Journal of Theoretical and Applied Information Technology

30th November 2023. Vol.101. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN OPTIMIZED HYBRID CUCKOO SEARCH BIO-INSPIRED DEEP BELIEF NETWORK FOR HEART DISEASE PREDICTION

M. SIVAKAMI[1], P. PRABHU[2]

[1]Research Scholar, Department of Computer Applications, Alagappa University, Karaikudi – 630 003, India

[2]Associate Professor, Department of Information Technology, Directorate of Distance Education, Alagappa University, Karaikudi – 630 003, India
Email: sivakamiurcw@gmail.com, prabhup@alagappauniversity.ac.in

Abstract

In today's situation the heart diseases are one of the most common problems for many people who is working under pressure environments. These cause severe life losses for many of the aged persons. So nowadays many techniques used for the classification and the prediction of the heart diseases among the affected persons. In this novel the machine learning is used for computing the data's directly from the ECD data and set the out of heart diseases predictions. Based on the analysis, about 55% of the men are affected by the heart diseases compared to the women. The idea behind the Deep Belief Network with suitable metaheuristic algorithm plays significant contribution to predict the heart disease. Neural convolution model with Adam Optimizer, Decision tree with grey wolf model, Support vector machine classifier using Bayesian Optimization techniques, cuckoo search bio inspired deep belief network is also used to predict heart disease with lesser accuracy. Tiny deviation accuracy in the prediction of heart disease causes to increase wrong treatment. This paper proposes the advanced technique of Cuckoo search bio inspired among deep belief network model with SVM classifier for the identification of the heart diseases. Hamming distance feature selection and Gaussian filter also deployed for preprocessing and data cleaning. The proposed optimized hybrid cuckoo search bio inspired Deep Belief Network (CS-DBN) improve accuracy to predict heart disease. Comparative analysis was made with existing Convolution Neural Model with Adam Optimizer (CNMAO), Decision tree with grey wolf Model (DT-GWN), Support vector Machine classifier with Bayesian Optimization algorithm (SVM-BOA), Convolution Neural network with Social Mimic Optimization (SVM-SMO). Compared to the previous statistics the proposed hybrid CS-DBN algorithm is one of the most efficient methods of the prediction and the classification of the heart diseases. Here about 99.5% accuracy of the disease's prediction can be get using this advanced proposed method. For performing the analysis of the damaged ECG signal among all the ECG signal data which is collected from the data set is done. Then for this type of analysis the python platform is used for the disease's predictions.

Keywords: *Support Vector Machine, Cuckoo Search Bio Inspired Model, Deep Belief Networks. Gaussian Filter*

1. INTRODUCTION

The heart disease data is collected from the patient is analyzed using the method of deep belief

networks and the cuckoo search algorithm is proposed in this study. [2] The deep belief networks are a deep learning method that helps in generates the graphical model for classifying the deep neural networks. This is used to solve the problem of the finding the formation of the ECG data that is collected the data of the patients is done. [3] Gaussian filter is used for the analysis of the pixels for performing the smoothing of the noise in the ECG data which is collected from the heart diseases patient is done. The Adam optimization is used for improving the speed and the accuracy based on the deep learning models which approaches the loss of functions is done in this manuscript. [4] Based on analysis the feature selection is done for the automatic analysis of the selecting the features is done in the machine learning model. Then the decision tree is used for the regression and the classification of the heart diseases is done. The collection of the ECG signal data is done and the analysis of the heart diseases is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

done for the data collected. Then the comparison of the data is done for the heart disease data.

The main contribution of the paper is below:

1. Preprocessing of the given data set by feature selection using hamming distance for data cleaning.
2. Gaussian filter to implemented to remove the noise.
3. Cuckoo search bio inspired with the combination of the deep belief network and SVM classifier for the identification of the heart diseases.
4. Decision tree and Adam optimization is used as a optimizer in this study for the disease prediction.

