International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Efficient Coronary Heart Disease Prediction: Enhanced Neural Network with Chaos Salp Decision and Feature Optimization

A. Asha[1], B. Dhiyanesh[2], G. Kiruthiga[3*], L. Shakkeera[4], Vinodkumar Jacob[5], Anita Venaik[6]

Submitted: 18/01/2024 **Revised**: 27/02/2024 **Accepted**: 04/03/2024

Abstract*:* In recent times, Coronary Heart Disease (CHD), with its high incidence and mortality risk, has emerged as a significant threat to human health. HD prediction is a challenging task that requires expertise and advanced learning, so physicians cannot predict it efficiently. Early HD diagnosis through clinical examination and simple physical indicators is necessary. Medical professionals face challenges in diagnosing HD. Nevertheless, comparing techniques to determine which one is faster or more accurate is a time-consuming and challenging task. However, since CHD HD is not a benign disease, finding and analysing CHD markers in screening is difficult. To solve this problem, we propose the Enhanced Elman Selfish Optimization Neural Network (EESONN) to find CHD prediction by selecting the best features based on the Chaos Salp Group Decision Function (CSSDF) method. Additionally, we collected data from Kaggle using the Cleveland Clinic Cardiology Dataset. We also process the data before Improved Quality Feature Group (IQFG), such as lack of significance, normalization and standardization. In addition, the Weighted Relief Algorithm (WRA) can be applied to eliminate redundant features and determine each feature's weight. Furthermore, optimal features can be selected using the CSSDF algorithm in the feature selection process. Finally, the proposed EESONN method to detect predictive data can effectively predict CHD classification. Furthermore, the proposed EESONN method enhances classifier prediction accuracy and performs well in HD detection. Additionally, the EESONN method enables the calculation of sensitivity, precision, specificity, accuracy, and F-score. This can be used to determine the accuracy and probability of the results.

Keywords: *Classification, Decision, Heart Disease, Neural Network, Optimization*

1. Introduction

Heart and blood vessel problems, collectively known as cardiovascular disease (CVD), are primarily brought on by smoking, eating poorly, and not exercising. Reducing the risk of long-term problems and death from CVD requires early diagnosis and treatment. Many open-source codes exist for accessing patient records, and research is ongoing to utilize computational techniques for accurate patient diagnosis, disease diagnosis, and mortality prevention. Hearts are muscular organs that pump blood throughout the body and are responsible for pumping oxygen to those cells. It promotes it as a vital component of the human cardiovascular system, including the lungs [1]. The human body has an extensive network of veins, arteries, and capillaries that aid in blood transportation throughout the body. A heart that does not receive enough blood can lead to irregularities in the physiology of different types of heart disease, commonly referred to as CHD Accordingly, early detection devices and it is possible to save lives by predicting heart disease and making effective treatment plans for heart disease, and eventually reduce CVD mortality. Advanced

A range of factors, such as personal habits, occupation habits, and genetic propensities, can cause heart disease. Furthermore, these include chronic risk factors such as excessive alcohol consumption, stress, and lack of exercise. Early detection of cardiac disease is crucial for survival. Additionally, data mining involves removing vital data from vast data sets across various fields such as medicine, business, and education [3].

Observing established risk factors for heart disease can make it difficult to predict heart disease in the future. Among these are hypertension, hypercholesterolemia, and erratic heart rates. Moreover, this kind of issue arises when the heart muscle's blood supply is obstructed, potentially harming nearby muscles. Diagnosing heart disease is crucial. Heart disease detection has led to invasive medical procedures like angiography. These procedures, despite being costly, can also cause side effects for diagnosed patients [4]

[1]*Professor, Department of Electronics and Communication Engineering, Rajalakshmi Engineering College, Chennai 602105, Tamil Nadu, India. Email: ngash78@gmail.com*
[2]*Associate Professor, Department of Computer Science and Engineering, Dr. N.G.P. Institute of Technology, Coimbatore 614048, Tamil Nadu, India. Email: dhiyanu87@gmail.com,*
[3]*Professor, Department of Computer Science and Engineering, Karpagam College of Engineering, Coimbatore 641032, Tamil Nadu, Email: kiruthiga.g@kce.ac.in**
[4]*Associate Professor, Department of Computer Science and Engineering, Presidency University, Bengaluru 560064, Karnataka, India. Email: shakkeera.l@presidencyuniversity.in*
[5]*Professor, Department of Electronics and Communication Engineering, M.A College of Engineering, Kothamangalam 686666, Kerala, India, Email: vkj@mace.ac.in*
[6]*Professor, Department of Information Technology, Amity Business School, Amity University, Noida 201313, India. Email: avenaik@amity.edu*
** Corresponding Author Email: kiruthiga.g@kce.ac.in*

Predicting heart disease is a challenging task in clinical data analysis. With proper analysis, the impact on patient health may be significant. It is imperative to acknowledge that cardiovascular disorders encompass a spectrum of ailments, including but not limited to vascular, rheumatic, and coronary heart diseases. Diabetes and high blood pressure are examples of genetic variables that raise one's risk of heart disease. Physical inactivity and an unhealthy diet are secondary factors that further increase the risk. CVD requires every moment for diagnosis and treatment. However, weariness, perspiration, back discomfort, and chest, shoulder, and arm pain are typical signs of CVD [5].

Data from the Cleveland Clinic Heart Disease Dataset is gathered via Kaggle. Additionally, missing values will be handled and the data will be optimized prior to preprocessing, the IDQFG method is implemented within the data pre-processing process. Additionally, WRA implements features with high-weight values selected, while low-weight values are discarded. Subsequently, the largest information coefficients are merged, and redundant features are removed. Additionally, the CSSDF algorithm is used in the feature selection phase to choose the most favourable features. Lastly, the findings of the suggested approach, EESONN, show increased inefficiency classifier prediction accuracy, suggesting appropriate CHD detection performance.

