Cardiovascular Disease Mortality Trends and Attribution of Traditional Risk Factors," *Curr. Diab. Rep.*, vol. 19, no. 7, p. 44, Jul. 2019, doi: 10.1007/s11892-019-1161-2.

[4] S. M. D. A. C. Jayatilake and G. U. Ganegoda, "Involvement of Machine Learning Tools in Healthcare Decision Making," *J. Healthc. Eng.*, vol. 2021, pp. 1–20, Jan. 2021, doi: 10.1155/2021/6679512.

[5] Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," *Database*, vol. 2020, p. baaa010, Jan. 2020, doi: 10.1093/database/baaa010.

[6] N. Rana, K. Sharma, and A. Sharma, "Diagnostic Strategies Using AI and ML in Cardiovascular Diseases: Challenges and Future Perspectives," in *Deep Learning and Computer Vision: Models and Biomedical Applications*, U. N. Dulhare and E. H. Houssein, Eds., in Algorithms for Intelligent Systems. , Singapore: Springer Nature Singapore, 2025, pp. 135–165. doi: 10.1007/978-981-96-1285-7_7.

[7] S. Abbas, G. A. Sampedro, S. Alsubai, A. Almadhor, and T. Kim, "An Efficient Stacked Ensemble Model for Heart Disease Detection and Classification," *Comput. Mater. Contin.*, vol. 77, no. 1, pp. 665–680, 2023, doi: 10.32604/cmc.2023.041031.

[8] H. Bihri, L. A. Charaf, S. Azzouzi, and M. E. H. Charaf, "A Robust Stacking-Based Ensemble Model for Predicting Cardiovascular Diseases," *AI*, vol. 6, no. 7, p. 160, Jul. 2025, doi: 10.3390/ai6070160.

[9] P. A. Omosebi, B. Idowu, and A. Folasade, "Classification and Detection of Coronary Artery Disease (CAD) Using Machine Learning Techniques," *NIPES J. Sci. Technol. Res.*, vol. 7, no. 1, Oct. 2025, doi: 10.37933/nipes/7.4.2025.SI22.

[10] A. Ghasemieh, A. Lloyed, P. Bahrami, P. Vajar, and R. Kashef, "A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients," *Decis. Anal. J.*, vol. 7, p. 100242, Jun. 2023, doi: 10.1016/j.dajour.2023.100242.

[11] S. Upadhyay, A. K. Sagar, and N. R. Roy, "Enhancing reliability in coronary artery disease prediction using ensemble learning for key attribute identification," *Life Cycle Reliab. Saf. Eng.*, vol. 14, no. 4, pp. 701–714, Dec. 2025, doi: 10.1007/s41872-025-00318-5.

[12]A. Zarrad, T. Armstrong, and J. Jemai, "A Hybrid Approach to Software Testing Efficiency: Stacked Ensembles and Deep Q-Learning for Test Case Prioritization and Ranking," *Comput. Mater. Contin.*, vol. 0, no. 0, pp. 1–10, 2025, doi: 10.32604/cmc.2025.072768.

[13]M. Gul, H. Ahmad, M. Z. Shafi, M. T. T. Bajwa, M. Ahsaan, and M. A. U. Rehman, "The Role of Reinforcement Learning in Advancing Artificial Intelligence: An Experimental Study with Q-Learning and DQN," *Asian Bull. Big Data Manag.*, vol. 5, no. 3, pp. 122–134, Aug. 2025, doi: 10.62019/zpq9nv97.

[14]M. D. Teja and G. M. Rayalu, "Optimizing heart disease diagnosis with advanced machine learning models: a comparison of predictive performance," *BMC Cardiovasc. Disord.*, vol. 25, no. 1, p. 212, Mar. 2025, doi: 10.1186/s12872-025-04627-6.

[15]X. Liu, S. Su, B. Wang, and X. Zhang, "Prediction of heart disease based on logistic regression and random forest models," *Highlights Sci. Eng. Technol.*, vol. 49, pp. 489–495, 2023.

[16]A. Anderies, J. A. R. W. Tchin, P. H. Putro, Y. P. Darmawan, and A. A. S. Gunawan, "Prediction of Heart Disease UCI Dataset Using Machine Learning Algorithms," *Eng. Math. Comput. Sci. EMACS J.*, vol. 4, no. 3, pp. 87–93, Sep. 2022, doi: 10.21512/emacsjournal.v4i3.8683.

[17]H. El-Sofany, B. Bouallegue, and Y. M. A. El-Latif, "A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method," *Sci. Rep.*, vol. 14, no. 1, p. 23277, Oct. 2024, doi: 10.1038/s41598-024-74656-2.

