



Renyi Entropy Predictive Data Mining and Weighted Xavier Deep Neural Classifier For Heart Disease Prediction

M. Revathy Meenal^{1*}, S. Mary Vennila²

¹Research Scholar, Department of Computer Science, Presidency College (Autonomous), Chennai, Tamilnadu-600005
Corresponding Author Email: revathymeenal@yahoo.co.in - ORCID: 0000-0003-1992-4919

²Associate Professor & Head, Department of Computer Science, Presidency College (Autonomous), Chennai, Tamilnadu-600005

Email: vennilarhymend@yahoo.com - ORCID: 0000-0002-6453-1789

Article Info:

DOI: 10.22399/ijcesen.1000
Received : 25 November 2024
Accepted : 02 February 2025

Keywords :

Frequent Pattern Mining,
Renyi Entropy,
Homogenized,
Weighted Xavier,
Swish Activation Function.

Abstract:

During the past few years, Frequent Pattern Mining (FPM) has received the interest of several researchers that necessitate extracting items from transactions, and sequences from datasets, clarifying heart disease diagnosis that materializes commonly, and recognizing specific arrangements. In this era with healthcare involving significant evolutions, the unforeseeable movement and enormous amount of data concerning the classification of heart disease lead the way to new issues in FPM, such as space and time complexity. However, most of the research work concentrates on identifying the healthcare patterns relating to heart disease that transpires frequently, where the patterns within every transaction were known a priori. To address such issues in the present scenario, selecting the predominant patterns or frequent patterns is essential using relevant FPM models. The primary objective of this work is to enhance FPM mining results and reduce the misclassification rate of Cardiovascular Disease (CVD) dataset samples. This work proposes a novel method called Renyi Entropy Homogenized Weighted Xavier-based Deep Neural Classifier (REHWX-DNC) for heart disease prediction. To tackle the first challenge, the Renyi Entropy-based Frequent Pattern Mining (RE-FPM) algorithm is proposed, which filters the low-quality features using the Renyi Entropy function. To handle the second issue, the HWX-DNC model is designed to assist in minimizing the misclassification rate by employing the Swish activation function. A dataset for CVD synthesis can be analyzed to obtain significant accuracy for this study, and REGEX-DNC can be improved with compared state-of-the-art methods. Some indicators, including prediction accuracy, time, misclassification level, and F1-total, are considered to calculate the predictor, checking that the REHWX-DNC method proposed is efficient and trustworthy for predicting heart disease..

1. Introduction

Data mining involves discovering hidden patterns and database trends to create predictive models. Data mining integrates statistical analysis, Deep Learning (DL), and database mechanisms for extracting hidden patterns and associations or correlations from wide-ranging-sized databases. According to World Health Statistics 2012, one in three adults worldwide suffers from acute cardiovascular disease (CVD). Early detection of unrevealed patterns concerning heart disease would assist in providing treatments, therefore reducing mortality to a greater extent. It is illustrated in [1] that using a hybrid method of Random Forest (RF)-

based Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) called GAPSO-RF to select an optimal feature can improve the prediction accuracy of heart disease. It is shown that multivariate statistical analysis can identify the most essential characteristics of the initial population. Second, separate global and local searches were performed using GA and PSO. Finally, with the aid of the PSO that had been rejected in the selection process. Owing to this, accuracy was significantly attained, and the number of groups, minimizing the completing time in the method, was found to be relatively large due to the lack of efficient feature or pattern selection. To solve this problem, the RE-FPM algorithm was

designed to extract frequent patterns for classification processing with the help of the Renyi entropy function. This, in turn, improves the prediction accuracy and the prediction time involved in heart disease prediction. Diseases can be predicted and diagnosed with the help of the Internet of Things (IoT) by developing intelligent automated systems to dedicate significant advances in the field of DL. Thus, a novel proposed implementation of an Enhanced DL-Aided Convolutional Neural Network (EDCNN) method to diagnose cardiac disease in patients [2] is presented. Moreover, the EDCNN method was a concentrated framework learning technique. Also, the subsequent fine-tuning of parameters achieved a more significant amount of precision. As a result, the attribute selection mechanisms have contributed to the accuracy factor. Despite improvements in accuracy with the fine-tuned feature selection model, misclassification was involved during the DL process. HWX-DNC model proposes that the fine-tuning of weight using Homogenized Weighted Xavier in the deep neural classifier reduces the misclassification rate to a greater extent. Machine Learning (ML) techniques learn patterns from the prevailing dataset. Next, it applies the method to an unknown dataset for prediction. On the other hand, specific classification technique needs more accuracy. An ensemble classification technique that combines multiple classifiers to improve classification accuracy is proposed [3]. A comparative analysis of the role of ensemble techniques in enhancing coronary heart disease prediction accuracy was investigated [4]. Optimization techniques are used to evaluate the prevailing data for obtaining optimal predictive results. A heart disease predictive framework has been proposed based on primary risk criteria based on multiple classification algorithms such as Naive Bayesian, Bayesian Optimal Support Vector Machines, K Nearest Neighbors, and Salp Swarm Optimal Neural Networks [4].

Deep Neural Classifier is studied widely [5-9]. A protocol for CVD prediction can be found based on the comparative analysis of ML and DL algorithms [10].

Yet another ensemble classification method based on a feature selection approach was proposed [11]. Furthermore, a classification model employing ensemble learning with GA was presented for obtaining biomedical assessment prices to identify heart disease.

To solve the issues above based on the classification Algorithms, this paper proposes an efficient REHWX-DNC for heart disease prediction.

1.1 Contributions to the work

- A novel DL method called REHWX-DNC for heart disease prediction is designed, with which prompt medication can be given to the patient accurately.
- To propose a RE-FPM algorithm for the CVD dataset to obtain the frequent patterns necessary for classification accurately and timely.
- To design HWX-DNC model for reducing misclassification rate and therefore improving F1-score.
- A heart disease prediction is conducted using the JAVA programming language to support our claims.
- Finally, we present a detailed comparative analysis between the proposed REHWX-DNC and existing methods. This analysis shows that our suggested REHWX-DNC method is more efficient and significant than the other methods regarding prediction accuracy, prediction time, misclassification rate, and F1 score.