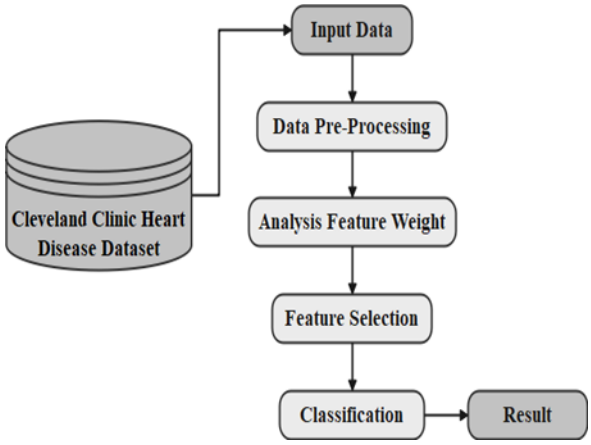

Fig 1. The Basic Architecture Diagram Coronary Heart Disease Detection

Figure 1 introduces a basic structural diagram for CHD diagnosis. The process includes data pre-processing, feature weight analysis, selection of optimal features, and prediction using the proposed classifier. The classifier is trained with a set of labelled samples. The model is then used for classifying new unseen samples. Metrics are used to assess the mode"s performance and determine how effective it is.

2. Literature Survey

The author stated that using a CNN method is compatible with current Heart Rate Variability (HRV) based methods. Furthermore, the raw electrocardiogram can accurately diagnose heart failure based on heart rate [6]. The author proposes ML classifiers as a means of identifying relevant features that can aid in heart disease detection and prediction. Ranking predictive models involves combining a variety of features with established classification algorithms [7]. The author proposes employing the Hybrid Decision Support System (HDSS) method, utilizing patient clinical characteristics, for early cardiac disease detection [8]. The author suggests using an ML algorithm to assess its application in heart failure diagnosis, classification, and prediction with clinician input [9]. The author used the Medical Diagnosis Support System (MDSS) patient clinical data to develop ML approaches for atherosclerosis prediction [10].

The author suggested that efficient CHD risk prediction can be achieved by training Deep Neural Networks (DNNs) on ordered datasets [11]. [12] proposed using ML techniques to diagnose Chronic Heart Failure (CHF). This novel proposes an ML algorithm for determining whether human heart disease is due to a heart condition based on the data. Combining clinical and pathological data enables heart disease identification [13]. The author suggested that ML algorithms can be utilized to uncover data set properties and relationships between them [14]. The novel proposes that the ML algorithm can specify the most accurate classifier. The performance and accuracy of CVD prediction can be approximated using supervised ML techniques [15].

The author suggests that differentiating recordings between patients and healthy subjects can provide promising results in detecting different CHF stages. Furthermore, these resources analyse heart failure based on heart sounds [16]. The novel suggests that feature engineering techniques and feature selection can enhance Principal Component Heart Failure (PCHF) performance [17]. The author proposes that machine learning approaches can be used to detect CVD and its risk factors. Furthermore, they can be analysed using predictive tests to predict heart disease [18]. The author proposes an effective and precise cardiac diagnosis method based on ML. A feature selection algorithm enhances accuracy and reduces runtime, depending on the classification system [19]. The author proposed a Dense Neural Network (DNN) approach, which implements 3–9 layers and employs different hidden layer formats. Identifying hidden patterns is challenging [20].

The novel suggests that Deep Learning (DL) videos can be analysed using models to identify changes in both shape and timing using foetal cardiac ultrasonography (US) technology [21]. The author suggests that ML-based high accuracy can be implemented to detect CVD prediction. However, identifying predictors of clinical

data processing for CVD is difficult [22]. The author introduced an enhanced DL-CNN (EDCNN) method to support and improve the needs of patients with heart disease [23]. The proposed novel implements pre-processing techniques and Genetic Algorithms (GA) to select features, improving prediction performance and time consumption [24]. A model combining various methods is suggested by the author as a means of predicting heart disease in an effective manner [25].

The author proposes an improved 3D U-CNN (U-CNN) method for cardiovascular and coronary artery segmentation using DL to predict disease risk [26]. The novel proposes an ML-based CVD Diagnosis Framework (MaLCaDD) for more precise and efficient CVD prediction [27]. The author suggests that Obstructive Sleep Apnea (OSA) can be diagnosed by relying on HRV features using various algorithms in ML [28]. The author suggests that the implementation is based on two layers: the base layer and the Meta layer [29]. The author suggested that the imbalance issue can be effectively detected and predicted using the SMOTE [30].

The author noted that the creation of large-scale datasets of biomedical measurements from the CVD Qatar Biobank (QBB) could be achieved based on anthropometric data, clinical biomarkers, and behavioral factors [31]. The author suggested that an intelligent system model could be used to diagnose CHD while considering diagnostic tests. However, their models for selecting features for diagnosis caused problems due to research cost considerations [32]. The authors proposed that the method could be implemented with an Optimal Configured Improved Deep Belief Network (OCI-DBN) technique [33]. The author proposed that regression and classification can be predicted using the CHD dataset by applying data mining techniques [34]. The author suggested implementing the Supervised Infinite Feature Selection (I-FS) method as an effective approach to identifying heart disease and improving patient survival. Machine learning faces persistent class inequality and high-dimensional problems. [35].

3. Proposed Methodology

To assess heart disease using our suggested method, we first present a Cleveland Clinic Heart Disease Dataset that we collected from Kaggle. Furthermore, operations such as missing values, normalization, and standardization can be performed using the IQFG method to pre-process the data based on the dataset. In addition, the WRA algorithm can be implemented to analyze feature weights to remove redundant features and optimize each feature's weight. Also, the CSSDF algorithm can be employed to select the optimal features after improving the feature weights. Finally, predictive data for CHD can be identified and predicted using the proposed EESONN classification. Thus, the proposed EESONN method performs well in HD detection, improving the classifier prediction accuracy.

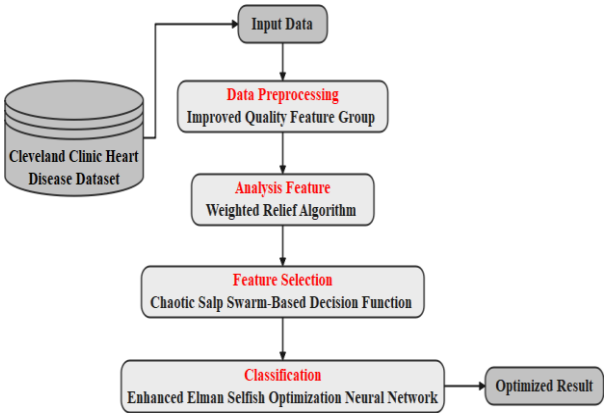

Fig 2. The Architecture Diagram for the Proposed EESONN Method

The EESONN method demonstrates suitable performance in detecting CHD and improves classifier prediction accuracy, as illustrated in Figure 2. The EESONN method also outperforms other traditional methods in terms of sensitivity and specificity. Furthermore, it is fast, robust, and computationally efficient, making it a viable option for CHD detection.