Related works:

Author	DESCRIPTION	Method	Algorithm
Boukhatem et al 2022	Analysis of the heart diseases using machine learning	The classification method is used for the analysis	Machine learning algorithm is used for the determination
Riyaz et al 2022	The prediction of the early symptoms of the heart diseases.	Feature selection method and the hybrid ANN approach	Genetic and the machine learning algorithm is used
Sun et al 2023	Heart diseases is analysed based on the physical conditions	RFRS method has been used for the analysis.	Machine learning algorithm is used for the analysis of the heart diseases.
Nandhakumar et al 2022	Cuckoo search enabled is used for the analysis in this study.	The deep belief network method is used for the prediction and the classification	The machine learning algorithm is used in this manuscript
Yadav et al 2021	Prediction of the heart diseases is done in this study.	The regression method is used for finding the previous symptoms.	Random forest and the decision tree algorithm is used for analysing the heart diseases.
Digumarthi et al 2022	Early prediction of the heart disease is done that is mentioned as the Cardiac Arrhythmia	Classification and the feature selection method for the images is done.	Bio inspired algorithm.

2. PROPOSED HYBRID CS-DBN METHOD

The proposed diagram is drawn in figure 1. This proposed method enables the feature selection method. Cleaning data t is proposed by the selection of the feature using hamming distance. For the data processing, the data handling and the cleaning of the data is done using the method of feature selection. The attribute of the hamming selection manages the selection of the feature method and the estimation of the distance among the heart is done. Heart disease can be analyzed by the classification of machine learning.

The validation of heart diseases achieved by testing and training the heart diseases data. The estimation for the diseases diagnosis is done by cross validation. The classification of the models with the decision trees and the support vector is done for the classification. The comparison of the results and the prediction of the testing and the training methods has been enabled in this manuscript. Also, the accuracy that enables the prediction of the cardiac diseases is proposed in this paper.

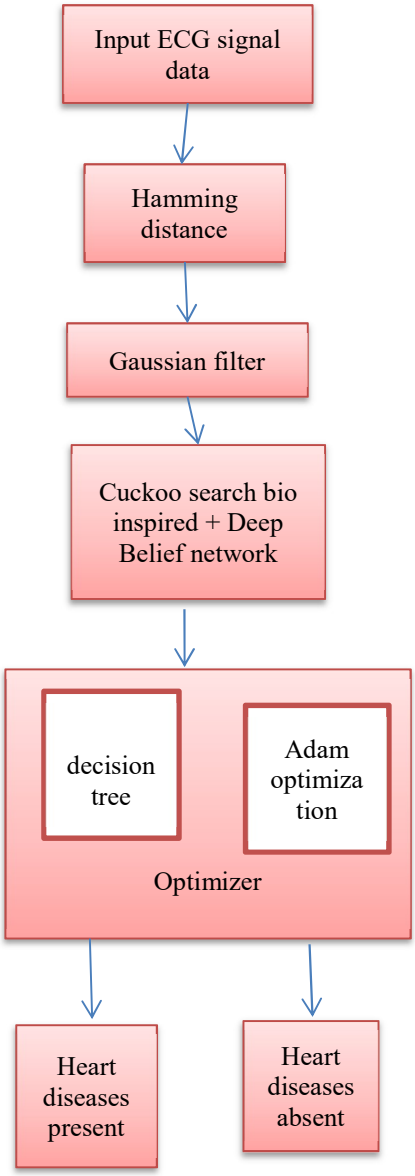


Figure 1 Proposed Hybrid CS-DBN Method

In this proposed work the data set for the process of feature selection has been introduced for the person those who are all affected by the heart diseases. The finding of the data that is missed and the reduction of the features is organized well for the selection process. Also, to identify that there are many different ways for the selection of hamming distance, the analysis and the selection process has been enabled. Some of the groups that enables the forecasting of the heart diseases is examined using the research objective and then the analysis of the sickness and the states has been made possible in this manuscript.

3. HAMMING DISTANCE FOR DATA CLEANING

The data set that is collected and then the cleaning process is done for the two binary vector and for finding the difference between the vectors. Cleaning of the non-linearity and the property of the data has been done the cleaning process. The length of the letter and the changes is the numbers has been enabled for another data preprocessing methods for the data analysis is done [5].

Let X be the symbol for the alphabet and Y denotes the subset of the X, here the words in the length denoted as a over X.