1.2 Paper organization

The article sequentially organizes relevant studies on cardiac or CVD diagnosis and prognostic processes mentioned in Section 2, comprehensively reviewing them. Section 3 of the REHWX-DNC is discussed. Section 4 illustrates the experiments, and a detailed discussion with a table and graph is given in Section 5. Finally, Section 6 represents the main conclusion.

2. Related works

Cardiovascular illness is said to be one of the leading causes of passing away in several countries. The disease is said to be diagnosed by physicians based on prevailing clinical tests and experiences facing similar symptoms. Also, patients suffering from heart disease have to be attended to promptly. A hybrid framework for heart disease finding based on medical voice records was presented in [12]. The hybrid framework included four layers. In the first layer, segmentation was suggested, followed by feature extraction was performed in the second layer. Next, in the third layer, the most pertinent configurations were selected based on the optimization technique, and finally, validation was performed in the fourth layer. With this hybrid framework, accuracy was said to be improved. However, the execution time could have been more focused. To concentrate on the time factor issue, strength scores for each feature were evaluated, and with this score, significant predictors were

identified [13]. The identified significant predictors were then utilized for predicting heart disease, with the period involved in heart disease prediction improved. The main goal is to improve feature selection and reduce feature count [14]. An empirical competition algorithm and metaheuristic technique are proposed to select the key features. As an outcome, the misclassification rate was said to be reduced. A systematic review of ML-based techniques employing distinct data modalities for heart disease prediction was investigated [15]. A novel mechanism to identify pertinent features via ML techniques was proposed that, in turn, caused the development of accuracy for CVD prediction [16]. An effective and precise mechanism to diagnose heart disease based on ML techniques was proposed [17]. Here, an attribute selection algorithm based on fast conditional mutual information was designed, followed by hyperparameter tuning, therefore achieving good accuracy. Yet another method employing nine classifiers was utilized to improve the accuracy rate of heart disease prediction [18]. Multiple data processing techniques were employed for handling imbalanced data issues during classification [19]. However, optimal feature selection remains a significant aspect that should have been more focused on during the category. A hybrid GA and PSO optimized technique using RF, referred to as GAPSO-RF, was proposed to address this aspect [20]. Here, optimal features were selected with the hybrid model, with which the heart disease prediction accuracy rate was also said to be improved. An exploration of DL techniques for CVD prediction was presented [21]. Complicated interactions between risk factors were analyzed for CVD prediction [22]. With this unnecessary treatment, other diseases were prevented to a greater extent. Chi-square tests were performed [23] to concentrate on the significant feature selection using the RF algorithm. A complete review of data mining and its presentation toward heart disease forecast was examined [24]. From the analysis of the above work, the research on heart disease prediction using DL has continued. In contrast, the study on heart disease prediction employing heart disease prediction has been moderately enhanced and improved. This REHWX-DNC for heart disease prediction is a novel DL technique based on previous ones that improve prediction accuracy and time and reduce the misclassification rate for heart disease datasets.

3. Materials and methods

Classification is pivotal in all domain areas of disease diagnosis, disease prediction, education,

and business. FPM is a frequently-used data mining function that can identify frequent co-occurrences in cardiovascular disease prediction, making it one of the most challenging problems. Nonetheless, involvements of infrequent patterns in heart disease prediction aren't preferably improving the accuracy owing to their confined predictive potentiality. This work presents a method called REHWX-DNC for heart disease estimates with minimum misclassification ratio and progress accuracy. The flow graph of REHWX-DNC is given in Figure 1. The figure 1 shows that the proposed REHWX-DNC method for heart disease prediction is split into two sections. First, frequent patterns are mined using the RE-FPM model. Second, heart disease prediction is performed by employing HWX-DNC with the obtained recurring patterns. The following segments offer a detailed description of the REHWX-DNC method, beginning with a dataset description.

3.1 Description of the dataset

The CVD dataset from the Kaggle was used for the trials [25]. The dataset contains 12 features or attributes and 70000 instances. They include three types of input aspects, which are objective (factual information), experimental (physical examination), and subjective (patient-provided information), respectively. The details of the data set are shown in Table 1 which shows that each patient's characteristics, age, height, systolic blood pressure (top number), and diastolic blood pressure (bottom number) are expressed as integers. Next, weight forms the float value, and gender represents male or female in categorical representation. Third, three types of cholesterol and glucose can be considered for heart disease prediction, with 1 representing normal, two representing above average, and three denoting well above normal, respectively. Fourth, binary representation of 0 or 1 representing the practice of smoking, intake of alcohol, and performing physical activity are employed where 0 means no and one represents yet. Finally, a binary target variable is created for the presence or absence of the disease.

3.2 Renyi Entropy-based Frequent Pattern Mining

A frequent pattern is a pattern that is supposed to occur frequently in a given dataset. Using routine pattern identification, strongly associated items are observed, therefore assisting in analyzing similar characteristics between them. Also, FPM is said to be employed in various fields of study like surveillance, disease diagnosis, e-commerce, etc.