3.1 Dataset Collection

In this dataset collection, Kaggle obtained the Cleveland Clinic Heart Disease dataset to predict CHD. For more information, please visit the link provided within: https://www.kaggle.com/datasets/aavigan/cleveland-clinic-heart-disease dataset. CHD is the most common heart disease, where plaque buildup in the heart arteries reduces blood flow to the heart muscles. However, this prediction method is costly and leads to morbidity and mortality in CHD patients. Developing non-invasive alternatives to the current gold standard would be advantageous. It would allow us to avoid potential complications and improve patient comfort during treatment. The dataset contains 304 observations, 13 features, and one target attribute. These 13 features include non-invasive prediction test results and additional relevant patient data.

age	Sex	cp	trestbps	chol	fbs	resteeg	thalach	exang	oldpeak	slope	ca	thal	num
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	2
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
56	1	3	130	256	1	2	142	1	0.6	2	1	6	2
44	1	2	120	263	1	0	173	0	0	1	0	7	0
52	1	3	172	199	0	0	162	0	0.5	1	0	7	0
57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
48	1	2	110	229	0	0	168	0	1	3	0	7	1
54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
48	0	3	130	275	0	0	139	0	0.2	1	0	3	0
49	1	2	130	266	0	0	171	0	0.6	1	0	3	0
64	1	1	110	211	0	2	144	1	1.8	2	0	3	0
58	0	1	150	283	1	2	162	0	1	1	0	3	0

Fig 3. Cleveland Clinic Heart Disease Dataset

As shown in Figure 3, they utilize Kaggle to evaluate its accuracy using the Cleveland Clinic Heart Disease dataset. As mentioned earlier, these 13 aspects contain non-invasive prediction test results and other relevant patient data.

S.No	Attribute Name	Description	Range of Values
1	Age	Age of the person in years	29 to 79
2	Sex	Gender of the person [1: Male, 0: Female]	0, 1
3	Cp	Chest pain type [1-Typical Type 1 Angina 2- Atypical Type Angina 3-Non-angina pain 4-Asymptomatic)	1, 2, 3, 4
4	Trestbps	Resting Blood Pressure in mm Hg	94 to 200
5	Chol	Serum cholesterol in mg/dl	126 to 564
6	Fbs	Fasting Blood Sugar in mg/dl	0, 1
7	Restecg	Resting Electrocardiographic Results	0, 1, 2
8	Thalach	Maximum Heart Rate Achieved	71 to 202
9	Exang	Exercise Induced Angina	0, 1
10	OldPeak	ST depression induced by exercise relative to rest	1 to 3
11	Slope	Slope of the Peak Exercise ST segment	1, 2, 3
12	Ca	Number of major vessels colored by fluoroscopy	0 to 3
13	Thal	3 – Normal, 6 – Fixed Defect, 7 – Reversible Defect	3, 6, 7

Fig 4. List of Feature Information Cleveland Heart Disease Dataset

Using Kaggle's Cleveland heart disease dataset as an example, figure 4 illustrates the dataset. Furthermore, it covers 303 observations with 14 features: eight in category and six in number.

3.2 Improved Quality Feature Group (IQFG)

The data pre-processing IQFG method can predict missing values, normalization, and standardization. Data pre-processing methods in IQFG often involve replacing missing values by using the average of other values in the attribute. However, this dataset contains a significant number of missing values and redundant and meaningless data, making it essential to transform the data. This process can be achieved using techniques such as data normalization, which is a common technique. Data normalization, in this IQFG procedure, involves converting data values into a range between 0 and 1. Numerical features are sorted by value and divided into equal-width bins. Furthermore, the range of sorted values is divided into predefined equal-frequency bins, each containing the same number of values.

Equation 1 calculates the distance and similarity of features to replace missing values. Let's assume W - Distance, A and B - values, and q - amount of data.

$$W_{[A,B]} = \sqrt{\sum_{l=1}^{q}(A_l - B_l)^2} \tag{1}$$

Calculate the missing values of the mean measurement as shown in Equation 2. Where P-mean value, Q - data, and A_l - data represent a row.

$$P(A) = \frac{\sum_{l-0}^{Q} A_l}{Q} \tag{2}$$

To find the minimum and maximum values of numerical features, refer to Equation 3. Let's assume we $\bar{X}$ - select the values for normalization, $\mathbb{mm}_{\mathbb{in}}$ - minimum value, and $\mathbb{mm}_{\mathbb{ax}}$ - maximum value.

$$\bar{A} = \frac{A - \mathbb{mm}_{\mathbb{in}}}{\mathbb{mm}_{\mathbb{ax}} - \mathbb{mm}_{\mathbb{in}}} \tag{3}$$

Equation 4 shows that normalization can be performed using zero mean or Z-score normalization. Let's assume A'_l −normalized data, $P_{(A)}$ − signifies mean value, $H_W(A)$ − standard deviation value.

$$A'_l = \frac{A_l - P_{(A)}}{H_W(A)} \tag{4}$$

Calculate the mean of the standard deviation shown in Equation 5. Where σ − sigma, σ − indicate standard deviation attribute.

$$\sigma = \sqrt{\frac{1}{Q-1}\sum_{l=1}^{Q}\left(A_l - \left(P_{(A)}\right)^2\right)} \tag{5}$$

Equations 6 and 7 calculate the threshold and rank values for equal-width bins. Let's assume the D-width of the bin, $F_{\mathbb{mm}_{\mathbb{in}}}$ − minimum value, $F_{\mathbb{mm}_{\mathbb{ax}}}$ − maximum value, J − numerical feature, Y_e − boundaries value,

$$D = \frac{F_{\mathbb{mm}_{\mathbb{ax}}} - F_{\mathbb{mm}_{\mathbb{in}}}}{J} \tag{6}$$

$$Y_e = F_{\mathbb{mm}_{\mathbb{in}}} + (l * D) \tag{7}$$

As shown in Equation 8, the quality of features can be improved by estimating thresholds that categorize values into predefined equal frequency bins. Let's assume v - equal frequency centralization, H_G − split range, s - sorting value, and $\frac{Q}{J}$ − equal frequency bins.

$$H_G = F\sum_H\left(\frac{Q}{J}\right) \tag{8}$$

Implementing data preprocessing permits teams to optimize data based on the methods involved in processing it.