Let x=(a_1 ,.....a_z) and y=(b_1……b_y) in the word B

Hamming distance is defined as the places count which indicates the words as x and y, can be vary from the equation as i=x_i not equal to y_i with respect to i= 1…n. Here the distance that is done using the hamming model is determined as the

C(x,y) is greater than are equal to 0 and C (a,b) if x is equal to y (1)

C(x,y) is equal to C(y,X); (2)

C(x,y) is less than or equal to the C(x,w) + C(w,y) (3)

W indicates the weight of the hamming and this represent the metric and C. The correction code for correcting the error in the words indicates the B and this transfer of the data and the changing in the symbols enables the error that is upto the order up and the errors in the strings is enables [6] . This defines the types of the string and the analysis of the length of the strings. For instance, the strings indicates the value of S_1 and S_2 this enables the hamming distance that indicates the H(S_1, S_2) is the different strings and the types of strings that enables the characters of the distinct. Here the word A is a letter enables the linear type of the vector space and the finite field for the determining of the weight and the hamming distance.

D (X, Y) is equal to w(X-Y) (4)

Here Y,X and the X,Y denotes the subset and the length of the words.

4. GAUSSIAN FILTER

Gaussian filtering is a common method for signal processing and can be applied to ECG data to smooth out noise and reduce artifacts. The Gaussian filtering is to convolve the signal with a Gaussian kernel, which is a bell-shaped curve with a mean and standard deviation that determine the degree of smoothing [7].To apply a Gaussian filter to ECG

data, first determine the appropriate standard deviation of the Gaussian kernel. This value should be selected based on the characteristics of the noise in the signal and the desired level of smoothing. A larger standard deviation results in more smoothing, but may also blur the details in the signal. When determined the appropriate standard deviation, apply the Gaussian filter using a convolution operation [8]

.

INPUT *Heart disease data matrix M of size m x n, Gaussian filter size k x k, and standard deviation sigma* ***OUTPUT*** *Smoothed heart disease data matrix M_s of size m x n*
set*Gaussian filter of size k x k with standard deviation sigma:* ***compute*** *center of the filter:* $(x_0, y_0) = ((k+1)/2, (k+1)/2)$ ***for*** *each element (x, y) of the filter* ***compute****Gaussian function:* $G(x, y) = \frac{1}{\delta\sqrt{2\pi}} e^{-((x-x_0)^2+(y-y_0)^2)/(2\delta^2)}$ *Pad(M with zeros)= (m + k - 1) x (n + k - 1)* ***end for*** ***for*** *each pixel (i, j) in the original matrix M* ***extract*** *submatrix centred at pixel (i, j)* $M\_w(i, j) = \sum_{x=1}^{k} \sum_{y=1}^{k} G(x, y) * M(i + x - 1, j + y - 1)$ $M\_s(i, j) = M\_w(i, j)$ *output matrix M_s =Trim (k-1 rows, columns of zeros from the edges)* ***Return*** *Smoothed heart disease data matrix M_s.*

In the case of heart disease data, the Gaussian filter can be used to reduce noise and make patterns in the data more visible. The algorithm starts by creating a Gaussian filter of size k x k with a given standard deviation sigma. The center of the filter is computed as

$$(x_0, y_0) = ((k+1)/2, (k+1)/2) \quad (1)$$

and for each element (x, y) of the filter, its value is computed using the Gaussian function. The filter is then normalized so that the sum of all its elements is equal to 1.

$$G(x, y) = \frac{1}{\delta\sqrt{2\pi}} e^{-((x-x_0)^2+(y-y_0)^2)/(2\delta^2)} \quad (2)$$

where, the δ represents the standard deviation or scale parameter of the Gaussian function, controlling the width of the distribution.The input matrix M of size m x n is then padded with zeros so that its size becomes (m + k - 1) x (n + k - 1). For each pixel (i, j) in the original matrix M, a k x k submatrix centered at that pixel is extracted from the padded matrix. The submatrix is multiplied by the Gaussian filter element-wise to obtain a weighted sum of the neighboring pixels, and the value of pixel (i, j) in the output matrix M_s is replaced with the weighted sum. Finally, the output matrix M_s is trimmed to its original size m x n by removing the k-1 rows and columns of zeros from the edges. This algorithm assumes that the heart disease data matrix M is a grayscale image represented as a matrix of intensity values, it can be adapted to work with other types of data as well. Overall, the Gaussian filter algorithm is a simple and effective way to smooth heart disease data and make patterns in the data more visible.

