



## ENHANCED AI-POWERED CORONARY ARTERY DISEASE PREDICTION USING SPIDER SWARM INTELLIGENCE WITH DEEP HYPER SCALAR CAPSULE GATED RECURRENT NEURAL NETWORK FOR EARLY TREATMENT RECOMMENDATION

D. Rajapriya\*<sup>1</sup> and M. Vigenesh<sup>2</sup>

<sup>1</sup>Assistant Professor, Department of Computer Science and Engineering, Karpagam College of Engineering, Coimbatore, India

<sup>2</sup>Professor, Department of Computer Science and Engineering, Karpagam Academy of Education, Coimbatore, India.

<sup>1</sup><https://orcid.org/0009-0003-6748-0256><sup>id</sup>, <sup>2</sup><https://orcid.org/0009-0001-8370-0277><sup>id</sup>

Email: \*[rajapriya546228@gmail.com](mailto:rajapriya546228@gmail.com), [mvigenesh@gmail.com](mailto:mvigenesh@gmail.com)

### ARTICLE INFO

#### Article History

Received: February 11, 2026

Reviewed: April 18, 2026

Accepted: May 18, 2026

Published: June 30, 2026

#### Keywords:

Prediction;  
Deep Learning;  
Machine Learning;  
Normalization;  
Neural Network.

### ABSTRACT

Nowadays, increasing numbers of heart attacks are commonly suffered by young age people due to unhealthy lifestyles. Especially Cardiovascular Disease (CVD) or coronary artery disease (CAD) are the most dangerous, deadliest diseases. Without knowing their symptoms, heart block attacks and lead to death. Unhealthy lifestyles are an essential aspect of this disease, which is analyzed by healthcare sectors through behavior and symptoms to identify with support of CT or angiogram support. So many traditional methods are implemented in Machine Learning (ML) and Deep Learning (DL) methods to identify heart disease by analyzing the disease's relational features and behaviors. Most of the techniques concentrate on disease margins to select non-relational features. So, the analysis degrades the accuracy of the performance and increases the false positive rates. Consequently, the precision, recall, and f1 attain poor performance. To resolve this problem, we proposed an advanced Artificial Intelligence-Powered Deep Hyper Spectral Learning Technique for heart disease prediction using Optimal Spider Swarm Intelligence Technique (OSSIT) Feature Selection with Deep Hyper Scalar Capsule Gated Recurrent Neural Network (DHSC-GRNN) for early prediction and recommendation. Initially, the Min Max normalization is applied to normalize the CAD heart disease dataset, and the Cardiac Feature Behavioral Impact Rate (CFBIR) is analyzed through decision tree ideal margins. By reducing the non-relational features and introducing the OSSIT, it is applied to select the mutual dependencies of feature patterns. Finally, the DHSC-GRNN is used to predict heart disease by class by risk. The feature patterns are scaled into capsule units from a related class cluster and fed into recurrent gated units to reduce non-scaling margins and disease severity resemblance to predict the risk. The proposed system achieves high performance compared to the other system in terms of higher precision rate and recall rate f1 measure by increasing the accurate favorable rates compared to the preliminary traditional methods.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

### I. INTRODUCTION

In recent years, the alarming rise in CVD, particularly CAD, has become a pressing public health concern. This trend is particularly evident among younger populations, where unhealthy lifestyles contribute significantly to the incidence of heart attacks. The consequences of these diseases can be dire, often leading to severe health complications or even death if not detected early. Traditional methods of diagnosing heart disease typically involve the use of imaging techniques such as CT scans or angiograms, coupled with an analysis of symptoms and behavioral factors.

However, these methods often fall short in terms of predictive accuracy, resulting in high false positive rates and inadequate early intervention. This essay proposes a novel system that utilizes advanced artificial intelligence (AI) techniques to enhance the prediction and early detection of heart disease, thereby improving patient outcomes.

## 1.1 BACKGROUND AND MOTIVATION

The increasing prevalence of unhealthy lifestyles, characterized by poor dietary choices, sedentary behavior, and high stress levels, has been linked to a surge in heart disease cases. In particular, the lack of awareness regarding the symptoms of heart disease exacerbates the situation, as individuals may ignore warning signs until it is too late. Traditional diagnostic methods, while effective to some extent, often rely on non-relational features that do not capture the complex interplay of factors contributing to heart disease. Consequently, there is a pressing need for innovative approaches that leverage ML and DL techniques to analyze relational features and behaviors associated with heart disease more effectively.

## II. LITERATURE SURVEY

Medical science is offered predictive models created with ML algorithms, allowing for timely identification of CAD and more helpful treatment results [1]. Furthermore, this proposed model is introduced as a possible open-source solution, serving as a valuable clinical tool for predicting CAD. The author [2] introduced ResNet-50, a noninvasive technique for early CAD diagnosis. It analyzes tongue image features through ML. Timely and precise diagnosis of CAD in hypertension patients is crucial for minimizing the risks and detrimental effects related to hypertension and CAD. Although invasive techniques are needed, this model can diagnose CAD using clinical data. The paper [3] uses the Z-Alizadeh Sani CAD dataset to support its suggestion that ML methods be used to identify CAD early. Diagnosing CAD involves significant time and resource commitments from competent professionals. This paper [4] proposed a data mining technique for CAD prediction using the Isfahan Cohort Study dataset and the Support Vector Machine (SVM) method. Nonetheless, the techniques for analyzing and managing this condition often have costly side effects. This study [5] described a Two-Layer Voting (TLV) model, a technique that combines hard and soft voting. Using the heart disease dataset from UCI and the CVD dataset from Kaggle, the proposed TLV technique and smooth voting demonstrate good accuracy. Appropriate analysis is crucial when discussing the new model based on the Two-Class Boost Decision Tree (two-class-BDT) approach. This study [6] has demonstrated the extreme potential of highly sophisticated ML models in CAD prediction and management. Develop prototype systems that employ different ML algorithms and assess their performance metrics to identify the optimal model. This study employed Artificial Neural Networks (ANN) to analyze the features of clinical datasets [7].

The Cleveland heart disease dataset from the UCI ML repository and Deep Neural Network (DNN) methodologies based on computational insights from different coronary heart disease forecasts were used to evaluate each method's performance metrics [8]. This study [9] aimed to detect early CAD using clinical and laboratory testing and ML algorithms. Artificial intelligence models perform better when high-quality virtual patient data is generated using Gradient Boosting (GB) techniques. This paper [10] described a dataset that combines five publicly available CAD-related datasets. Using the open-source framework AutoGluon, an ensemble model was created to evaluate AutoML's potential in developing disease prediction models for cardiovascular medicine. This descriptive study analyzed 303 records with 26 characteristics, all chosen as features of interest based on expert recommendations. Furthermore, a key aspect of the K-Nearest Neighbor (KNN) algorithm is its application in developing CAD detection models [11]. Using the proposed random forest (RF) approach, a model was created to predict CAD, and its performance was evaluated using the ROC curve. This paradigm provided significant clinical advantages and helped with primary prevention and care of individuals with coronary heart disease [12]. Various Convolutional Neural Network (CNN)-based ensemble voting classifier models were developed and examined [13]. The main goal of this study is to present an improved approach for enhancing the accuracy of CAD predictions within a classification framework using DL algorithms. Several ensemble voting classifier models based on CNNs were created and analyzed.

This study [14] aimed to introduce an enhanced method for increasing the accuracy of Computer-Aided Diagnosis (CAD) predictions within a classification framework utilizing DL algorithms. This paper [15] explored various elements of heart disease using ML models, including the Naive Bayes (NB) algorithm. However, in recent decades, heart disease has emerged as a significant cause of mortality. Predictive models are constructed using ML algorithms, such as SVM and KNN algorithms, and assessed using Receiver Operating Curve (ROC) performance evaluation approaches and accuracy [16]. This study [17] introduced the Classification and Regression Tree (CART) algorithm as a simpler, more reliable alternative to existing diagnostic methods for CAD. In contrast, these traditional approaches often incur costs and health risks. Moreover, the Least Absolute Summation Selection Operator (LASSO) assesses the importance of the features. The author [18] discussed a majority voting approach to pinpoint the primary features. Additionally, they introduce an epoch training technique similar to simulated annealing to improve classification accuracy. The suggested methodology is a hybrid binary objective Particle Swarm Optimization (PSO) algorithm that integrates many techniques for numerical particle classification and encoding and velocity calculations [19]. Randomly created particles that take random values within the designated limitation attribute are the source of the rules. Precision metrics assessed the system's performance and were implemented using the publicly available UCI Cardiology dataset [20]. These results, showing 93.6% accuracy, highlight the system's potential as a crucial tool for early detection and prevention, which could enhance clinical outcomes in treating CAD.