3.3 Weighted Relief Algorithm (WRA)

The WRA algorithm removes redundant features and determines the weight of each feature. In WRA, features in each dataset are assigned weights automatically based on their importance in the dataset. High weighted features are considered more significant than low weighted features. Moreover, WRA also considers the distance between features when assigning weights. If the distance between some features is greater than the size of those features, the feature's weight is increased. Conversely, when the distance between some features is lower, the feature's weight is decreased. WRA also optimizes the random sample of the training set to ensure that its samples from the minority class. The feature weights for samples from the majority class are calculated for this purpose. The average of feature weights increases the probability of selecting boundary samples.

Algorithm 1

Input : Improved Quality of feature H_G

Output : weight for each feature

Begin

Step 1 : Find the total amount of training samples

Step 2 : Estimate the quantity of dimensions of a feature

Step 3 : Computational weight of feature set

$D(X) = 0.0$

for $J \leftarrow 1$ to p do

Step 4 : Choose a "target" sample and weight at random.

Step 5 : Evaluate detected near hits and near miss.

for $X \leftarrow 1$ to p do

$$D(X) \leftarrow D(X) - \mathcal{D}_{iff}(X,\ G_J, S)/p + \mathcal{D}_{iff}(X,\ G_J, P)\ /p$$

End for

End for

Step 6 : Return← D [A feature vector to compute the feature quality]

Stop

The WRA algorithm iterates with random training samples and calculates each sample's weight using a "target." Algorithm 1 shows WRA's pseudocode. Let's assume m-parameter, w-weight, and R_k - a random training sample.

To calculate the weighted average of the individual feature weights, Equation 9 can be used. Let's assume P-nearest miss, S-nearest hit, k-feature sample, $A_l(k)$ −sample value, p −number of random extractions.

$$D_k = D_k + \frac{w(A_l(k), p_l(k))}{p}, \frac{w(A_l(k), S_l(k))}{p} \quad (9)$$

Compute the numerical feature distance function as shown in equation 10.

$$W(A_l(k), P_l(k)) = \left[\frac{A_l(k) - P_l(k)}{max(k) - min(k)}\right] \quad (10)$$

As shown in Equation 11, estimate the distance of the measured sample from the threshold. Let's assume δ −large the value, $\mathbb{N}\mathbb{u}\mathbb{m}_Q$ − number of sample different class.

$$\delta = \frac{\mathbb{N}\mathbb{u}\mathbb{m}_Q}{J} \quad (11)$$

To calculate the importance of each feature, refer to Equation 12 for the corresponding weights. Let's assume m-positively correlated, u-weight.

$$m = \frac{1}{1+u^{-\delta}} \quad (12)$$

In this method, feature weights are estimated to activate selected models closer to the threshold. Additionally, the probability value of the selected model increases with the feature weight.

3.4 Chaotic Salp Swarm-Based Decision Function (CSSDF)

The CSSDF algorithm can be applied to determine the most appropriate parameter for a given choice. Feature selection is a critical process that involves identifying, removing, and iterating on irrelevant features. One way to improve this process is to employ CSSDF optimization. These methods make feature selection faster and more efficient. The CSSDF algorithm can also help reduce residual data and identify the most relevant features. Additionally, random variables can be replaced with chaotic variables using CSSDF chaos graphs. This feature can help determine the potential impact of the selection. A weighting scale can be used to combine two targets, representing the dimensions of selected attributes and total attributes. In addition, weights controlling the importance and accuracy of specific features can also be estimated. Furthermore, the CSSDF can be improved by choosing the most appropriate optimization method to evaluate the weighting parameters.

Calculate the crawling behaviour of the process salp as shown in Equation 13. Where l–dimension, B_l^i −position of leader, ey_l & iy_l − lower and upper boundary value, V_l −food's position, g_1, g_2, g_3 − random number.

$$B_l^i = m\begin{cases} V_l + g_1\big((ey_l - lb_i)g_2 + iy_l\big) & g_3 \geq 0 \\ V_l - g_1\big((ey_l - lb_i)g_2 + iy_l\big) & g_3 \geq 0 \end{cases} \quad (13)$$

Equation 14 computes a random number to balance attack relationships. Let's assume $\mathcal{T}$ -haphazard number u-extreme amount of iteration.

$$g_2 = 2_u^{-(4t/\mathcal{T})^2} \quad (14)$$

Calculate the power level of the current iteration as shown in Equation 15. Let's assume β_0 −initial speed, j-position.

$$B_l^k = \frac{1}{2}\alpha i^2 + \beta_0 i \quad (15)$$

Estimate the equivalent iteration time in the optimization process, as shown in Equation 16.

$$B_l^k = \frac{1}{2}\big(B_l^k + B_l^{k-1}\big) \quad (16)$$

Equation 17 shows the estimate of the update factors based on the chaotic graph.

Where $\Psi_H -$ swarm value.

$$G_2 = \Psi_H \quad (17)$$

Equation 18 shows the result of calculating the updated state of the chaotic map. Where t-iteration, $\Psi^t -$ gained value.

$$B_l^1 = m\begin{cases} V_l + g_1\big((ey_l - iy_l)\Psi^t + iy_l\big) & g_3 \geq 0 \\ V_l - g_1\big((ey_l - iy_l)\Psi^t + iy_l\big) & g_3 \geq 0 \end{cases} \quad (18)$$

Equations 19 and 20 show the computation of an agent transfer from one consecutive binary location. Let's assume Y- an arbitrary number.

$$B_l^t = \begin{cases} 1 & if\big(H(B_l^t)\big) \geq Y \\ 0 & otherwise \end{cases} \quad (19)$$

$$H(B_l^t) = \frac{1}{1+u^{10\left(B_l^t - 0.5\right)}} \quad (20)$$

Equation 21 computes a fitness function to improve feature size accuracy. Let's assume $\mathcal{A}_{\mathbb{CC}}$ - accuracy, $I_v -$ feature length, $I_t -$ total number of features, $d_v -$ weight factor, $v_{\mathbb{v}}$ −fitness function.

$$v_{\mathbb{v}} = \mathbb{mm}_{\mathbb{ax}}\left[d_v \times \mathcal{A}_{\mathbb{CC}} + (1 - d_v) \times \left(1 - \frac{I_v}{I_t}\right)\right] \quad (21)$$

To calculate the feature weighting factors, refer to Equation 22. Where y-bias value, Q -number of features.

$$v(A) = \sum_{l=1}^{Q} d_l A_l + Y \quad (22)$$

As shown in Equation 23, the combined sum is computed by removing the empty set from the attributes. Let's assume $\mathcal{T}$ - total value.

$$\mathcal{T} = 2^q - \left(\frac{q!}{1!(q-1)!}\right) - \left(\frac{q!}{1!(q-1)!}\right) - 1 = 2^q - \left(\frac{\mathbb{q}^2 + \mathbb{q}}{2} + 1\right) \quad (23)$$

In this sense, feature classification involves identifying and removing irrelevant features.