Over the past few years, FPM has been essential in cardiovascular disease prognosis. By identifying frequent patterns in the dataset, prediction accuracy and time are likely improved. Figure 2 shows the structure of the RE-FPM model. As shown in the figure 2 a dataset called the CVD dataset from Kaggle is first obtained as input. The patient records are initially obtained and stored in a matrix for further processing. Next, to evaluate the pattern frequency, the Renyi Entropy function is applied to the patient records, and based on the probability of occurrence of the pattern, weighted support is formulated. Third, with the initialized user-defined threshold support, minimum weighted approval is obtained with which the frequency patterns are evolved in a computationally efficient manner. Let us consider the CVD dataset 'DS' and 'P = {P₁, P₂, ..., P_m}' denote the set of 'm' definite patterns and 'C = {C₁, C₂, ..., C_l}' denotes the set of 'l' classes. Also, let us assume that 'DS' comprises 'n' instances such that 'DS = {A_i, B_i}' where 'A_i ⊆ P' and 'B_i ⊆ C.' In our work, the set of 'm' definite patterns includes the 11 distinct features mathematically stated as given below.

$$PR = PatR = \begin{bmatrix} P_1F_1 & P_1F_2 & \dots & P_1F_n \\ P_2F_1 & P_2F_2 & \dots & P_2F_n \\ \dots & \dots & \dots & \dots \\ P_mF_1 & P_mF_2 & \dots & P_mF_n \end{bmatrix} \quad (1)$$

From the above equation (1), patient records 'PatR' are provided where the first row specifies the pattern and features of the first patient. Similarly, 'm' patterns and features of the 'nth' patients are obtained from the CVD dataset 'DS.' Similarly, 'SS = {SS₁, SS₂, ..., SS_k}' represents the subset belonging to pattern set 'P.' Given the CVD dataset 'DS,' the set that has 'SS' is represented as 'DS_{SS} = {(A_i, B_i) | SS ⊆ A_i}', and 'SS' is said to be frequent if 'α = $\frac{DS_{SS}}{DS} \geq \alpha_0$ ', where 'α' is referred to as the relative support of 'SS' and 'SS₀' denotes the minimum support 'min_{sup}' threshold and the frequent pattern sets are represented as 'FP' A smaller value of biased estimations, in addition to the Renyi Entropy function, is utilized in evaluating the frequency of the pattern. The entropy indicates the goodness of a dynamic collection of sample sets and is mathematically considered as given below.

$$H_\beta(A) = \frac{1}{1-\beta} \log \left(\sum_{i=1}^n Prob_i^\beta \right) \quad (2)$$

From the above equation (2), 'A' refers to the dynamic collection of sample sets with probable

feature outcomes in the set 'P = {PR₁, PR₂, ..., PR_m}' and the respective probabilities 'p_i = Prob(A = PR_i)'. If the probabilities are 'p_i = $\frac{1}{n}$, for all i = 1, 2, ..., n', it is called a weighted pattern set. Then, if the probability of occurrence of the pattern is more frequent, then the weighted pattern set is high, or else it is low. Then, for a design with patient records, the weighted support is mathematically formulated below.

$$WSUP(PR) = SUP(PR) * W(PR) \quad (3)$$

From the above equation (3), the weighted support 'WSUP(PR)' is measured based on the support of the pattern with patient records 'SUP(PR)' and the weight 'W(PR)' of the pattern with patient records, respectively. This is followed by the weight for each pattern with patient records using 'n' differences features, mathematically formulated as given below.

$$W(PR_{kml}) = \frac{\sum_{k=1}^F W(PR_{kml})}{WSUP(PR)} \quad (4)$$

Next, with the above weight 'W(PR_{kml})' as given in equation (4), minimum weighted support is mathematically stated as given below.

$$MINSUP(PR_{kml}) = \begin{cases} W(PR_{kml}) * H_\beta(A), & \text{when } SUP(PR) > TSUP \\ H_\beta(A), & \text{Otherwise} \end{cases} \quad (5)$$

Finally, with the above minimum weighted support resultant value 'MINSUP(PR_{kml})', the frequent patterns are obtained as given below.

$$FP = WSUP(PR_{kml}) \geq \min \left[\begin{matrix} MINSUP(p_{1ml}), MINSUP(p_{2ml}), \\ \dots, MINSUP(p_{kml}) \end{matrix} \right] \quad (6)$$

Based on the above minimum weighted support 'WSUP(PR_{kml})' for each patient record, computationally efficient frequent patterns (i.e., features) are obtained. The pseudo-code representation of RE-FPM is given below.

Input: Dataset 'DS,' Patterns 'P = {P ₁ , P ₂ , ..., P _m }'
Output: Computationally efficient frequent patterns 'FP' (i.e., features)
1: Initialize set of 'l' classes, set of 'm = 11' definite patterns, instances 'n,' subset 'k.'
2: Initialize user-defined threshold support 'TSUP'
3: Begin
4: For each Dataset 'DS' with Patterns 'P'

5: Direct patient records 'PatR' as given in equation (1)
 6: Formulate the Renyi Entropy function as given in equation (2)
 7: Calculate weighted support as given in equation (3)
 8: Evaluate the weight for each pattern with patient records as given in equation (4)
 9: Estimate minimum weighted support as given in equation (5)
 10: **Return** frequent patterns 'FP' as given in equation (6)
 11: **End for**
 12: **End**

Algorithm 1. Renyi Entropy-based Frequent Pattern Mining (frequent patterns with examination features, i.e., ID number, systolic blood pressure, diastolic blood pressure, cholesterol, and glucose)

As given in the above algorithm, for each CVD dataset provided as input, the objective remains in designing computationally efficient features (i.e., frequent patterns) using the Renyi Entropy function. The advantage of applying the Renyi Entropy function is that the probable feature outcomes were said to preserve the additivity for independent events (i.e., independent feature outcomes). This, in turn, improves the time and the accuracy at which the heart disease prediction is said to be made.