Table 1: Survey of ML and DL methods for CVD prediction.

| Author/Year | Used Methodology                                      | Used Dataset                    | Limitations                                                                                                                                                                          | Accuracy |
|-------------|-------------------------------------------------------|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|
| [21]        | Self-Supervised Learning (SSL)                        | Longitudinal laboratory data    | SSL identified difficulties with longitudinal laboratory data, including as irregularity, temporality, absenteeism, and sparsity.                                                    | 0.90     |
| [22]        | Light Gradient Boosting Machine (LightGBM)            | Publicly available CVD dataset  | It may not include all patient demographics or unusual conditions. The model's performance on external datasets or in clinical situations was not addressed.                         | 94.5%    |
| [23]        | Ensemble method                                       | Clinical data                   | It may have constraints, such as potential overfitting from the ensemble technique and the necessity for big, varied datasets to assure model generalizability.                      | 88%      |
| [24]        | Random Forest and SVM (RF-SVM)                        | Electrocardiogram (ECG)         | RF-SVM may be restricted by ECG data quality and variability, possible class imbalances, and the necessity for validation on a wide range of patient groups.                         | 93.7%    |
| [25]        | Heartbeat Dynamics                                    | MIT-BIH Arrhythmia Database     | Potential drawbacks include the necessity for significant computing resources for feature extraction and validation on external datasets to ensure generalizability.                 | 95.2%    |
| [26]        | Neuro-fuzzy method                                    | Clinical datasets               | The suggested technique may be constrained by the quality and completeness of the clinical data, potential biases in patient selection, and the requirement for external validation. | 91%      |
| [27]        | DL methods                                            | ECG data                        | Using these methods, we face a number of issues, including dealing with imbalanced data, potential overfitting, and the requirement for validation on bigger, more diverse datasets. | 89%      |
| [28]        | Ensemble method-based Multilayer Dynamic System (MDS) | Patient data                    | The proposed method may have limitations such as the complexity of the ensemble model, potential overfitting, and the requirement for big datasets to assure robustness.             | 92%      |
| [29]        | Supervised ML (SVL)                                   | Clinical datasets               | The SVL was trained and tested on a specific dataset, which may limit its applicability to other populations or settings.                                                            | 87.4%    |
| [30]        | RF                                                    | Cleveland Heart Disease dataset | The RF focused on a single dataset, which may limit the generalizability of the findings.                                                                                            | 86.6%    |

Source: Authors, (2026).

Table 1 presents a comprehensive survey of various machine learning (ML) and deep learning (DL) methodologies that have been utilized to predict cardiovascular disease (CVD).

### III. PROPOSED METHODOLOGY

The proposed system centres on an advanced Artificial Intelligence-Powered Deep Hyper Spectral Learning Technique for heart disease prediction. This system integrates the OSSIT for feature selection and the DHSC-GRNN for predictive modelling. The following sections outline the key components of the proposed system and its operational framework.

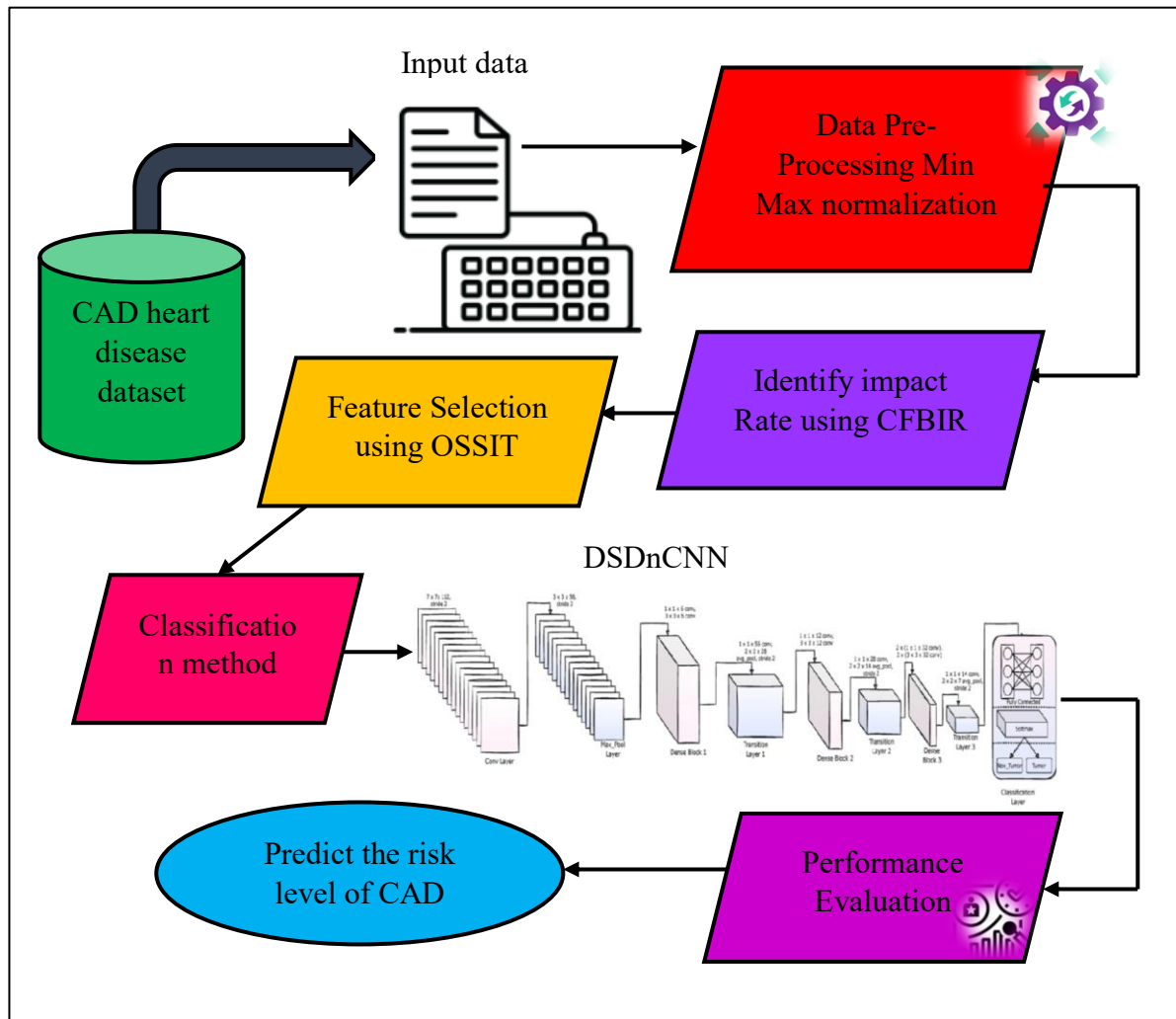


Figure 1: The Proposed Method Architecture Diagram.

Source: Authors, (2026).

The proposed architecture for heart disease prediction, illustrated in Figure 1, employs an advanced AI-powered deep Hyper Spectral Learning Technique. It utilizes OSSIT for feature selection and DHSC-GRNN to enhance predictive accuracy. The CAD heart disease dataset is first normalized using Min-Max scaling, followed by CFBIR analysis through decision tree ideal margins. OSSIT effectively eliminates non-relational features, enabling the identification of mutual dependencies within feature patterns. The DHSC-GRNN then processes these patterns into capsule units provided into recurrent gated units to reduce non-scaling margins and accurately predict disease risk.