3.5 Enhanced Elman Selfish Optimization Neural Network (EESONN)

The EESONN method used in the study successfully identified predictive data that accurately predicts CHD classification. A classification algorithm can be developed to improve EESONN accuracy, and the integer of selected attributes obtained from feature selection can be used as input to the algorithm. The proposed EESONN method comprises four layers: an input, a hidden, an output, and a context layer. The input layer transforms data, the hidden layer maps it with weights, and the output layer linearly processes it. The proposed method of adding a context layer to EESONN allows it to respond to subsequent moments of the hidden layer. It also stores output data from previous moments. This is particularly useful for the EESONN classifier, as it enables the most significant data to be fed back into the network. In this way, the model data can be analysed in more detail in order to prove its validity.

Compute the strongly optimal sub-memory as shown in Equations 24 to 26. Let's assume the $A(t) -$ hidden layer, $B(t) -$ output layer, $A_z(t) -$ context layer, e-input vector, $D_1 -$ connection weight in the context layer, $D_2 -$ weight value of the input layer, $D_3 -$ hidden layer weight, v (.) - input layer of the transfer function, and R (.) - a transfer function's output layer.

$$A(\mathbb{t}) = \mathbb{T}\left(d_1 A_z(\mathbb{t}) + d_2\big(e(\mathbb{t}-1)\big)\right) \quad (24)$$

$$B(t\mathbb{t}) = R\big(d_3 A(\mathbb{t})\big) \quad (25)$$

$$A_z(\mathbb{t}) = a(\mathbb{t}-1) \quad (26)$$

Compute the constant of rate value as shown in Equation 27.

$$A_z(t) = A(\mathbb{t}-1) + \alpha A_z(\mathbb{t}-1) \quad (27)$$

The cross-entropy calculated in Equation 28 is used as a loss function. Let's assume $w_k(t) -$ value of raw data, $B_k(t) -$ gained output value,

$$i(w,B) = -\sum_{k=1}^{q}\big[w_k(t) log B_k(t) + \big(1 - w_k(t)\big) log\big(1 - B_k(t)\big)\big] \quad (28)$$

Calculate the updates of the new location of the selfish, as shown in Equation 29. Where HV-selfish, $A^t_{elite\_HV^{\lambda_l}} -$ elite selfish best position, $A^l_{inj\_HV} -$ position of injured sardine, $A^l_{old\_HV} -$ current position selfish.

$$A^t_{new\_HV} = A^t_{elite\_HV^{\lambda_l}} \times \left(\alpha \times \frac{A^t_{elite\_HV^{\lambda_l}} + A^l_{inj\_HV}}{2}\right) - A^l_{old_{HV}} \quad (29)$$

As shown in Equation 30, evaluate the random numbers within a range of coefficients. Let's assume $\lambda_i -$ coefficient, $M_w -$ important factor.

$$\lambda_i = (2 \times \alpha \times M_w) - M_w \quad (30)$$

Calculate the critical coefficients for the optimal sail condition as described in Equation 31. Let's assume $Q_{HV} -$ number of selfish, $Q_H -$ number of sardines.

$$M_w = 1 - (Q_{HV} / Q_{HV} + Q_H) \quad (31)$$

The new position of the sardine can be calculated using Equation 32. Let's assume X_m -selfish attack power.

$$A^t_{new\_H} = g \times \left(A^t_{elite\_HV}\lambda_l - A^l_{old\_HV} + X_m\right) \quad (32)$$

Calculate selfish attack force as shown in equation 33. Let's assume $i_{\mathcal{TR}}$ - current iteration, X and ε −decrease the power attack of two factors.

$$X_m = \times \left(1 - (2 \times i_{\mathcal{TR}} \times \varepsilon)\right) \quad (33)$$

Calculate the number of sardines that improve the condition given by equation 34. Let's assume $\eta -$ selected number.

$$\eta = Q_H \times X_m \quad (34)$$

Calculate the alternative states of the swordfish state and the modern state of the hunted sardines, as shown in Equation 35. Let's assume

$$A^l_{HV} = A^l_H if\ v(H_l) < v(HV_l) \quad (35)$$

Estimate the mean value weight of the selfish and sardine positions as described in Equation 36. Let's assume P - mean value, $m_{oH} -$ position of selfish, $Mo^H -$ position of sardine.

$$M = \frac{M_{oH} + Mo^H}{2} \quad (36)$$

The condition of swordfish is likely to change based on the current state of the prey fish, increasing the chances of finding new prey. Additionally, in the EESONN classifier, the average of the two levels varies as the weight value.

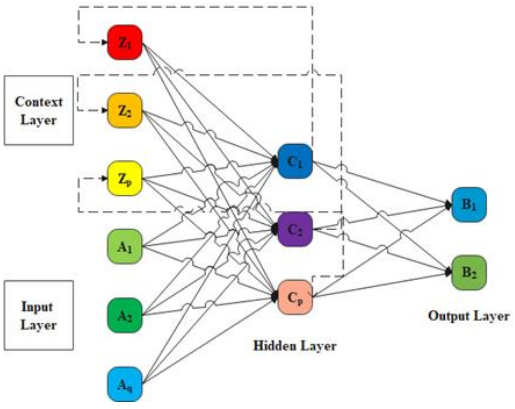

Fig 5. Elman Neural Network Structure

The Ellman Neural Network architecture includes four layers: input, output, hidden, and context layers, as illustrated in Figure 5. ENNs possess higher computational power than feed-forward neural networks.