3.3 Homogenized Weighted Xavier-based Deep Neural Classifier

With the resultant frequent pattern mining in the above section, classification has to be made for significant heart disease prediction. Researchers have also designed numerous predictive methods. In this section, a HWX-DNC model is developed. Figure 3 shows the structure of the HWX-DNC model. As shown in the figure 3, the HWX-DNC model utilizes a feed-forward neural network, with the basic structure split into three layers. They are one input layer, five hidden layers (i.e., 'n=5000' in the first hidden layer, 'n=4000' in the second hidden layer, 'n=3000' in the third hidden layer, 'n=2000' in the fourth hidden layer, and 'n=1000' in the fifth hidden layer) and one output layer respectively. Input features (i.e., five frequent patterns) in the input layer are then extracted by the hidden layer, and the classification results are obtained in the output layer. As illustrated in the figure 4, we input the frequent patterns selected in the FPM model into the HWX-DNC model. The HWX-DNC model has seven hidden layers and finally obtains two outputs, according to the Swish Scores interval belonging to each class. The input layer and hidden layer of our HWX-DNC utilize the softmax function as given below.

$$\sigma (FP)_i = \frac{\exp^{FP_i}}{\sum_{j=1}^n \exp^{FP_j}} \quad (7)$$

From the above equation (7), a standard exponential function 'exp' is applied to each frequent pattern 'FP_i' of the input vector 'FP' that, in turn, normalizes the recurring patterns by dividing by the sum of all these exponentials. Our work utilizes a Weight initializer by employing the Homogenized Weighted Xavier model. This model aims to update the parameters in a feed-forward neural network using a backpropagation process, which is a complex process. Therefore, a weight initialization using the Homogenized Weighted Xavier model is employed that not only avoids saturation of activation values in each layer but also prevents the activation value of each layer from being zero. As a result, the misclassification rate can be reduced to a greater extent. The Homogenized Weighted Xavier is mathematically stated as given below.

$$HW = U \left[- \left(\frac{SQRT(6)}{SQRT(n+m)}, \frac{SQRT(6)}{SQRT(n+m)} \right) \right] \quad (8)$$

From the above equation (8), the Homogenized Weighted Xavier function 'HW' is calculated with a uniform probability distribution 'U' with 'n' representing the number of inputs (i.e., frequent patterns) to the node (i.e., regular patterns in the previous layer) and 'm' denoting the number of outputs from the layer (i.e., number of frequent patterns in the current layer). Finally, the swish function is utilized in the output layer for mapping the output to the range of '(0,1)'. Moreover, selecting the swish scores in the output layer makes the output easier to control. This is mathematically stated as given below.

$$SWISH(FP) = FP \text{ Sigmoid} (\alpha FP) = \frac{FP}{1 + \exp^{-\alpha FP}} \quad (9)$$

From the above equation (9), 'α' refers to the trainable parameter ranging between '0 and 1'. The output result obtained by swish scores is the probability of belonging to the corresponding class. The pseudo-code representation of Homogenized Weighted Xavier-based Deep Neural Classification (HWX-DNC) for heart disease prediction is below.

Input: Dataset 'DS,' Patterns 'P = {P ₁ , P ₂ , ..., P _m }'
Output: Precise heart disease prediction
1: Initialize frequent patterns 'FP'
2: Begin
3: For Dataset 'DS' with frequent patterns 'FP'
4: Formulate softmax function as given in equation (7) //input layer
5: Formulate Homogenized Weighted Xavier function as given in equation (8) //weight updates
6: Formulate swish scores as given in equation (9) //output layer
7: If 'SWISH(FP) ≥ 1'
8: Then disease predicted
9: End if
10: If 'SWISH(FP) < 1'

```

11: Then no disease predicted
12: End if
13: End for
14: End
    
```

Algorithm 2. Homogenized Weighted Xavier-based Deep Neural Classification for heart disease prediction

As assumed in the algorithm, a Deep Neural Network (DNN) method was implemented using a uniformly weighted Xavier function. These can reduce the misclassification rate in the heart disease prognostic process. First, in the input layer, the frequent patterns obtained are provided as input employing the softmax function. Second, with five hidden layers designed, each the number of hidden neurons ‘n’ keeps decreasing in succeeding layers to get closer to the pattern for identifying the objective class (i.e., existence or nonexistence of cardiovascular disease). This is performed by employing the Homogenized Weighted Xavier function. Finally, it helps reduce the misclassification rate with the advantage of multi-layer neuron stacking, and the final target variable is derived from the swish score in the output layer.

4. Experimental setup

This Section Presents A Survey And Assessment Of The Existing State-Of-The-Art Techniques To Analyze Our Proposed REHWX-DNC For Heart Disease Prediction. By Analyzing The Prediction Time, Accuracy, Misclassification Rate, And F1 Score, The Test Evaluation Uses Java With Three Input Features: Objective, Exploratory, And Subjective.

5. Discussion

5.1 Case scenario 1: Prediction accuracy

First, prediction accuracy involved in heart disease using the data mining concept is provided in this section. With the CVD dataset supplied as input and patterns obtained through eleven distinct features, the prediction accuracy is measured as follows.

$$PA_{HD} = \left(\frac{Pd_i \text{ correctly classified}}{n} \right) * 100 \quad (10)$$

From the above equation (10), the prediction accuracy for heart disease ‘PA_{HD}’ is estimated based on the percentage ratio of patient data correctly classified ‘Pd_i correctly classified’ and the number of patterns ‘n’ involved in the simulation process. It is measured in terms of percentage (%). To obtain a more suitable method for the method, some experiments were conducted to test the prediction performance of the REHWX-

DNC method and the prediction performance of GAPSO-RF [1] and EDCNN [2], and some results are shown in Table 2. The REHWX - DNC method outperforms [1] and [2], and the benchmark chart is shown in Figure 4, which compares the methods in detail.