#### III.1 DATA PREPROCESSING

The initial step in the proposed system involves data preprocessing to ensure the quality and consistency of the dataset used for training the predictive model. The Min-Max normalization technique is employed to scale the CAD heart disease dataset, ensuring that all features are within a specified range. This normalization process is crucial for the subsequent analysis, as it helps mitigate the influence of outliers and enhances the model's performance. The preprocessing method of Min-Max Normalization applies to CVD datasets so they can scale numerical attributes between specified ranges, which commonly include  $[0,1]$  or  $[-1,1]$ . Computing the new feature values requires subtracting the minimum point followed by the range (max-min) division for each feature so data distribution remains intact. Normalizing the data range makes the model training process more stable, and features with larger magnitudes will not overshadow features with smaller magnitudes. CVD prediction models that normalize medical attributes can improve blood pressure, cholesterol levels, and heart rate quality. The normalization process is most valuable for CVD datasets because it scales features, including blood pressure, cholesterol levels, BMI, and heart rate, which have different units. The normalization process makes all features have an equal impact on model training while improving the overall computational speed. In equation 1 we perform general formula of Min-Max Normalization,

$$I' = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \quad (1)$$

Let assume,  $I$  as original value (a patient's systolic blood pressure),  $I_{\min}$  as minimum value of the feature in CVD dataset (lowest systolic blood pressure recorded),  $I_{\max}$  as maximum value of feature in CVD dataset (highest systolic blood pressure recorded), and  $I'$  as normalized value in the range  $[0,1]$ . The equation performs a rescaling operation on all systolic blood pressure readings to create range values from 0 through 1 which helps maintain fairness when comparing patients. The range of normalization can differ from original features to  $[-1,1]$  when we aim to achieve enhanced numerical stability. In equation 2 we perform modified Min-Max for a custom range  $[u, v]$ ,

$$I' = u + \frac{(I - I_{\min})(v - u)}{I_{\max} - I_{\min}} \quad (2)$$

here, we use  $u, v$  as the desired minimum and maximum values of the normalized range. In this equation the cholesterol level is transformed to 0.2, ensuring that it remains within the new range while preserving its relative magnitude. The normalization process transforms individual features independently such as  $BMI' = \frac{BMI - BMI_{\min}}{BMI_{\max} - BMI_{\min}}$  by adjusting the BMI value. This method operates on body mass index separately when combined with other measurements including Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Cholesterol Levels (Chol), and Body Mass Index (BMI). Independently applying transformations to features ensures the numerical attributes maintain uniform scaling that allows min-max normalization to work without influence from original unit differences. The normalization process that utilizes minimum and maximum values shows sensitivity to outliers within the dataset. Extremely high or low values in a dataset can substantially change the normalization level. The application of Min-Max Normalization follows percentile thresholding to remove outliers before normalization occurs,

$$I' = \frac{I - I_{q5}}{I_{q95} - I_{q5}} \quad (3)$$

Let assume,  $I_{q5}$  as 5th percentile of the feature (to exclude extreme low values), and  $I_{q95}$  as 95th percentile of the feature (to exclude extreme high values). When DBP contains rare high measurements, such as 250 mmHg, from measurement errors, the transformation using the 5th and 95th percentiles provide better strength. Separate Min-Max Normalization should be used for demographic groups like age groups because this preserves the relationship between values within each grouping structure

$$I'_g = \frac{I_g - I_{\min,g}}{I_{\max,g} - I_{\min,g}} \quad (4)$$

where normalization is applied within each subgroup. Suppose we have two groups: Young Patients (18-40 years): Blood pressure ranges from 95 to 150 mmHg, Elderly Patients (60+ years): Blood pressure ranges from 120 to 190 mmHg. This equation separately ensures fair comparisons within each demographic group instead of forcing all patients into a single distribution.

### III.2 CARDIAC FEATURE BEHAVIORAL IMPACT RATE (CFBIR)

To improve predictive accuracy, the CFBIR is analyzed using decision tree ideal margins. This analysis identifies the most relevant features that contribute to heart disease risk. However, traditional feature selection methods often overlook the interdependencies among features, leading to suboptimal model performance. The Sequential Pattern Algorithm (SPA) is vital for discovering repeated medical event sequences in the CVD dataset. SPA reveals standard patterns of disease development together with treatment response, as well as time-dependent patterns of risk elements. A sequential pattern algorithm within CVD dataset records would reveal the frequent sequence, including Hypertension followed by High Cholesterol and progressing to Atrial Fibrillation, then Heart Failure. Evidence shows hypertension patients tend to face elevated cholesterol levels that might result in atrial fibrillation before the onset of heart failure. Healthcare professionals need these insights to intervene early in implementing prevention programs that enhance patient outcomes. Medical researchers and healthcare professionals can use CVD dataset analysis with SPA to discover major disease evolution patterns and develop superior beneficial plans and model-based cardiovascular health assessments. The chronological order of patient events forms an ordered list and is a sequence  $Q$ . In equation 5 we perform the  $Q$ ,

$$Q = \{c_1, c_2, c_3, \dots, c_n\} \quad (5)$$

were, each event  $c_x$  represents a medical condition, symptom, diagnosis, or treatment observed in a patient's medical history. The equation analyzes a patient's case by tracking their diagnosed medical conditions, beginning with hypertension on Time: 2020-01-10 and high cholesterol on Time: 2021-03-15, among other events. A progressive change in CVD emerges through such a  $Q$  value. The support  $S$  of the  $Q$  in a dataset is the proportion of patients who exhibit that sequence. In equation 6 we computed the proportion of patients,

$$S = \frac{Q_N}{I_T} \quad (6)$$

Let assume,  $Q_N$  as number of patients with  $Q$ , and  $I_T$  as Total number of patients in dataset  $I$ . This means that 25% of patients in the dataset follow this disease progression pattern, making it a significant sequential pattern for cardiovascular risk assessment. In equation 7, we perform confidence measure  $C$  for sequential patterns, which measures the probability that a specific event follows a given sequence,

$$C(Q_1 \Rightarrow Q_2) = \frac{S(Q_1 \cup Q_2)}{S(Q_1)} \quad (7)$$

Let assume,  $Q_1$  as a partial sequence (e.g., "Hypertension, High Cholesterol"),  $Q_2$  as the predicted next event (e.g., "Arrhythmia"), and  $Q_1 \cup Q_2$  as the full sequence (e.g., "Hypertension, High Cholesterol, Arrhythmia"). The statistical confidence stands at 75%, with 40% of patients having Hypertension and High Cholesterol and 30% following Hypertension and High Cholesterol, which leads to Arrhythmia. The analysis shows that Arrhythmia develops in 75% of patients with Hypertension, followed by High Cholesterol, thus indicating strong confidence in the sequence pattern. By following we perform Prefix-Projected Sequential Pattern (PrefixSpan) through equation 8,

$$Q = \set{1} \quad (8)$$

The widespread application of this equation occurs in sequential pattern mining operations within big medical databases. The algorithm performs pattern growth through prefix-based database projection. Through pattern  $Q = (\text{Hypertension, High Cholesterol})$ , the database shows all records beginning with these conditions and ending with Arrhythmia, Heart Failure, or Stroke. The PrefixSpan finds high-support sequential patterns efficiently by expanding its frequent prefixes through its iterative process. CVD requires researchers to pay attention to the duration between different medical events. PrefixSpan uses time constraints to identify meaningful clinical patterns from the patient data. It is defined as,

$$Z_{min} \leq (Z_{c_x} - Z_{e_{x-1}}) \leq Z_{max} \quad (9)$$

Let assume,  $Z_{c_x}$  as timestamp of current medical event,  $Z_{e_{x-1}}$  timestamp of previous event,  $Z_{min}$  as minimum allowable time gaps among event, and  $Z_{max}$  as maximum allowable time gaps among event. For the time gap window between 6 months and 5 years, only the "Hypertension, High Cholesterol, Arrhythmia, Heart Failure" pattern is valid when all transitions happen within this duration. The calculated equation avoids the detection of patterns when medical significance exists above this specified timeframe. The extracted frequent sequences become ready for future disease risk prediction models. An application of Markov Chain modeling allows researchers to measure the likelihood that patients transition from disease stage to disease stage using,

$$G(c_{x+1}|c_x) = \frac{C(c_x \Rightarrow c_{x+1})}{C(c_x)} \quad (10)$$

Let assume,  $G$  as probability,  $c_{x+1}$  next event,  $c_x$  as current event,  $C(c_x \Rightarrow c_{x+1})$  as the number of patients who progressed from  $c_x$  to  $c_{x+1}$ , and  $C(c_x)$  as the total number of patients with  $c_x$ . This equation is used to assess the likelihood of future CVD complications based on current conditions. By using the SPA we can enhance CVD prediction models, improve clinical decision-making, and ultimately reduce CVD morbidity and mortality.