Sardine schools avoid predation by changing their positions when one of their members is injured. Additionally, the population sets the maximum number of iterations and coefficient values. Figure 6 illustrates the flow chart for weight estimation during self-interest optimization.

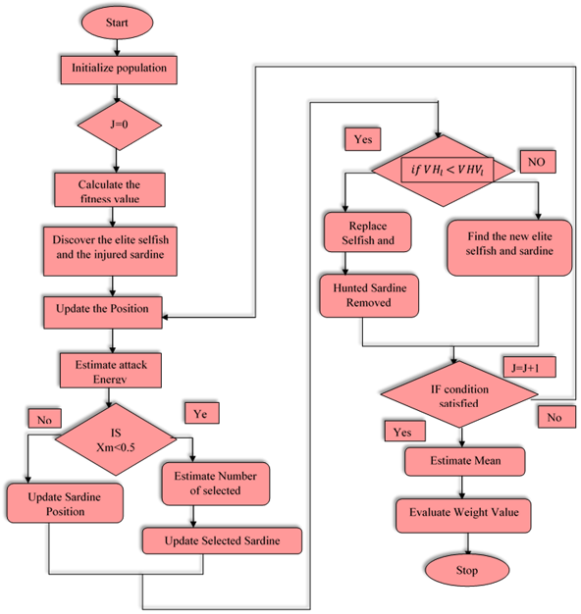

Fig 6. Selfish Optimization Structure

4. Result and Discussion

The EESONN method proposed in this study utilizes various performance measures to detect heart disease. It can detect CHD using a Python-based Jupiter notebook. With them, F1-Score, specificity, sensitivity, precision, and accuracy can be calculated to obtain an accurate estimate of heart disease risk. In addition, a confusion matrix can be used to measure the parameters of the heart disease classifier.

Table 1. Simulation Parameter

Simulation	Values
Name of the Dataset	Cleveland Clinic Heart Disease Dataset
Total number of datasets	304
Training	208
Testing	96
Tool	Jupiter
Language	Python

As mentioned in Table 1, they use a Python-based Jupiter Notebook to get the correct accuracy and detect CHD with the support of the dataset.

		Predicted Condition	
	Total Population = P+N	Positive (PP)	Negative (PN)
Actual Condition	Positive (P)	True Positive (TP)	False Negative (FN)
	Negative (N)	False Positive (FP)	True Negative (TN)

Fig 7. Confusion Matrix Diagram

A confusion matrix can not only show errors but also provide various metrics, as illustrated in Figure 7. Each metric has a specific significance and should be taken into account in certain situations.

$$Accuracy = \frac{\text{TrPos} + \text{TrNeg}}{TrPos + TrNeg + FalPos + FalNeg}$$

$$Sensitivity = \frac{\text{TrPos}}{TrPos + FalNeg}$$

$$Specificity = \frac{\text{TrNeg}}{TrNeg + FalPos}$$

$$Precision = \frac{\text{TrPos}}{TrPos + FalPos}$$

$$F1Score = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$$

Figure 8 displays a study of the heart disease prediction dataset's accuracy provided by Kaggle. In addition, a confusion matrix view displays errors and provides various metrics and significance levels for each measurement. Therefore, with the proposed EESONN method, sensitivity analysis can be conducted among these metrics to obtain the CHD predictor and accuracy. Compared to DNN, GS, and UCNN methods analysed from the literature with the proposed method, they are 43% lower. However, the EESONN method improved CHD prediction by 63%.

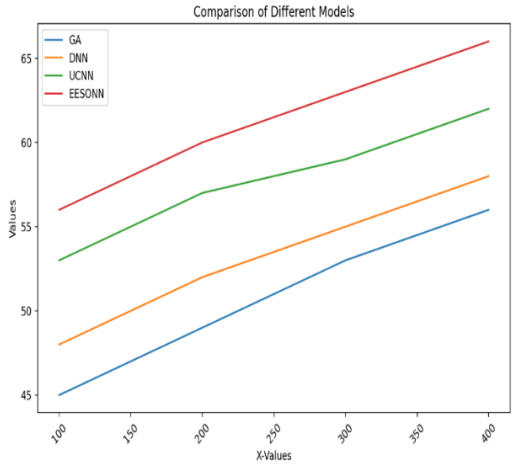

Fig 8. Analysis of Sensitivity

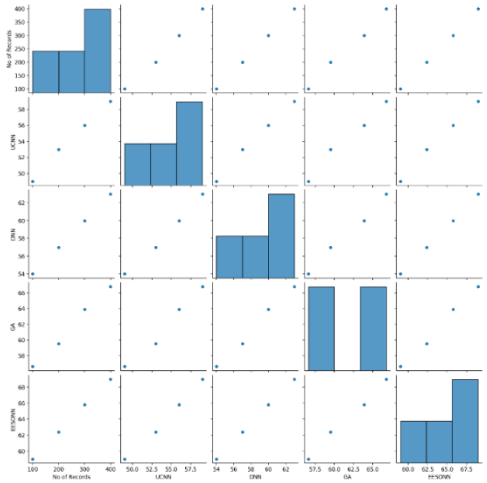

Fig 9. Analysis of Specificity

Figure 9 illustrates how Kaggle can be utilized to obtain the dataset for heart disease detection. Accuracy scores can also be obtained. The confusion matrix view not only shows errors but also provides various metrics as well as the specific meaning of each measurement. With the proposed EESONN method, CHD predictors can be detected with specific analysis and accuracy. Compared to literature analysis methods using GS, UCNN, and DNN, the proposed method is 49% lower. However, EESONN improves CHD prognosis by 69%.

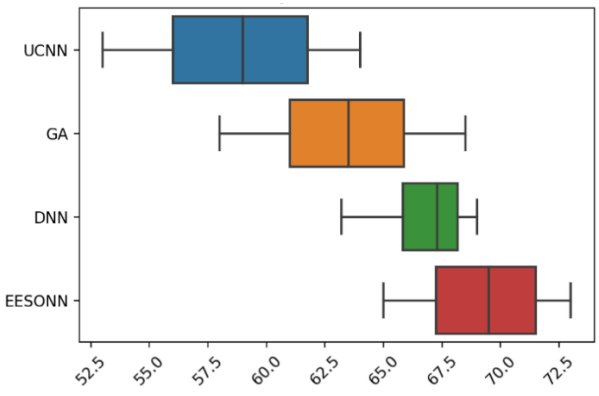

Fig 10. Analysis of Precision

The dataset used to forecast heart illness is sourced from Kaggle, as Figure 10 illustrates. The proposed EESONN method accurately identifies CHD prognostic factors through fine-grained analysis. Compared to literature analysis methods that use GS, UCNN, and DNN, EESONN improves CHD prediction by 73% despite having a lower success rate of 53%. With this method, a precision score can be obtained. The confusion matrix displays not only errors but also various metrics and the specific meaning of each metric.