Table 1. Attribute information of the cardiovascular disease dataset

S. No	Feature of attribute	Type of feature	
1	Age	Objective feature	Integer
2	Height	Objective feature	Integer
3	Weight	Objective feature	Float
4	Gender	Objective feature	Categorical code
5	Systolic blood pressure	Examination feature	Integer
6	Diastolic blood pressure	Examination feature	Integer
7	Cholesterol	Examination feature	1: Normal, 2: Above normal, 3: Well above normal
8	Glucose	Examination feature	1: Normal, 2: Above normal, 3: Well above normal
9	Smoking	Subjective feature	Binary
10	Alcohol intake	Subjective feature	Binary
11	Physical activity	Subjective feature	Binary
12	Presence or absence of CVD	Target variable	Binary

Table 2. Performance of prediction accuracy for heart disease prediction

Number of patterns	Prediction accuracy (%)		
	REHWX-DNC	GAPSO-RF	EDCNN
500	98	95	92
1000	96.35	92.15	90.35
1500	95	91	88.15
2000	94.65	90.85	85.35
2500	94.15	90.35	82.15
3000	93.25	89.15	82
3500	93	89	81.15
4000	91.55	87.35	78.35
4500	89.25	86	75
5000	87	85.25	72

5.2 Case scenario 2: Prediction time

It is noted that extracting patterns and obtaining significant customs takes a little time during heart disease forecasting. These are called prediction times, and heart prediction times are shown mathematically.

$$PT = n * t \text{ (classifying one pattern)} \tag{11}$$

From the above calculation (11), the prediction time ‘PT’ involved in heart disease prediction is obtained based on the number of patterns ‘n’ and the time consumed in getting frequent patterns for further classification ‘t (classifying one pattern).’ Table 3 presents the table result analysis of heart disease prediction time for three methods, REHWX-DNC, GAPSO-RF [1], and EDCNN [2], and can be measured in milliseconds (ms).

Table 3. Performance of prediction time for heart disease prediction

Number of patterns	Prediction time (ms)		
	REHWX-DNC	GAPSO-RF	EDCNN
500	125	150	225
1000	145	175	240
1500	160	225	295
2000	175	240	335
2500	190	275	350
3000	215	280	385
3500	235	315	415
4000	250	335	435
4500	295	350	485
5000	315	385	525

In Figure 5, heart disease prediction time performances were compared for distinct patterns ranging between 500 and 5000. The proposed REHWX-DNC method shows better heart disease prediction time and consistent decline than [1] and [2]. Furthermore, the proposed REHWX-DNC method takes less time to predict heart disease in model data samples with three feature types: objective, exploratory, and subjective. Besides, we can realize from the figure that the proposed method exhibits higher stability as the sample grows. However, the possibility of predicting heart disease [1] and [2] is essential in analysis. Therefore, the REHWX-DNC method works with the original pattern as input by frequently expressing the sample patterns. For this reason, heart disease prediction time is significantly reduced as the number of model patterns increases. This is because, as mentioned earlier, based on the

RE-FPM algorithm, probable feature outcomes were first said to preserve the additivity for independent events via the application of the Renyi Entropy function. The REHWX-DNC method reduced heart disease prediction time by 23% compared to [1] and 43% compared to [2], respectively, while creating better frequent patterns in less time.

5.3 Case scenario 3: Misclassification rate

A misclassification rate is a performance measure representing the percentage of incorrect predictions without distinguishing between positive and negative predictions. Furthermore, the rate of misclassification can be calculated using the following method.

$$MR = \frac{Number_{IP}}{TP} * 100 \tag{12}$$

Based on the calculation (12) above, the misclassification rate “MR” is measured as the ratio of the number of incorrect predictions “ [(Number) _IP” to the total forecast “TB.” This is calculated as a percentage (%). Table 4 lists the misclassification rates for the three different methods (REHWX-DNC, GAPSO-RF [1] and EDCNN [2]) using the above equation (12).

Table 4. Misclassification rate using REHWX-DNC, GAPSO-RF [1] and EDCNN [2]

Number of patterns	Misclassification rate (%)		
	REHWX-DNC	GAPSO-RF	EDCNN
500	6	8	10
1000	6.35	8.55	10.34
1500	7	8.95	11
2000	7.25	9.15	11.35
2500	7.55	9.45	11.55
3000	8.15	9.85	12
3500	8.45	10.15	12.15
4000	8.95	10.35	12.45
4500	9.35	10.85	12.55
5000	9.85	11.35	13

In Figure 6, there is a visual display of misclassification rates that were obtained through the use of three distinct methods. From the above figure, the horizontal axis represents the number of patterns provided as input, ranging between 500 and 5000. On the other hand, the vertical axis represents the misclassification rate arrived at using the formula mentioned above. Also, the misclassification rate was found to increase with the increase in frequent patterns. The reason was

due to the large number of patterns and increases in the total predictions; the predictions to be made also see a downward trend, increasing a significant number of incorrect predictions. However, with 500 frequent patterns and a complete forecast of 50, the inaccurate predictions using the three methods, REHWX-DNC, GAPSO-RF [1], and EDCNN [2], were found to be 3, 4, and 5. With this, the misclassification rate was observed to be 6%, 8%, and 10%, respectively. The significant decrease in misclassification rate using the REHWX-DNC method was due to incorporating the Homogenized Weighted Xavier function in the DL process. With this function, not only is the saturation of activation values avoided in each layer but also the activation value of each layer is zero. From this, the REHWX-DNC method's misclassification rate is reduced by 19% compared to [1] and 33% compared to [2].

5.4 Case scenario 4: F1-score

Finally, it is evaluated based on the accuracy and recall of the test. Furthermore, the F1 score indicates the degree of test accuracy. Here, precision refers to the number of accurate positive results (i.e., heart disease patients identified with as it is) divided by the number of all positive outcomes (i.e., diseased patients). The recall measure is calculated by dividing the number of correctly identified positive results by the total number of favorable results in the sample.

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

Finally, the F1-score is described in Table 5 with numbers of patterns in the range of 500 to 5000, and comparisons are made between three methods, REHWX-DNC, GAPSO-RF [1], and EDCNN [2]. Finally, the graphical representation of the F1 score can be introduced in REHWX-DNC, GAPSO-RF [1], and EDCNN [2] using three different methods, as shown in Figure 7 above. Let us assume that the overall frequent patterns are 30, with regular patients being 24 and diseased patients being 6. Then, true positives, false negatives, and false positives were detected as '5', '1', and '4', respectively, using REHWX-DNC. Similarly, the true positive, false negative, and false positive using GAPSO-RF [1] were found to be '4', '3' and '4'; using EDCNN [2], it was found to be '3', '3' and '6', then, the F1-Score using the three methods were measured. Comparative analysis from the figure 7 showed better results using the REHWX-DNC method than the advanced methods. The purpose behind the progress was the optimal selected frequent patterns performed with the use of HWX-

DNC. With this, the F1-measure was found to be comparatively better using the REHWX-DNC method by 14% and 34% upon comparison with the two methods, [1] and [2].