### III.3 OPTIMAL SPIDER SWARM INTELLIGENCE TECHNIQUE (OSSIT)

To address this limitation, the OSSIT is introduced. This technique employs a swarm intelligence approach to select mutual dependencies among feature patterns, effectively reducing the number of non-relational features while retaining those that are critical for accurate prediction. The Social Spider Optimization (SSO) technique model's spider cooperative behaviours found in nature to find food and survive. SSO selects the most pertinent features to enhance ML model performance for network intrusion detection with UNSW-NB15 datasets. The algorithm turns individual spiders into potential answers by showing possible feature sets. The algorithm lets spiders move between webs using three movement types related to how real-life spiders interact. Each spider's quality stands against an established measure that analyses performance results and resource usage during evaluation. The spiders move towards an effective answer during multiple rounds by removing useless feature elements from the feature set. At UNSW-NB15, SSO helps address excessive feature dimensions to detect cyber-attacks more effectively while working faster. Through SSO, spiders show potential feature subsets  $Q_x = [q_{x1}, q_{x2}, \dots, q_{xt}]$   $Q$  as a group of solutions represented by  $x$  with  $t$  as the total features. The selection of features follows a binary process denoted by  $q_{xy} \in \{0,1\}$ . A zero value represents non-selection, while a one shows chosen features. The dataset UNSW-NB15 contains 49 features for which the value  $t=49$  applies. Encoding determines which features (e.g., src\_ip, dst\_port) are part of the reduced set. In equation 11 we perform fitness function  $F$  to evaluate the fitness value of the  $Q$ ,

$$F(Q_x) = \alpha \cdot A(Q_x) - \beta \cdot \frac{|Q_x|}{t} \quad (11)$$

Let assume,  $A(Q_x)$  as classification accuracy of selected features,  $|Q_x|$  as number of selected features, and  $\alpha, \beta$  as weighting factors which is used to balance feature reduction and accuracy. For UNSW-NB15,  $A(Q_x)$  measures how well the reduced feature set classifies normal vs. attack traffic. Then in the SSO method the male spiders moved on the basis of attraction or repulsion so in the equation 12 we compute the movement of male spider  $S$ ,

$$S_x = Q_x + \eta \cdot (Q_{g_{best}} - Q_x) \quad (12)$$

Let assume,  $\eta$  as random number which is used to control exploration, and  $Q_g$  as position of the global best spider which is used to find the best solution in the UNSW-NB15 dataset. In this equation the movement toward  $Q_g$  ensures that the SSO explores feature subsets that better separate normal and attack classes in UNSW-NB15. By following we compute the movement of female spider  $D$  through equation 13 because  $D$  communicate via vibrations to refine their positions,

$$D_x = Q_x + \gamma \cdot (Q_{n_{best}} - Q_x) \quad (13)$$

Let assume,  $\gamma$  as random number which is used to influencing local search, and  $Q_{n_{best}}$  as best neighbor solution on the basis of fitness.

This equation helps female spiders to identify feature subsets with improved fitness, focusing on critical features such as packet size, protocol, or attack labels. By the equation 4 we share the social knowledge through equation 14 to balance exploration and exploitation,

$$Q_x^{i+1} = Q_x^i + \delta \cdot K_x \quad (14)$$

Let assume,  $i$  as current iteration or time step of the SSO method,  $\delta$  as control parameter to knowledge integration, and  $K_x$  as combination of local and global information. In the UNSW-NB15 dataset, this equation ensures that relevant features like service, sload, or ct\_state\_ttl are consistently explored by the population. Then we perform stopping criterion function of the SSO method through equation 15,

$$\text{Stop if: } i \geq I_{max} \text{ or } F(Q_{best}) \geq F_T \quad (15)$$

Let assume  $i$  as current iteration,  $I_{max}$ , maximum number of iterations,  $F$  as fitness value,  $Q_{best}$  as best solution of spider, and  $T$  as threshold which is used to stop the SSO method. With UNSW-NB15, the equation becomes the target accuracy score or maximum number of iterations as desired milestones. Through UNSW-NB15, the SSO algorithm systematically improves feature subsets using social behaviors and direct knowledge transfers to blend performance enhancement with feature selection. This approach maintains top-quality security performance while minimizing platform demands across real-world network protection scenarios.

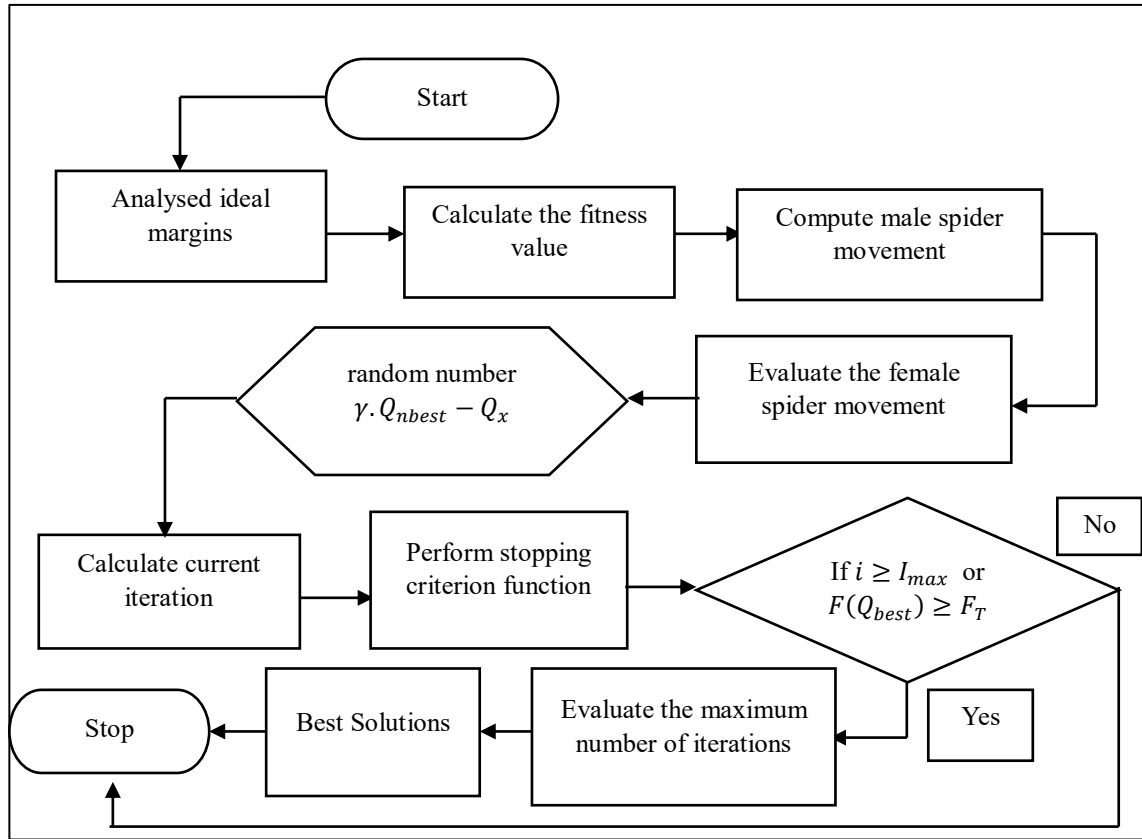


Figure 2: The Flowchart Diagram based on OSSIT.

Source: Authors, (2026).

Figure 2 in the Flowchart Diagram depicts the OSSIT method, providing strong security performance while minimizing platform requirements in practical network protection situations.