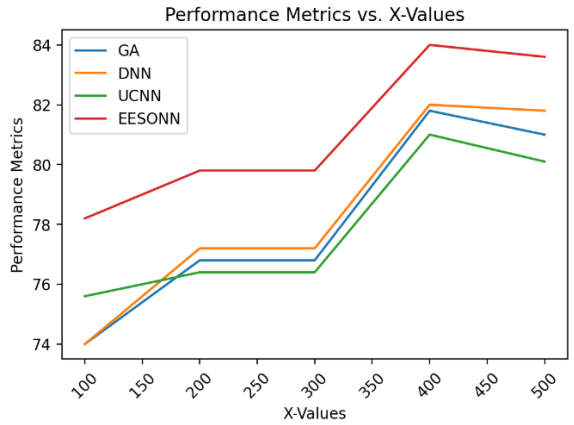

Fig 11. Analysis of F1-Score

Fig. 11 illustrates the use of F1-score performance analysis for predicting heart disease prognosis based on the Kaggle dataset. The accuracy score obtained from the dataset is accompanied by a confusion matrix display. This shows not only the error but also the performance impact of various metrics and the F1 score for each. CHD prediction can be verified by analysing the F1 score and the accuracy of the proposed EESONN method. EESONN showed an 83% improvement in CHD prediction compared to data analysis methods using DNN, UCNN, and GA. However, the success rate of this method is 79%.

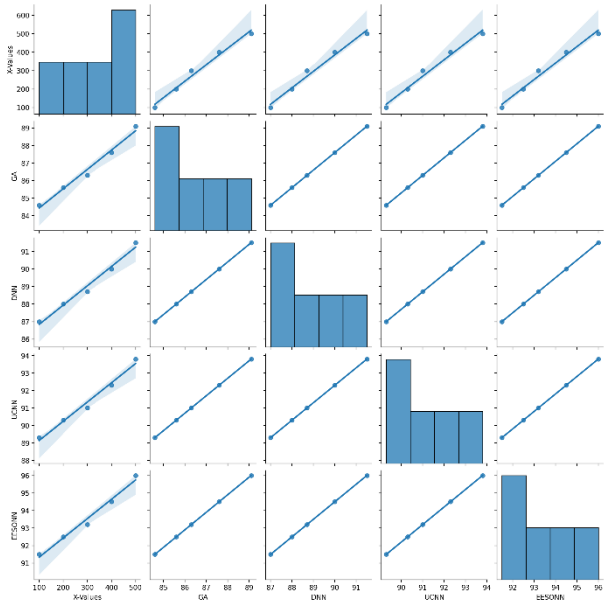

Fig 12. Analysis of Accuracy

To evaluate the performance index, we first estimated cardiovascular disease. This is accomplished by processing Kaggle data and analysing various methods based on the collected data. Figure 12 shows the EESONN method for predicting CHD. Accuracy scores obtained from the data can be compared with the confusion matrix. Additionally, when we tested the accuracy of methods such as DNN, UCNN, and GA based on data, it was 81%. However, when we tested the plan's accuracy, its score increased to 96%.

5. Conclusion

An effective method for predicting CHD is proposed, including optimal feature selection based on CSSDF and improved EESONN classification. Preprocessing, feature weighting, feature selection, and data classification are all part of the process in order to achieve accurate predictions. At the start of the project, the team collected data from the Cleveland Clinic Heart dataset through Kaggle. To improve data quality, they implemented the IDQFG method in data pre-processing to handle missing values. The team used the WRA method to select features based on their weights. This method helped them eliminate underweight symptoms and identify overweight symptoms. To further refine the features, the team combined the largest information coefficients and removed redundant features. Their final model was based on features selected using the CSSDF algorithm. The proposed EESONN method showed a significant improvement in prediction accuracy over the inefficient classifier, indicating good CHD detection performance. Based on the simulation results, the CSSDF-EESONN method outperforms other classifiers and advanced methods in terms of all performance metrics when tested on the datasets. The accuracy score is then compared to the confusion matrix for further evaluation. After testing, the proposed method had 96% accuracy. In the future, more developed methods and algorithms will be utilized to predict CHD in the short term.

Reference

[1] Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, and Parneet Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning," Volume 2021.

[2] Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam, Mohammad Hasan Imam, "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System," World Journal of Engineering and Technology > Vol.6 No.4, November 2018.

[3] Shah, D., Patel, S. & Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN COMPUT. SCI. 1, 345 2020.

[4] Santhosh Gupta Dogiparthi, Jayanthi K, and Ajith Ananthakrishna Pillai, "A Comprehensive Survey on Heart Disease Prediction Using Machine Intelligence," Research - International Journal of Medical Research & Health Sciences, Volume 10, Issue 7, 2021.

[5] Hassan CAU, Iqbal J, Irfan R, Hussain S, Algarni AD, Bukhari SSH, Alturki N, Ullah SS. Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers. Sensors (Basel). 22 (19):7227. 2022.

[6] M. Porumb, E. Iadanza, S. Massaro and L. Pecchia, "A convolutional neural network approach to detect congestive heart failure," Biomed—signal Process. Control, vol. 55, Jan. 2020.

[7] C. A. U. Hassan, J. Iqbal, R. Irfan, S. Hussain, A. D. Algarni, S. S. H. Bukhari, et al., "Effectively predicting the presence of coronary heart disease using machine learning classifiers", Sensors, vol. 22, no. 19, pp. 7227, Sep. 2022.

[8] P. Rani, R. Kumar, N. M. O. S. Ahmed, and A. Jain, "A decision support system for heart disease prediction based upon machine learning," J. Reliable Intell. Environments, vol. 7, no. 3, pp. 263-275, Sep. 2021.