Table 5 F1-score using REHWX-DNC, GAPSO-RF [1] and EDCNN [2]

Number of patterns	F1-score		
	REHWX-DNC	GAPSO-RF	EDCNN
500	0.661	0.528	0.39
1000	0.715	0.635	0.455
1500	0.735	0.655	0.535
2000	0.785	0.685	0.585
2500	0.825	0.715	0.625
3000	0.8	0.7	0.6
3500	0.755	0.685	0.635
4000	0.785	0.695	0.655
4500	0.825	0.725	0.685
5000	0.855	0.775	0.715

6. Conclusion

With the swift evolution of FPM and DL techniques, researchers started to focus on integrating and applying these two in several fields. In this paper, we use the RE-FPM model for obtaining frequent patterns. Next, we propose a HWX-DNC model that not only performs to get more and more close to the pattern for identifying the target class but also stacks multiple layers of neurons and is hence capable of predicting heart disease. From the performance analysis, it is known that the proposed REHWX-DNC NN method has better performance than the conventional methods, such as GA and PSO based on RF, and Improved DL assisted CNN in terms of prediction accuracy, prediction time, misclassification rate, and F1-score.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.

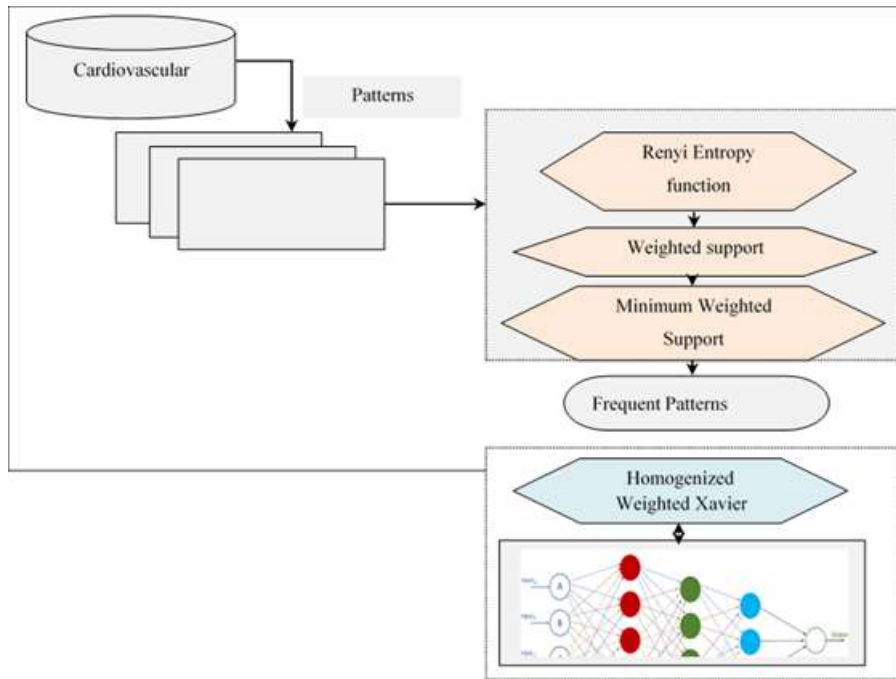


Figure 1. Block diagram of Renyi Entropy Homogenized Weighted Xavier-based Deep Neural Classifier

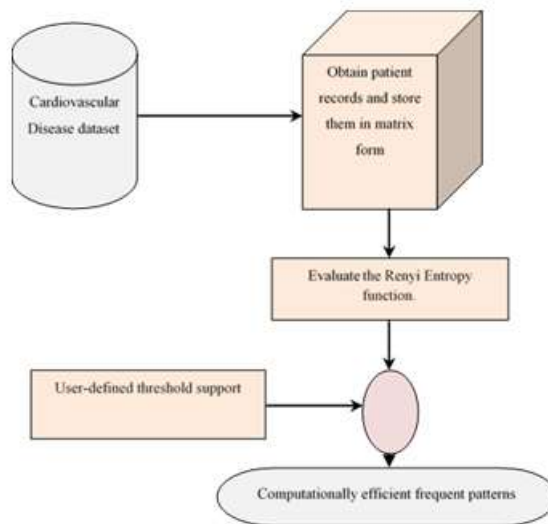


Figure 2. Structure of Renyi Entropy-based Frequent Pattern Mining

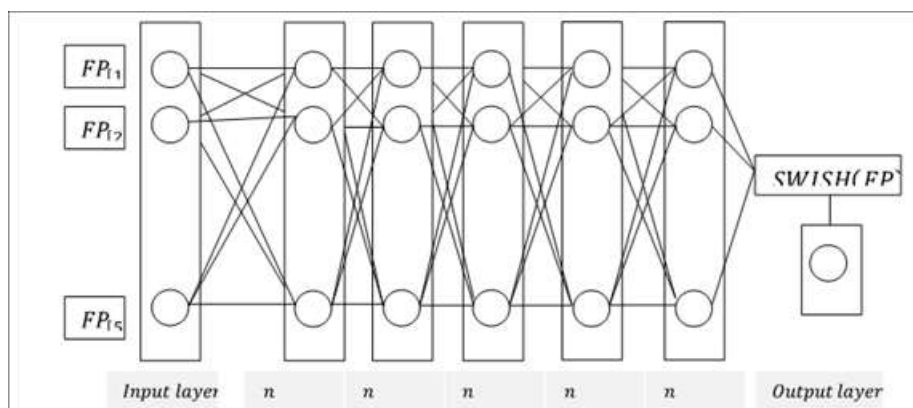


Figure 3. Structure of Homogenized Weighted Xavier-based Deep Neural Classifier model