#### III.4 DEEP HYPER SCALAR CAPSULE GATED RECURRENT NEURAL NETWORK

Once the relevant features have been identified, the DHSC-GRNN is employed as the core predictive model. This innovative architecture combines the strengths of capsule networks and gated recurrent neural networks (GRNNs) to enhance the model's ability to capture complex patterns in the data. The feature patterns are scaled into capsule units, which represent related class clusters. These capsule units are then fed into recurrent gated units, allowing the model to learn temporal dependencies and variations in the data. This approach significantly reduces non-scaling margins and improves the model's ability to predict disease severity and risk accurately. LSTM is a unique Recurrent Neural Network (RNN) built to solve the gradient disappearance issue in traditional RNNs when processing sequential data. Using an LSTM network for UNSW-NB15 illustrates how traffic patterns evolve to find essential features that reveal normal and aberrant activities. Using LSTMs enables the organization of input features by letting their logs capture and gate mechanisms focus on what is necessary and ignore optional elements. Because of this, LSTMs show their effectiveness in finding crucial features that strengthen the performance of security system anomaly detectors. Through equation 16, our method performs the forget gate to determine which prior cell state  $P_{(t-1)}$  information needs to be discarded.

$$f_t = \sigma(U_f \cdot [h_{t-1}, a_t] + b_f) \quad (16)$$

Let assume,  $f_t$  as forget gate,  $U_f$  as weight of the forget gate,  $h_t$  as hidden gate,  $a_t$  as input gate, and  $b_f$  as biased gate. The forget gate in the UNSW-NB15 dataset can decide whether or not features like "service," "protocol," or "srcip" from earlier time steps (or records) should be eliminated or remain useful for identifying abnormalities. For instance, the forget gate deactivates its weight if the "srcip" pattern isn't useful for anomaly identification. Then, using equation 17, we execute the input gate to determine which new data should be stored in the cell state  $\tilde{Q}_t$ .

$$\begin{aligned} i_t &= \sigma(U_i \cdot [h_{t-1}, a_t] + b_i) \\ \tilde{Q}_t &= \tanh(U_c \cdot [h_{t-1}, a_t] + b_q) \end{aligned} \quad (17)$$

Let assume,  $i_t$  as input gate, and  $\tilde{Q}_t$  as candidate values which is used to update the state. The above equation can determine which attributes (such as "attack\_cat" or "sreport") are relevant for anomaly identification and how much weight to give them while refreshing the memory. For instance, it gives a new feature combination more weight if it implies malevolent behavior. By following we update the  $Q_t$  by combining information retained from the  $f_t$  and  $i_t$  through equation 18,

$$Q_t = f_t \cdot Q_{t-1} + i_t \cdot \tilde{Q}_t \quad (18)$$

Due to this equation, the LSTM can combine information from previously added and freshly introduced feature values. For example, it updates its knowledge of anomalous traffic patterns by combining recent and historical observations of "ct\_dst\_ltm" (count of destination connections). The perform output gate  $o_t$  through equation 19 which is used to determine what part of the  $Q_t$  will influence the current  $h_t$  which is passed to the next step or used for predictions.

$$\begin{aligned} o_t &= \sigma(U_o \cdot [h_{t-1}, a_t] + b_o) \\ h_t &= o_t \cdot \tanh(Q_t) \end{aligned} \quad (19)$$

The equation identifies vital patterns in time-based attributes like sttl (source-to-destination time-to-live) and dttl (destination-to-source time-to-live). It forwards the information to detect abnormalities while discarding unnecessary background data. An LSTM uses these formulas to locate time-based patterns such as series changes in(ct\_srv\_dst) items and repeated attacks in(attack\_cat) categories. This methodology allows LSTM models to identify essential features for spotting dangerous behaviors and create more reliable anomaly detection solutions.

### III.5 GENERATIVE ADVERSARIAL NETWORKS (GAN) METHOD

Generative Adversarial Networks (GANs) contain a framework of two models, a generator and a discriminator, that compete against one another. GAN-based technology helps us find essential features in the UNSW-NB15 dataset that best predict network behavior. The generator builds realistic traffic data by understanding dataset patterns, and the discriminator tests if inputs match real networks or generator outputs. The discriminator uses this competitive training method to understand specific network features (protocol, flow duration, and packet size) for detecting real fake data samples. GANs evolve discriminator and generator repeatedly to find key network features that detect differences between benign and malicious traffic from the UNSW-NB15 dataset. In this method in equation 20 we perform generator objective  $G$ ,

$$v \sim q_v(v), \quad G(v; \theta_G) \rightarrow i_s \quad (20)$$

Let assumes,  $v$  as noise vector,  $q_v(v)$  as prior distribution,  $i$  as input data,  $i_s$  as synthetic data, and  $\theta_G$  as parameter of the generator. The generator takes a random  $v$ , sampled from  $aq_v(v)$ , and transforms it into synthetic data  $i_s$ , which mimics the real data distribution. In this equation the  $G$  transforms into synthetic network traffic data, including features such as "srcip" (source IP address), "proto" (protocol type), or "ct\_dst\_sport\_ltm" (count of destination ports). The goal is to generate realistic traffic patterns that resemble actual data, capturing both benign and anomalous behaviours. By following we perform discriminator objective  $D$  through equation 21,

$$D(i; \theta_D) \rightarrow [0,1] \quad (21)$$

The discriminator takes an  $i$ , which can either be real data from the dataset or  $i_s$  generated by  $G$ . It outputs a probability value indicating whether the data is real (close to 1) or fake (close to 0). In this equation  $i$  includes real network traffic records with features like "ct\_srv\_dst" (connections per service), "sbytes" (source bytes), or "attack\_cat" (attack category). The discriminator evaluates whether the input data aligns with the characteristics of real network traffic or is generated by  $G$ . After we perform the adversarial loss function through equation 22,

$$\min_G \max_D \mathbb{E}_{i \sim p_{data}}(i) [\log D(i)] + \mathbb{E}_{v \sim q_v}(v) [\log (1 - D(G(v)))] \quad (22)$$

This is the core adversarial loss function. The  $D$  maximizes the probability of correctly classifying real and fake data, while the  $G$  minimizes it by trying to fool  $D$ . The two networks play a minimax game. In this equation the  $D$  attempts to distinguish between real traffic records (e.g., actual values of "dstip," "sttl," or "dttl") and fake traffic generated by  $G$ . The generator builds successive versions of human-like data while keeping the essential patterns required to spot everyday behavior from atypical activity intact. Noisy signal initializes the generator to create synthetic versions of the entire UNSW-NB15 dataset entries, including "proto," "ct\_flw\_http\_mthd," and "dbytes" information.

The discriminator detects which traffic features "service," "state," and "ct\_src\_ltm" best differentiate real from synthetic samples. The feedback helps refine feature weights. Training with additional features causes reductions when these features have minimal effects on identifying anomalies. This method reveals which metrics make the most significant difference in finding anomalies. The selected features from the refined dataset help improve intrusion detection through effective classification methods. GAN networks pinpoint relevant behavioural distinctions through their adversarial loss performance improvement. The technique identifies important dataset features that enhance model performance when used to detect intrusions on the UNSW-NB15 dataset.

#### IV. PERFORMANCE EVALUATION

The effectiveness of the proposed system is evaluated through a series of performance metrics, including precision, recall, and F1 score. By comparing the results of the DHSC-GRNN with those obtained from traditional methods, it is anticipated that the proposed system will demonstrate superior performance in terms of higher precision rates, improved recall rates, and enhanced F1 measures. The focus on relational features and the incorporation of advanced AI techniques are expected to lead to a substantial increase in accurate favorable rates, thereby facilitating early intervention and improving patient outcomes. In this section, evaluation metrics can be used to analyze the effectiveness of classification systems for heart disease. In addition, the proposed method can provide statistical evaluation analysis by computing the heart disease. The features of the heart disease dataset can be evaluated by comparing the proposed DHSC-GRNN method with ANN, LightGBM and DNN methods from the literature. The system's efficacy is validated through various evaluation methods:

- **Cross-Validation:** Implementing k-fold cross-validation ensures that the model is robust and can generalize well to unseen data.
- **Comparison with Existing Models:** The new system's performance is benchmarked against traditional stroke prediction models to quantify improvements in prediction accuracy and reliability.