The preprocessing function involves taking ECG-based CSV data and then training using a proposed method using Python framework that is represented as follows,

$$F(D)= D\{1,2,3\ldots m\} \quad (1)$$

where the dataset is denoted by D, and the data training function is represented by F. The raw database contains normal and noisy features [9] , where i and the noisy features represent normal features by H, which refers to as D(i,H). The noise elimination (N_E) process is described using Equation (2).

$$N\_E= D(i,H)-D(H) \quad (2)$$

The feature extraction process involves eliminating unwanted features and tracking disease features. The chimp fitness process executes this process, where meaningless features are removed using Equation (3). The feature analysis function is represented by P, Q represents meaningless features monitoring process by Q, R represents useless features, and S refers tracked heart functioning feature.Here is a list of commonly tracked heart functioning features, Heart Rate, R-R Interval, P-Wave Duration, QRS Complex Duration, QT Interval, ST Segment, T-Wave Morphology, Heart Rate Variability (HRV), P-Q Interval, and ST-T Wave Changes.

$$P=|D(S,R) - Q(R)| \quad (3)$$

The disease type (T_d) classification is based on heart failure and arrhythmia, where the arrhythmia is specified in three classes: fusion beats, ventricular ectopic, and supraventricular ectopic.

$$T_d= = \begin{cases} Normal, & P = 0 \\ Abnormal, & P \neq 0 \end{cases} \quad (4)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Equation (4) is used to classify the disease type, If the traced feature falls under the 0th class, it returns, the ECG signal is a normal ECG signal, while if it falls under the 1st class, it returns, the ECG signal is an abnormal ECG signal data.

Algorithm for Classification

```
Input: Raw ECG-based CSV data
Output: ECG signal classification
# Feature extraction:
function preprocess(data):
        preprocessed_data = apply_equation(data)
        noise_removed_data = remove_noise(preprocessed_data)
return noise_removed_data
processed_data = train(data)
function remove_noise(data):
        N_E= D(i,H)-D(H)
        normal_features = track_normal_features(data)
        noisy_features = track_noisy_features(data)
        noise_removed_data=remove_noise (data,normal_features,noisy_features)
return noise_removed_data
#Classification function:
Functionclassify(data):
        heart_function_rate = trace_heart_function_rate(data)
        if (heart_function_rate =normal)
                return "Normal"
        else if (heart_function_rate= abnormal)
                disease_type = classify_disease_type(heart_function_rate)
        End if
        if (disease_type = arrhythmia)
                arrhythmia_type = classify_arrhythmia_type(heart_function_rate)
                return arrhythmia_type
        else if (disease_type = heart_failure)
                return "Heart Failure"
        End if
Function classify_arrhythmia_type(heart_function_rate)
        if (heart_function_rate = fusion_beats)
                return "Fusion Beats"
        else if (heart_function_rate = ventricular_ectopic)
                return "Ventricular Ectopic"
        else if (heart_function_rate = supraventricular_ectopic)
                return "Supraventricular Ectopic"
return arrhythmia_type
Function classify_disease_type(heart_function_rate):
        if (heart_function_rate= abnormal)
                return "Abnormal ECG signal data"
        else if (heart_function_rate= normal)
                return "Normal ECG signal data "
return Normal or abnormal ECG signal data
```

The classification of neural networks and the theorem that allows for the universal approximation of continuous functions with real values [10]. Additionally, the deep belief network method is employed, which serves as a classification model for analyzing datasets and making predictions. The study addresses the challenges associated with classification and prediction, aiming to achieve higher accuracy in both areas. The approach involves procedures for data collection, testing the dataset, calculating voting time, and determining average accuracy.

5. DEEP BELIEF NETWORKS WITH RESPECT TO THE OPTIMIZATION ALGORITHM:

The deep belief networks analyses the alternative factorization of the requested networks has been structured. The model for finding the probabilities and the layers that enables the deep belief networks and the analysis of the whole information makes the process of clustering, identification of the data and the images processing. The optimization algorithm that is used for the mathematical problems and the feasible solution for the analysis of the stochastic and the deterministic of the optimization problem. The procedure of the data to find the analysis of the data can be done in the problem-solving method.