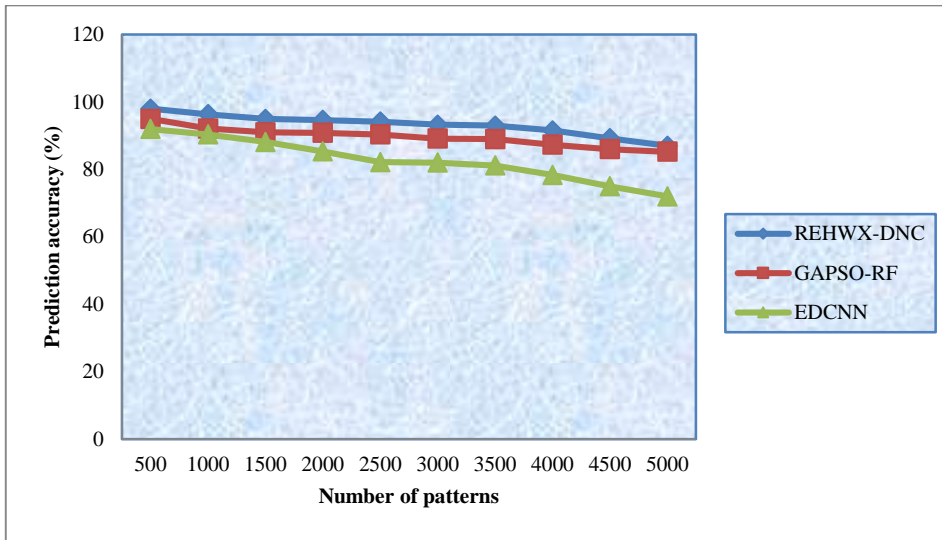


Figure 4. Performance analysis of prediction accuracy

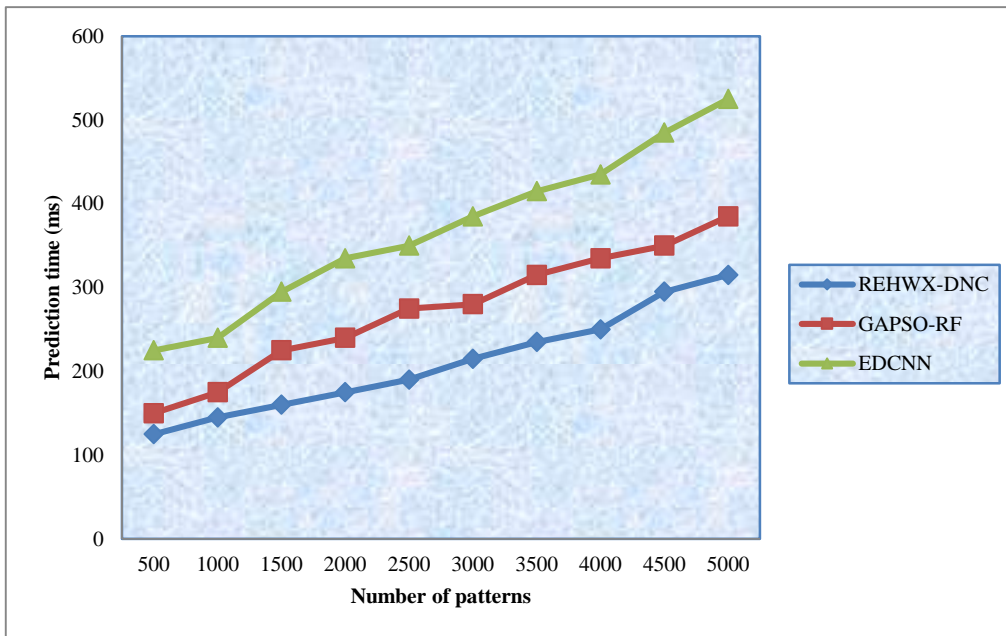


Figure 5. Performance analysis of prediction time

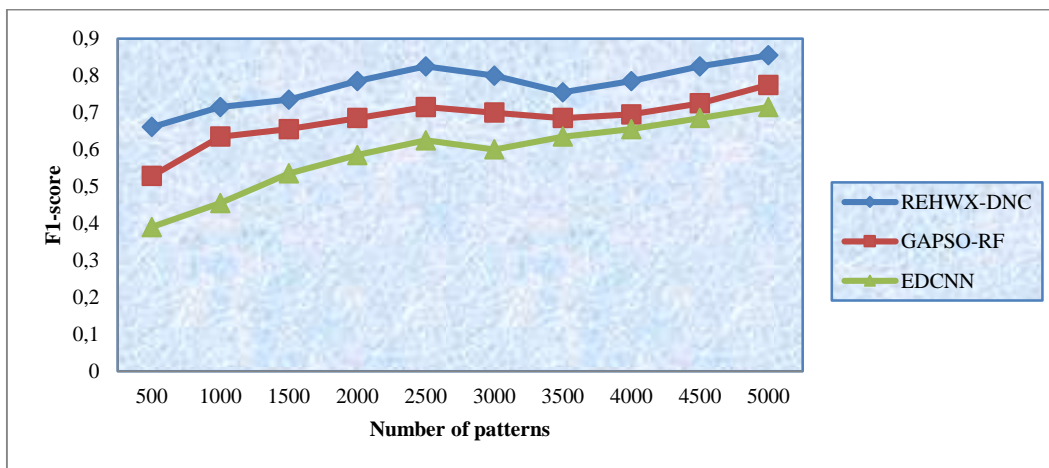


Figure 7. Performance analyses of F1-score

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

- [1] Mohamed G. El-Shafiey, Ahmed Hagag, El-Sayed A. El-Dahshan, Manal A. Ismail, (2022). A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest, *Multimedia Tools and Applications*, 81;18155–18179. <https://doi.org/10.1007/s11042-022-12425-x>
- [2] Y. Pan, M. Fu, B. Cheng, X. Tao and J. Guo, (2020). Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform, in *IEEE Access*, 8;189503-189512, doi: 10.1109/ACCESS.2020.3026214
- [3] C. Beulah Christalin Latha, S. Carolin Jeeva, (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques, *Informatics in Medicine Unlocked*, <https://doi.org/10.1016/j.imu.2019.100203>
- [4] Vardhan Shorewala, (2021). Early detection of coronary heart disease using ensemble techniques, *Informatics in Medicine Unlocked*, 26;100655 <https://doi.org/10.1016/j.imu.2021.100655>
- [5] LAVUDIYA, N. S., & C.V.P.R Prasad. (2024). Enhancing Ophthalmological Diagnoses: An Adaptive Ensemble Learning Approach Using Fundus and OCT Imaging. *International Journal of Computational and Experimental Science and Engineering*, 10(4). <https://doi.org/10.22399/ijcesen.678>
- [6] Nawaz, S. M., K. Maharajan, N. N. Jose, & R.V.S. Praveen. (2025). GreenGuard CNN-Enhanced Paddy Leaf Detection for Crop Health Monitoring. *International Journal of Computational and Experimental Science and Engineering*, 11(1). <https://doi.org/10.22399/ijcesen.1027>
- [7] Jha, K., Sumit Srivastava, & Aruna Jain. (2024). A Novel Texture based Approach for Facial Liveness Detection and Authentication using Deep Learning Classifier. *International Journal of Computational and Experimental Science and Engineering*, 10(3). <https://doi.org/10.22399/ijcesen.369>
- [8] T. Deepa, & Ch. D. V Subba Rao. (2025). Brain Glial Cell Tumor Classification through Ensemble Deep Learning with APCGAN Augmentation. *International Journal of Computational and Experimental Science and Engineering*, 11(1). <https://doi.org/10.22399/ijcesen.803>
- [9] Johnsymol Joy, & Mercy Paul Selvan. (2025). An efficient hybrid Deep Learning-Machine Learning method for diagnosing neurodegenerative disorders. *International Journal of Computational and Experimental Science and Engineering*, 11(1). <https://doi.org/10.22399/ijcesen.701>
- [10] Swathy, M., & Saruladha, K. (2021). A comparative study of classification and prediction of Cardiovascular Diseases (CVD) using Machine Learning and Deep Learning techniques. *ICT Express*, 8, 109-116. <https://doi.org/10.1016/j.ict.2021.08.021>
- [11] Abdollahi, J., Nouri-Moghaddam, B. A hybrid method for heart disease diagnosis utilizing feature selection based ensemble classifier model generation. *Iran J Comput Sci* 5, 229–246. <https://doi.org/10.1007/s42044-022-00104-x>