After that, the performance of the feature selection technique can be determined using the elements of the confusion matrix. Moreover, heart disease data can be analyzed using various evaluation metrics such as precision, false score, precision and specificity. Furthermore, these features can be analyzed using the link <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset> for the heart disease dataset. Based on ML techniques, a total of 1026 records have been analyzed on the heart disease dataset collected from the Kaggle repository, including 743 training samples and 283 records to predict heart disease by testing.

Table 2: Simulation Parameter.

| Parameter         | Value                 |
|-------------------|-----------------------|
| Dataset Name      | Heart disease dataset |
| Number of Dataset | 1026                  |
| Tool              | Jupyter               |
| Language          | Python                |
| Training          | 743                   |
| Testing           | 283                   |

Source: Authors, (2026).

As shown in table 2, Jupiter notebooks can be created and analyzed by implementing the Python programming language while using the simulation parameters for comparative analysis. Additionally, heart disease datasets can be evaluated to identify patients with heart disease and healthy patients. The classification problem's performance ratio can be calculated by analyzing the values of TP, FP, TN, and FN, as presented. These factors can evaluate the classification model's performance derived from the confusion matrix.

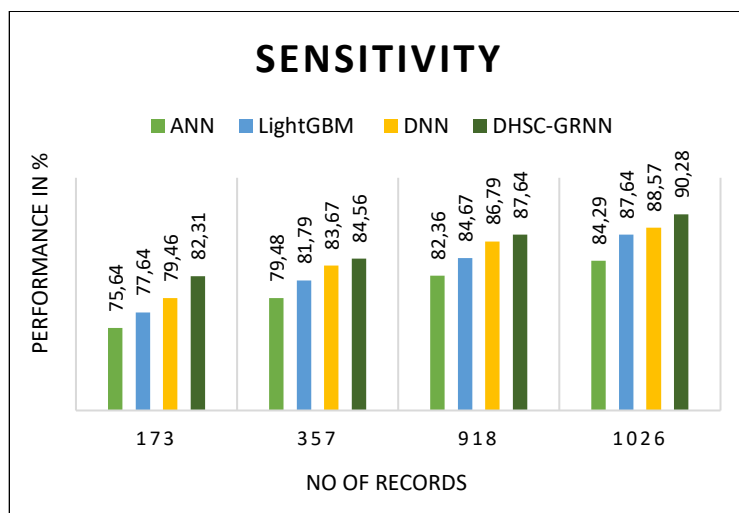


Figure 3: Analysis of Sensitivity.

Source: Authors, (2026).

As shown in figure 3, the accuracy can be predicted by performing the sensitivity analysis and analysing the heart disease datasets of various selected feature subsets. Furthermore, their accuracy increased to 90.28% in predicting the accuracy of sensitivity analysis using the proposed ESVMRF method.

Table 3: Analysis of Comparison Methods.

| Number of methods /records | ANN   | LightGBM, | DNN   | DHSC-GRNN |
|----------------------------|-------|-----------|-------|-----------|
| Accuracy                   | 83.78 | 84.23     | 90.23 | 94.97     |
| F1-Score                   | 81.19 | 84.19     | 90.12 | 93.78     |
| Recall                     | 82.69 | 83.84     | 89.45 | 94.11     |
| Specificity                | 82.63 | 82.56     | 87.91 | 95.66     |
| Precision                  | 85.22 | 83.75     | 90.44 | 95.89     |
| Sensitivity                | 89.12 | 85.67     | 90.21 | 95.12     |
| False rate                 | 17.81 | 15.23     | 11.89 | 6.56      |

Source: Authors, (2026).

Moreover, comparing the proposed method with the literature-derived methods such as DNN, XGBoost, and SVM, their accuracy improved to 82.31, 84.56, and 87.64%, respectively. Similarly, the techniques presented can enhance the accuracy of sensitivity analysis using heart disease datasets as shown in table 3.

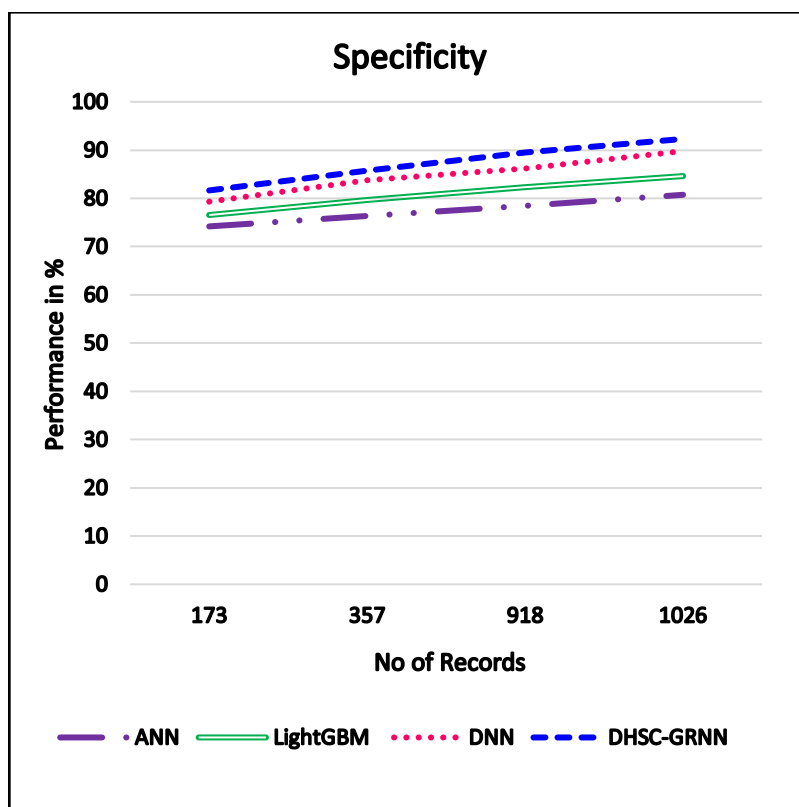


Figure 4: Analysis of Specificity.

Source: Authors, (2026).

As depicted in Figure 4, accuracy prediction involves conducting specificity analysis on the heart disease dataset across various selected feature subsets. When compared to previous methods like ANN, DNN, LightGBM from existing literature, the accuracy enhancements are significant, reaching 80.73%, 84.61%, and 89.79%, respectively. These DHSC-GRNN techniques showcased here have the potential to enhance specificity in heart dataset analyses. Furthermore, employing the proposed DHSC-GRNN method for accuracy prediction in specificity analysis led to a significant improvement, achieving an accuracy of 92.31%.