Pseudocode for Cuckoo search algorithm with a Deep Belief Network

```
function CuckooSearch (heartData, maxGenerations, populationSize, numNests)
        nests = InitializeNests(populationSize)
        EvaluateNests (nests, heartData)
        SortNestsByFitness (nests)
        globalBest = nests [0]
        generation = 1
```

```
        while generation <= maxGenerations
                newNests = GenerateNewNests (nests, numNests)
                EvaluateNests (newNests, heartData)
                mergedNests = MergeNests (nests, newNests)

        SortNestsByFitness(mergedNests)
                nests = ReplaceNests (mergedNests, numNests)
                if nests [0]. fitness < globalBest.fitness
                        globalBest = nests [0]
                        generation = generation + 1
                end if
        end while
return globalBest
function InitializeNests(populationSize)
        nests = empty list
        for i = 1 to populationSize
                nest = randomly initialize a solution
                nests. append (nest)
        end for
return nests
function EvaluateNests (nests, heartData)
        for each nest in nests
                nest. fitness = evaluateFitness (nest, heartData)
        end for
end function
function GenerateNewNests (nests, numNests)
        newNests = empty list
        for i = 1 to numNests
                newNest = performCuckooSearchOperations (nests)
                newNests.append(newNest)
        end for
return newNests
function MergeNests (nests, newNests)
        mergedNests = nests + newNests
return mergedNests
function ReplaceNests (mergedNests, numNests)
        sortedNests = SortNestsByFitness(mergedNests)
        selectedNests = sortedNests [: numNests]
return selectedNests
function evaluateFitnessWithDBN (nest, heartData)
        modelConfig = convertNestToModelConfig(nest)
        trainedDBN = trainedDBN (modelConfig, heartData)
        fitness = calculateFitness (trainedDBN, heartData)
return fitness
```

The Cuckoo Search algorithm is a metaheuristic optimization algorithm inspired by the behavior of cuckoo birds and their brood parasitism strategy. It is commonly used to solve optimization problems. In the context of heart disease classification, the Cuckoo Search algorithm can be integrated with a Deep Belief Network (DBN) to improve the accuracy of classification.

A DBN is a type of artificial neural network that consists of multiple layers of hidden units. It is trained using unsupervised learning techniques, such as Restricted Boltzmann Machines (RBMs), followed by fine-tuning using supervised learning [12] . DBNs are known for their ability to capture complex patterns and hierarchical representations in data.

The integration of Cuckoo Search algorithm with a DBN for heart disease classification involves the following steps, randomly initialize a population of nests, where each nest represents a potential solution or configuration for the DB. Evaluate the initial nests by training the corresponding DBN models using the heart disease data. The fitness of each nest is determined based on the accuracy or other performance metric achieved by the DBN when classifying the heart disease data. Sort the nests in descending order of fitness, with the best performing nests at the beginning of the list. Set the best nest found so far (the nest with the highest fitness) as the global best. Begin the optimization loop, which typically involves generating new nests using Cuckoo Search operations, evaluating their fitness using the DBN, and updating the population of nests based on their performance. Generate new nests by performing Cuckoo Search operations, such as Levy flight or random perturbations, on the existing nests. These operations explore the search space and potentially discover better configurations for the DB. Train the DBN models corresponding to the new nests using the heart disease data and evaluate their fitness based on the achieved classification performance. [13] Merge the existing nests with the new nests and sort them based on their fitness values. Select the best nests from the merged list to replace the worst nests in the population. [14] This ensures that the population maintains diversity and includes the most promising solutions. If the best nest in the new population has higher fitness than the current global best, update the global best with this nest. Repeat steps until a termination condition is met, such as

reaching the maximum number of generations or achieving a desired level of performance. Return the best nest (DBN configuration) found during the optimization process, which represents the optimal solution for heart disease classification. [15] By integrating the Cuckoo Search algorithm with a DBN, this approach combines the exploration and exploitation capabilities of Cuckoo Search with the powerful pattern recognition and classification abilities of the DBN, leading to improved accuracy and performance in heart disease classification tasks.