[18]N. P.L., S. Terence, A. Raj, J. Immaculate, and D. Mishra, "Enhanced Heart disease prediction using LLM ranked feature selection, Dynamic custom Kernel," *Mach. Learn. Appl.*, vol. 23, p. 100860, Mar. 2026, doi: 10.1016/j.mlwa.2026.100860.

[19]M. Hasan, M. A. Sahid, M. P. Uddin, M. A. Marjan, S. Kadry, and J. Kim, "Performance discrepancy mitigation in heart disease prediction for multisensory inter-datasets," *PeerJ Comput. Sci.*, vol. 10, p. e1917, Mar. 2024, doi: 10.7717/peerj-cs.1917.

[20]A. A. Lamir, S. Razzagzadeh, and Z. Rezaei, "A Comprehensive Machine Learning Framework for Heart Disease Prediction: Performance Evaluation and Future Perspectives," May 15, 2025, *arXiv*: arXiv:2505.09969. doi: 10.48550/arXiv.2505.09969.

[21]W. S. Andras Janosi, "Heart Disease." UCI Machine Learning Repository, 1989. doi: 10.24432/C52P4X.

[22]V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY: Springer New York, 1995. doi: 10.1007/978-1-4757-2440-0.

[23]A. A. S. R. De Sousa, J. Da Silva Coelho, M. R. Machado, and M. Dutkiewicz, "Multiclass Supervised Machine Learning Algorithms Applied to Damage and Assessment Using Beam Dynamic Response," *J. Vib. Eng. Technol.*, vol. 11, no. 6, pp. 2709–2731, Sep. 2023, doi: 10.1007/s42417-023-01072-7.

[24]L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[25]I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.

[26]N. K. Basha, S. B. Khan, A. Kumar, and A. Mashat, *Deep Learning and Its Applications Using Python*, 1st ed. Wiley, 2023. doi: 10.1002/9781394167791.

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com



## الخلاصة

**الخلفية**: لا تزال أمراض القلب والأوعية الدموية تشكل هاجسًا صحيًا عالميًا رئيسيًا، مما يُبرز الحاجة المُلحة إلى أدوات تشخيصية موثوقة وقابلة للتطبيق على نطاق واسع. يهدف هذا العمل إلى تقديم نظام متطور للكشف عن أمراض القلب، يدمج البيانات السريرية من أربع مجموعات بيانات دولية رئيسية: كليفلاند، وهنغاريا، وسويسرا، ولونغ بيتش (VA)، بالإضافة إلى قاعدة بيانات Statlog لأمراض القلب سواءً كانت مصحوبة بمرض الشريان التاجي أو غير مصحوبة به. يستخدم النظام 14 سمة معيارية ذات صلة سريرية، ويعتمد على مسار معالجة مسبقة آلي يتضمن استكمال البيانات الرقمية بالوسيط، واستكمال البيانات الفئوية بالبيانات الأكثر تكرارًا، والتحجيم المعياري لضمان سلامة البيانات.

**طرق العمل :** تُنفذ المنهجية المقترحة على مرحلتين. في المرحلة الأولى، يتم التعلم باستخدام شبكة Q العميقة، والتي حققت دقة عالية مقارنةً بالأساليب الأخرى. أما المرحلة الثانية، فهي عبارة عن خوارزمية تصويت التجميع المكدس، والتي يتم تركيبها باستخدام الانحدار اللوجستي، وآلة المتجهات الداعمة (SVM)، و(XGB) XGBoost، والشبكة العصبية متعددة الطبقات، والغابة العشوائية (RF). لقد حسّن النموذج المصمم دقة الأداء باستخدام ضبط المعلمات الفائقة المكثف لخوارزمية GridSearchCV. تُظهر النتائج دقة عالية مقارنةً بالدراسات الأخرى.

**النتائج:** حقق النموذج الأول دقة 92.65%، بينما حقق النموذج الثاني دقة 92.52% باستخدام طريقتي DQN والتجميع على التوالي. أظهر النموذج المقترح أداءً عاليًا وموثوقًا، مما يدل على كونه أداة قوية وموثوقة للكشف المبكر عن أمراض القلب، وخاصةً لتصنيف المرضى إلى مرضى الشريان التاجي (CAD) وغير المصابين به، بالإضافة إلى إمكانية تطبيقه في أنظمة دعم القرار السريري في الواقع العملي.

**الكلمات المفتاحية**:مرض الشريان التاجي، معالجة البيانات المسبقة، التعلم العميق المعزز ، التعلم الآلي، التعلم الجماعي، دعم القرار السريري