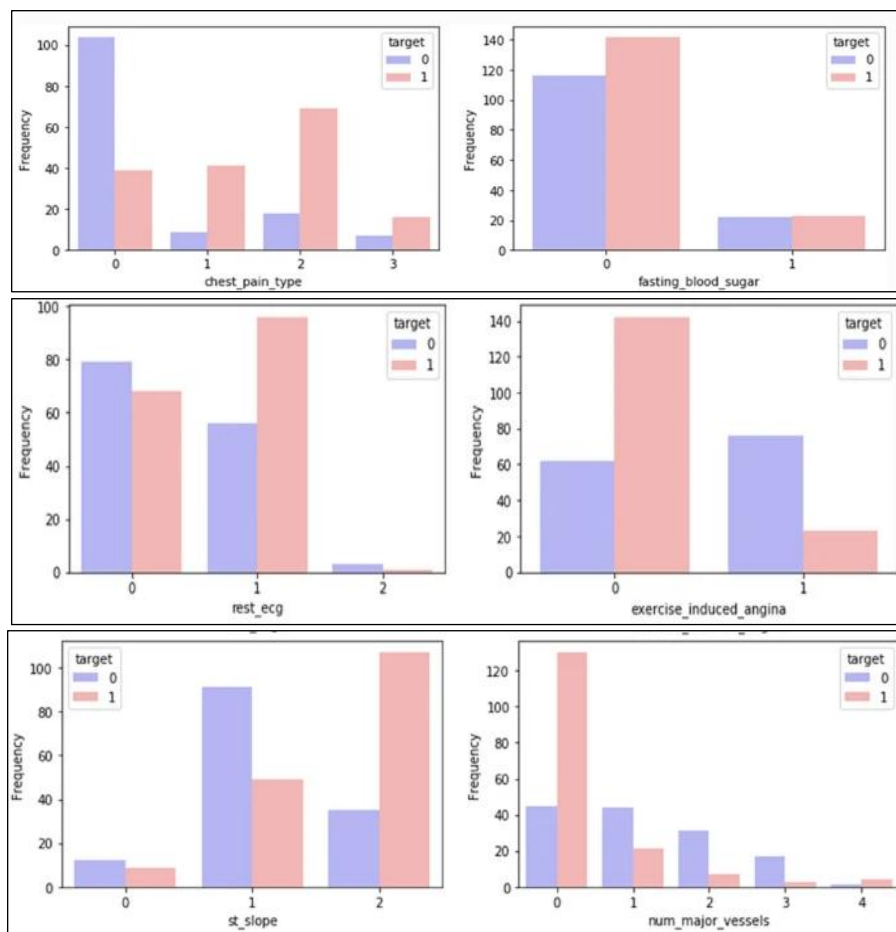


Figure 5: Frequency Absolute Rate.

Source: Authors, (2026).

In Figure 5, accurate prediction is demonstrated through a detailed analysis of the heart disease dataset across different selected feature subsets. A significant improvement can be achieved by evaluating the accuracy levels achieved by previous methods, such as ANN-87%, LightGBM-90.18%, and DNN-91.23% with the proposed methods. Furthermore, the proposed methods can be used to analyze a heart disease dataset and provide an assessment of improving accuracy. Additionally, utilizing the offered DHSC-GRNN technique for predictive precision analysis resulted in a significant accuracy enhancement of 93.46%.

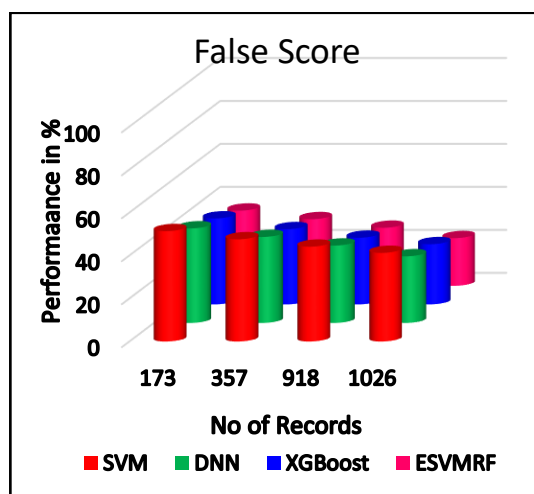


Figure 6: Analysis of False Rate.

Source: Authors, (2026).

As described in Figure 6, testing the features on the heart disease dataset involves evaluating the false scores on subsets of their selected features to provide accurate prediction. When compared to other methods like ANN, DNN, LightGBM from existing literature, the proposed method shows lower accuracy scores of 41.23%, 31%, and 28%, respectively. Similarly, employing the proposed DHSC-GRNN method for accuracy prediction in error score analysis results in a 22.18% lower accuracy in emissions classification. These techniques showcased have the potential to enhance false scores when assessing heart disease datasets.

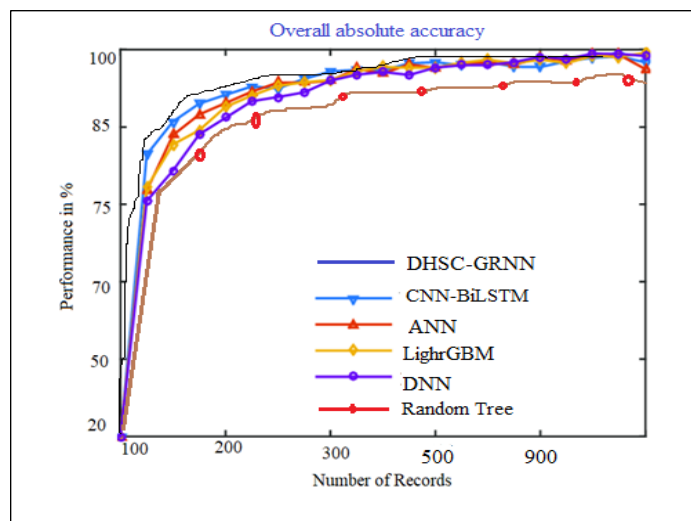


Figure 7: Analysis of Accuracy.  
Source: Authors, (2026).

Figure 7 demonstrates that utilizing features from a heart disease dataset can offer accurate analysis for heart disease patients utilizing selected feature subsets. Additionally, the suggested DHSC-GRNN approach can select heart data features and enhance accuracy to 95.4%. Furthermore, when compared to previous methods, the accuracy improves to ANN-91.39%, LightGBM-93.17%, and DNN-93.17%.

## V.CONCLUSION

The proposed system represents a significant advancement in the field of heart disease prediction by leveraging cutting-edge AI techniques to address the limitations of traditional diagnostic methods. By employing the OSSIT for feature selection and the DHSC-GRNN for predictive modelling, this system aims to enhance the accuracy and reliability of heart disease predictions. As the prevalence of CVD continues to rise, the implementation of such innovative approaches is essential for improving early detection and intervention strategies. Ultimately, this proposed system has the potential to save lives by enabling healthcare providers to identify at-risk individuals more effectively and implement timely preventive measures. As research in this domain progresses, ongoing refinement and validation of the system will be crucial to ensure its efficacy and applicability in real-world clinical settings. In addition to the benefits outlined, the proposed DHSC-GRNN method can effectively select relevant features from heart data, leading to a notable enhancement in accuracy, which reaches an impressive level of 95.4%. This is particularly significant when we compare this new approach to previously established methods. The DHSC-GRNN exhibited a performance increase over traditional techniques such as ANN, which achieved an accuracy of 91.39%, and LightGBM and DNN, both of which demonstrated accuracy rates of 93.17%. Each of these comparisons highlights the effectiveness of the DHSC-GRNN method in achieving superior results in assessing heart data.

## VI. AUTHOR'S CONTRIBUTION

**Conceptualization:** D. Rajapriya and Dr. M. Vigenesh.

**Methodology:** D. Rajapriya and Dr. M. Vigenesh.

**Investigation:** D. Rajapriya and Dr. M. Vigenesh.

**Discussion of results:** D. Rajapriya and Dr. M. Vigenesh.

**Writing – Original Draft:** D. Rajapriya and Dr. M. Vigenesh.

**Writing – Review and Editing:** D. Rajapriya and Dr. M. Vigenesh.

**Resources:** D. Rajapriya and Dr. M. Vigenesh.

**Supervision:** D. Rajapriya and Dr. M. Vigenesh.

**Approval of the final text:** D. Rajapriya and Dr. M. VigeneshS

## VII. REFERENCES

- [1] Akella, A., & Akella, S. (2021). Machine learning algorithms for predicting coronary artery disease: Efforts toward an open-source solution. *Future Science OA*, 7(6), FSO698. <https://doi.org/10.2144/fsoa-2020-0206>.
- [2] Duan, M., Zhang, Y., Liu, Y., Mao, B., Li, G., Han, D., & Zhang, X. (2024). Machine learning aided non-invasive diagnosis of coronary heart disease based on tongue features fusion. *Technology and Health Care*. [https://doi.org/10.3233\\_THC-230590](https://doi.org/10.3233_THC-230590).
- [3] Sayadi, M., Varadarajan, V., Sadoughi, F., Chopannejad, S., & Langarizadeh, M. (2022). A Machine Learning Model for Detection of Coronary Artery Disease Using Noninvasive Clinical Parameters. *Life*, 12(11), 1933. <https://doi.org/10.3390/life12111933>.
- [4] Saeedbakhsh S, Sattari M, Mohammadi M, Najafian J, Mohammadi F. Diagnosis of Coronary Artery Disease based on Machine Learning algorithms Support Vector Machine, Artificial Neural Network, and Random Forest. *Adv Biomed Res*. 2023 Feb 25; 12:51. doi: 10.4103/abr.abr\_383\_21. PMID: 37057235; PMCID: PMC10086656.
- [5] D. Y. Omkari and K. Shaik, "An Integrated Two-Layered Voting (TLV) Framework for Coronary Artery Disease Prediction Using Machine Learning Classifiers," in *IEEE Access*, vol. 12, pp. 56275-56290, 2024, doi: 10.1109/ACCESS.2024.3389707.