6. RESULT AND DISCUSSION

The dataset is a combination of two collections of heartbeat signals, derived from the MIT-BIH Arrhythmia Dataset and the PTB Diagnostic ECG Database. It contains a total of 123,998 samples, which are preprocessed and segmented into individual heartbeats. [16] The Arrhythmia Dataset has 109,446 samples with 5 categories - 'N', 'S', 'V', 'F', and 'Q', representing normal cases and different types of arrhythmias and myocardial infarction. The PTB Diagnostic ECG Database has 14,552 samples with 2 categories - normal and abnormal cases. [17] The signals in both datasets have a sampling frequency of 125Hz, making them suitable for training deep belief neural networks. This dataset has been widely used for exploring heartbeat classification using deep belief neural network architectures and transfer learning. Overall, this dataset offers a valuable resource for researchers and practitioners interested in developing machine learning models for heart disease diagnosis and classification based on ECG signals. The following figure 1 and 2 depicts that abnormal and normal persons ECG visualization

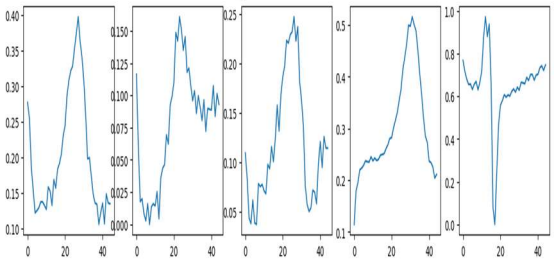


Figure.1ECG Visualization Of Abnormal Persons

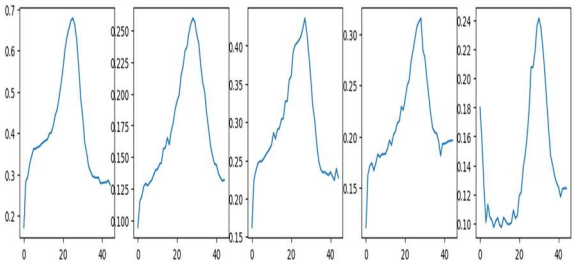


Figure 2.ECG visualization of Normal persons

In this paper, hybrid CS-DBN model is proposed to predict the heart diseases with high accuracy. In this model Cuckoo Search bio-inspired Algorithm (CSA) for optimization, SVM for classification, and Deep Belief Networks (DBN) for feature extraction has been done. [18] To prepare the data, the authors used the Hamming distance for feature selection and applied a Gaussian filter for noise removal. [19] The proposed model achieved promising results, indicating that the combination of CSA, SVM, and DBN is effective in identifying heart diseases from ECG data. To compare the proposed model with other state-of-the-art models, such as CNMAO and DT-GWM. CNMAO uses a neural convolution model with Adam Optimizer, while DT-GWM uses a decision tree with a grey wolf model. [20] Convolutional Neural Network with Social Mimic Optimization (CNN-SMO) and Support Vector Machine classifier using Bayesian Optimization algorithm (SVM-BOA). The comparison showed that the proposed model outperformed than others in terms of accuracy and performance.

There are several metrics that can be used to evaluate the performance of the proposed method for heart disease classification of ECG dataset. Here are some commonly used metrics:

A confusion matrix is a table that shown in Table 1, it used to evaluate the performance of a classification model by comparing the predicted and actual class labels of a dataset. The table consists of four cells that represent the following:

True Positive (TP): Instances that are correctly classified as positive.

False Positive (FP): Instances that are incorrectly classified as positive.

True Negative (TN): Instances that are correctly classified as negative.

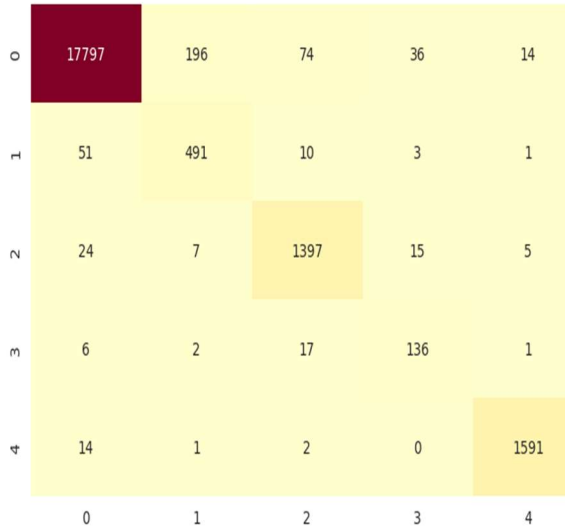
False Negative (FN): Instances that are incorrectly classified as negative.