- [12] Balaha, H.M., Shaban, A.O., El-Gendy, E.M., & Saafan, M.M. (2022). A multi-variate heart disease optimization and recognition framework. *Neural Computing and Applications*, 34, 15907 - 15944. DOI:10.1007/s00521-022-07241-1
- [13] Yazdani, A., Varathan, K.D., Chiam, Y.K. et al. (2021). A novel approach for heart disease prediction using strength scores with significant predictors. *BMC Med Inform Decis Mak* 21;194 <https://doi.org/10.1186/s12911-021-01527-5>
- [14] Nourmohammadi-Khiarak, J., Feizi-Derakhshi, MR., Behrouzi, K. et al. (2020). New hybrid method for heart disease diagnosis utilizing optimization algorithm in feature selection. *Health Technol.* 10, 667–678 <https://doi.org/10.1007/s12553-019-00396-3>
- [15] Ashir Javeed, Shafqat Ullah Khan, Liaqat Ali, Sardar Ali, Yakubu Imrana, and Atiqur Rahman, (2022). Machine Learning-Based Automated Diagnostic Systems Developed for Heart Failure Prediction Using Different Types of Data Modalities: A Systematic Review and Future Directions, *Comput Math Methods Med* 2022;9288452. doi: 10.1155/2022/9288452.
- [16] S. Mohan, C. Thirumalai and G. Srivastava, (2019), Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, in *IEEE Access*, 7;81542-81554, doi: 10.1109/ACCESS.2019.2923707.
- [17] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, (2020). Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare, in *IEEE Access*, 8;107562-107582, doi: 10.1109/ACCESS.2020.3001149.
- [18] Abdul Saboor, Muhammad Usman, Sikandar Ali, Ali Samad, Muhammad Faisal Abrar, and Najeeb Ullah, (2022). A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms, *Mobile Information Systems*, (15);1-9 DOI:10.1155/2022/1410169
- [19] Nabaouia Louridi, Samira Douzi, Bouabid El Ouahidi, (2021) Machine learning-based identification of patients with a cardiovascular defect, *Journal of Big Data*, 8;133 <https://doi.org/10.1186/s40537021005249>
- [20] El-Shafiey, M.G., Hagag, A.M., El-Dahshan, E.A., & Ismail, M.A. (2022). A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest. *Multimedia Tools and Applications*, 81, 18155 - 18179.
- [21] Kusuma S, Divya Udayan J., (2020). Analysis on Deep Learning methods for ECG based Cardiovascular Disease prediction *Scalable Computing Practice and Experience* 21(1);127-136 DOI:10.12694/scpe.v21i1.1640

- [22] Stephen F. Weng, Jenna Reips, Joe Kai, Jonathan M. Garibaldi, Nadeem Qureshi, (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data?, *PLOS ONE* | <https://doi.org/10.1371/journal.pone.0174944>
- [23]K. Karthick, S. K. Aruna, Ravi Samikannu, Ramya Kuppusamy, Yuvaraja Teekaraman, and Amruth Ramesh Thelkar, (2022). Implementation of a Heart Disease Risk Prediction Model Using Machine Learning, *Comput Math Methods Med* 2022;6517716. doi: 10.1155/2022/6517716.
- [24]Lamido Yahaya, Nathaniel David Oye, Etemi Joshua Garba, (2020) A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques, *American Journal of Artificial Intelligence*, 4(1);20-29 DOI:10.11648/j.ajai.20200401.12
- [25]The CVD dataset <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>.