[9] N. S. Mansur Huang, Z. Ibrahim, and N. Mat Diah, "Machine learning techniques for early heart failure prediction," Malaysian J. Comput. (MJoC), vol. 6, no. 2, pp. 872-884, 2021.

[10] Terrada O, Hamida S, Cherradi B, Raihani A, Bouattane O (2020) Supervised machine learning-based medical diagnosis support system for prediction of patients with heart disease. Adv Sci Technol Eng Syst J 5(5):269–277.

[11] T. Amarbayasgalan, V. Pham, N. Theera-Umpon, Y. Piao and K. H. Ryu, "An efficient prediction method for coronary heart disease risk based on two deep neural networks trained on well-ordered training datasets," IEEE Access, vol. 9, pp. 135210-135223, 2021.

[12] D. K. Plati, E. E. Tripoliti, A. Bechlioulis, A. Ramos, I. Dimou, L. Lakkas, et al., "A machine learning approach for chronic heart failure diagnosis", Diagnostics, vol. 11, no. 10, pp. 1863, Oct. 2021.

[13] S. Sarah, M. K. Gourisaria, S. Khare and H. Das, "Heart disease prediction using core machine learning techniques—A comparative study" in Advances in Data and Information Sciences, Springer, pp. 247-260, 2022.

[14] D. C. Yadav and S. Pal, "Prediction of heart disease using feature selection and random forest ensemble method," Int. J. Pharmaceutical Res., vol. 12, no. 4, pp. 56-66, 2020.

[15] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," Comput. Biol. Med., vol. 136, Sep. 2021.

[16] M. Gjoreski, A. Gradišek, B. Budna, M. Gams and G. Poglajen, "Machine Learning and End-to-End Deep Learning for the Detection of Chronic Heart Failure From Heart Sounds," IEEE Access, vol. 8, pp. 20313-20324, 2020.

[17] A. M. Qadri, A. Raza, K. Munir, and M. S. Almutairi, "Effective Feature Engineering Technique for Heart Disease Prediction With Machine Learning," in IEEE Access, vol. 11, pp. 56214-56224, 2023.

[18] R. Katarya and S. K. Meena, "Machine learning techniques for heart disease prediction: A comparative study and analysis," Health Technol., vol. 11, no. 1, pp. 87-97, Jan. 2021.

[19] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in IEEE Access, vol. 8, pp. 107562-107582, 2020.

[20] A. A. Almazroi, E. A. Aldhahri, S. Bashir and S. Ashfaq, "A Clinical Decision Support System for Heart Disease Prediction Using Deep Learning," in IEEE Access, vol. 11, pp. 61646-61659, 2023.

[21] T. Rahman, M. K. A. A. Al-Ruweidi, M. S. I. Sumon, R. Y. Kamal, M. E. H. Chowdhury and H. C. Yalcin, "Deep Learning Technique for Congenital Heart Disease Detection Using Stacking-Based CNN-LSTM Models From Fetal Echocardiogram: A Pilot Study," IEEE Access, vol. 11, pp. 110375-110390, 2023.

[22] A. Kumar, K. U. Singh and M. Kumar, "A Clinical Data Analysis Based Diagnostic Systems for Heart Disease Prediction Using Ensemble Method," Big Data Mining and Analytics, vol. 6, no. 4, pp. 513-525, 2023.

[23] Y. Pan, M. Fu, B. Cheng, X. Tao, and J. Guo, "Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform," IEEE Access, vol. 8, pp. 189503-189512, 2020, doi: 10.1109/ACCESS.2020.3026214.

[24] S. E. A. Ashri, M. M. El-Gayar and E. M. El-Daydamony, "HDPF: Heart Disease Prediction Framework Based on Hybrid Classifiers and Genetic Algorithm," IEEE Access, vol. 9, pp. 146797-146809, 2021.

[25] P. Ghosh et al., "Efficient Prediction of Cardiovascular Disease Using Machine Learning

Algorithms With Relief and LASSO Feature Selection Techniques," IEEE Access, vol. 9, pp. 19304-19326, 2021.

[26] C. Xiao, Y. Li and Y. Jiang, "Heart Coronary Artery Segmentation and Disease Risk Warning Based on a Deep Learning Algorithm," IEEE Access, vol. 8, pp. 140108-140121, 2020.

[27] A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim and A. W. Muzaffar, "An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases," IEEE Access, vol. 9, pp. 106575-106588, 2021.

[28] D. Padovano, A. Martinez-Rodrigo, J. M. Pastor, J. J. Rieta and R. Alcaraz, "On the Generalization of Sleep Apnea Detection Methods Based on Heart Rate Variability and Machine Learning," IEEE Access, vol. 10, pp. 92710-92725, 2022.

[29] J. Wang et al., "A Stacking-Based Model for Non-Invasive Detection of Coronary Heart Disease," IEEE Access, vol. 8, pp. 37124-37133, 2020.

[30] A. Abdellatif, H. Abdellatef, J. Kanesan, C. -O. Chow, J. H. Chuah and H. M. Gheni, "An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods," IEEE Access, vol. 10, pp. 79974-79985, 2022.

[31] H. R. H. Al-Absi, M. A. Refaee, A. U. Rehman, M. T. Islam, S. B. Belhaouari and T. Alam, "Risk Factors and Comorbidities Associated to Cardiovascular Disease in Qatar: A Machine Learning Based Case-Control Study," IEEE Access, vol. 9, pp. 29929-29941, 2021.

[32] Wiharto, E. Suryani, S. Setyawan and B. P. Putra, "The Cost-Based Feature Selection Model for Coronary Heart Disease Diagnosis System Using Deep Neural Network," IEEE Access, vol. 10, pp. 29687-29697, 2022.

[33] S. A. Ali et al., "An Optimally Configured and Improved Deep Belief Network (OCI-DBN) Approach for Heart Disease Prediction Based on Ruzzo–Tompa and Stacked Genetic Algorithm," IEEE Access, vol. 8, pp. 65947-65958, 2020.

[34] M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani and A. Shaikh, "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks," IEEE Access, vol. 11, pp. 121574-121591, 2023.

[35] A. Abdellatif, H. Abdellatef, J. Kanesan, C. -O. Chow, J. H. Chuah and H. M. Gheni, "Improving the Heart Disease Detection and Patients' Survival Using Supervised Infinite Feature Selection and Improved Weighted Random Forest," IEEE Access, vol. 10, pp. 67363-67372, 2022.