- [6] T. Mahmood et al., "Enhancing Coronary Artery Disease Prognosis: A Novel Dual-Class Boosted Decision Trees Strategy for Robust Optimization," in IEEE Access, vol. 12, pp. 107119-107143, 2024, doi: 10.1109/ACCESS.2024.3435948.
- [7] Dipto IC, Islam T, Rahman HM, Rahman MA, Dipto IC, Islam T, et al. Comparison of different machine learning algorithms for the prediction of coronary artery disease. [Last accessed on 2021 Sep 07]; J Data Anal Inf Process. 2020 8:41–68. Available from: <http://www.scirp.org/journal/PaperInformation.aspx?PaperID=99402>.
- [8] Safial Islam Ayon, Md. Milon Islam & Md. Rahat Hossain (2020): Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques, IETE Journal of Research, DOI: 10.1080/03772063.2020.1713916.
- [9] Koloi, A., Loukas, V. S., Hourican, C., Sakellarios, A. I., Quax, R., Mishra, P. P., Lehtimäki, T., Raitakari, O. T., Papaloukas, C., Bosch, J. A., März, W., & Fotiadis, D. I. (2024). Predicting early-stage coronary artery disease using machine learning and routine clinical biomarkers improved by augmented virtual data. European Heart Journal - Digital Health, 5(5), 542-550. <https://doi.org/10.1093/ehjdh/ztae049>
- [10] Wang, J., Xue, Q., Zhang, C. W., Wong, K. K., & Liu, Z. (2024). Explainable coronary artery disease prediction model based on AutoGluon from AutoML framework. Frontiers in Cardiovascular Medicine, 11, 1360548. <https://doi.org/10.3389/fcvm.2024.1360548>.
- [11] Garavand, A., Salehnasab, C., Behmanesh, A., Aslani, N., Zadeh, A. H., & Ghaderzadeh, M. (2021). Efficient Model for Coronary Artery Disease Diagnosis: A Comparative Study of Several Machine Learning Algorithms. Journal of Healthcare Engineering, 2022(1), 5359540. <https://doi.org/10.1155/2022/5359540>
- [12] Wang, C., Zhao, Y., Jin, B., Gan, X., Liang, B., Xiang, Y., Zhang, X., Lu, Z., & Zheng, F. (2021). Development and Validation of a Predictive Model for Coronary Artery Disease Using Machine Learning. Frontiers in Cardiovascular Medicine, 8, 614204. <https://doi.org/10.3389/fcvm.2021.614204>.
- [13] Mansoor, C. M. M., Sarat Kumar Chettri, and H. M. M. Naleer. "Development of an efficient novel method for coronary artery disease prediction using machine learning and deep learning techniques." Technology and Health Care Preprint (2024): 1-25.
- [14] Ghrabat, M. jalil J., Mohialdin, S.H., Abdulrahman, L.Q., Al-Yoonus, M.H., Abduljabbar, Z.A., Honi, D.G., Nyangaresi, V.O., Abduljaleel, I.Q. and Neamah, H.A. 2024. Utilizing Machine Learning for the Early Detection of Coronary Heart Disease. Engineering, Technology & Applied Science Research. 14, 5 (Oct. 2024), 17363–17375. DOI: <https://doi.org/10.48084/etasr.8171>.
- [15] Hosur, R., Mahendrakar, P., Nagaral, M., Hiremath, A., & Patil, S. B. (2024). Comparative Analysis for Prediction of Coronary Artery Disease Using Machine Learning Algorithms. International Journal of Intelligent Systems and Applications in Engineering, 12(12s), 735 –. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/4559>.
- [16] Muhammad, L.J., Al-Shourbaji, I., Haruna, A.A. et al. Machine Learning Predictive Models for Coronary Artery Disease. SN COMPUT. SCI. 2, 350 (2021). <https://doi.org/10.1007/s42979-021-00731-4>.
- [17] Ghiasi, Mohammad M., Sohrab Zendejboudi, and Ali Asghar Mohsenipour. "Decision tree-based diagnosis of coronary artery disease: CART model." Computer methods and programs in biomedicine 192 (2020): 105400.
- [18] Dutta, Aniruddha, et al. "An efficient convolutional neural network for coronary heart disease prediction." Expert Systems with Applications 159 (2020): 113408.
- [19] Zomorodi-moghadam, Mariam, et al. "Hybrid particle swarm optimization for rule discovery in the diagnosis of coronary artery disease." Expert Systems 38.1 (2021): e12485.
- [20] El-Ibrahimi, A., Daanouni, O., Alouani, Z., El Gannour, O., Saleh, S., Cherradi, B., & Bouattane, O. (2024). Fuzzy based system for coronary artery disease prediction using subtractive clustering and risk factors data. Intelligence-Based Medicine, 11, 100208. <https://doi.org/10.1016/j.ibmed.2025.100208>.
- [21] L. -C. Chen et al., "Self-Supervised Learning-Based General Laboratory Progress Pretrained Model for Cardiovascular Event Detection," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 12, pp. 43-55, 2024, doi: 10.1109/JTEHM.2023.3307794.
- [22] J. J. Gabriel and L. Jani Anbarasi, "Accurate Cardiovascular Disease Prediction: Leveraging Opt\_hpLGBM With Dual-Tier Feature Selection," in IEEE Access, vol. 12, pp. 142427-142448, 2024, doi: 10.1109/ACCESS.2024.3470537.
- [23] A. Kumar, K. U. Singh and M. Kumar, "A Clinical Data Analysis Based Diagnostic Systems for Heart Disease Prediction Using Ensemble Method," in Big Data Mining and Analytics, vol. 6, no. 4, pp. 513-525, December 2023, doi: 10.26599/BDMA.2022.9020052.
- [24] S. Bhattacharyya, S. Majumder, P. Debnath and M. Chanda, "Arrhythmic Heartbeat Classification Using Ensemble of Random Forest and Support Vector Machine Algorithm," in IEEE Transactions on Artificial Intelligence, vol. 2, no. 3, pp. 260-268, June 2021, doi: 10.1109/TAI.2021.3083689.
- [25] X. Dong and W. Si, "Heartbeat Dynamics: A Novel Efficient Interpretable Feature for Arrhythmias Classification," in IEEE Access, vol. 11, pp. 87071-87086, 2023, doi: 10.1109/ACCESS.2023.3305473.
- [26] Taylan, O., Alkabaa, A. S., Alqabaa, H. S., Pamukçu, E., & Leiva, V. (2022). Early Prediction in Classification of Cardiovascular Diseases with Machine Learning, Neuro-Fuzzy and Statistical Methods. Biology, 12(1), 117. <https://doi.org/10.3390/biology12010117>
- [27] Rath, A., Mishra, D., Panda, G., & Satapathy, S. C. (2021). Heart disease detection using deep learning methods from imbalanced ECG samples. Biomedical Signal Processing and Control, 68, 102820. <https://doi.org/10.1016/j.bspc.2021.102820>
- [28] Uddin, M. N., & Halder, R. K. (2020). An ensemble method based multilayer dynamic system to predict cardiovascular disease using machine learning approach. Informatics in Medicine Unlocked, 24, 100584. <https://doi.org/10.1016/j.imu.2021.100584>
- [29] Dritsas, Elias, Sotiris Alexiou, and Konstantinos Moustakas. "Cardiovascular Disease Risk Prediction with Supervised Machine Learning Techniques." ICT4AWE 1 (2022): 315-321.
- [30] Rani, P., Kumar, R., Ahmed, N.M.O.S. et al. A decision support system for heart disease prediction based upon machine learning. J Reliable Intell Environ 7, 263–275 (2021). <https://doi.org/10.1007/s40860-021-00133-6>