Journal of Theoretical and Applied Information Technology

30th November 2023. Vol.101. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Confusion Matrix For Heart Disease Classification



When visualizing a confusion matrix, a color range can be used to represent the values in the cells. The color range is typically chosen to provide a visual gradient that corresponds to the magnitude of the values in the matrix. This color range can be defined based on a continuous scale, such as a gradient from light to dark or from low to high intensity. The specific color palette used can vary depending on the visualization tool or personal preference.

The dataset has the Classes of ['N': 0, 'S': 1, 'V': 2, 'F': 3, 'Q': 4]. here, 0-4 counts show true and predicted label of data in the datasets.

The table 1shows the comparison of the accuracy values for five different methods used for heart disease classification. The methods are Neural Convolution Model with Adam Optimizer (CNMAO), Decision Tree with Grey Wolf Model (DT-GWM), Support Vector Machine classifier using Bayesian Optimization algorithm (SVM-BOA), Convolutional Neural Network with Social Mimic Optimization (CNN-SMO), and a proposed hybrid method (CS-DBN)

Accuracy: This metric can be used to measure the overall performance of the model in correctly identifying the different classes of heart diseases.

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{TrueNegative} + \text{FalsePositive} + \text{False Negative}}$$

Precision= True positive/True positive+False Positive

Table 2: Comparison Of Performance Metrics For Different Methods

Methods	Accuracy Values
Neural Convolution Model with Adam Optimizer (CNMAO)	93.5%
Decision Tree withGrey Wolf Model (DT-GWM)	87.2%
Support Vector Machine classifier using Bayesian Optimization algorithm (SVM-BOA).	89.4%
Convolutional Neural Network with Social Mimic Optimization(CNN-SMO)	90.4%
Hybrid CS-DBN model	95.4%

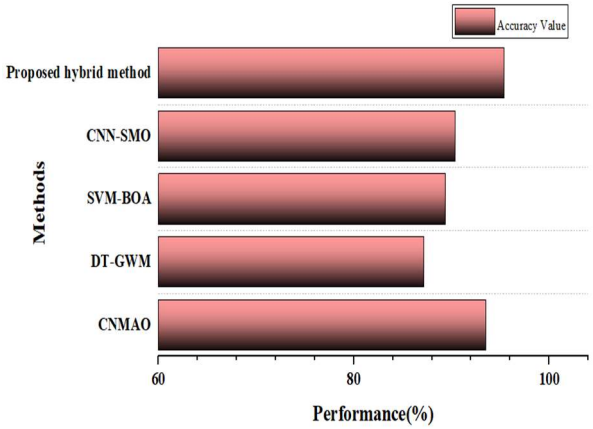


Figure 3: Accuracy Score Of Various Methods

The accuracy values for each method are represented in figure 3. The y-axis represents the methods, and the x-axis represents the accuracy values in percentage.The graph shows that the proposed Hybrid (CS-DBN) Method has the highest accuracy value of 95.4%, followed by the Neural Convolution Model with Adam Optimizer (CNMAO) with an accuracy value of 93.5%. [20] The Decision Tree with Grey Wolf Model (DT-GWM) has the lowest accuracy value of 87.2%, while the Support Vector Machine classifier using Bayesian Optimization algorithm (SVM-BOA) and the Convolution Neural Network with Social Mimic Optimization(CNN-SMO) have accuracy values of 89.4% and 90.4%, respectively.In this case, the proposed Hybrid (CS-DBN) Method appears to have the highest accuracy value among the methods being compared.

Precision: Precision can be used to evaluate the proportion of true positives (correctly identified instances of a particular heart disease) among all

instances that the model has classified as that particular disease [21] .

Table 3:Comparison Of Existing And Proposed Method Based On Precision

Epochs	Precision				
	CNMAO	DT-GWM	CNN-SMO	SVM-BOA	Proposed hybrid method
5	0.70	0.64	0.68	0.65	0.75
10	0.73	0.65	0.70	0.67	0.77
15	0.76	0.68	0.74	0.70	0.79
20	0.79	0.70	0.76	0.73	0.82
25	0.81	0.72	0.78	0.75	0.84
30	0.85	0.76	0.82	0.78	0.88
35	0.90	0.80	0.87	0.83	0.92
40	0.92	0.82	0.89	0.86	0.94
45	0.93	0.84	0.91	0.88	0.95
50	0.94	0.85	0.92	0.89	0.96

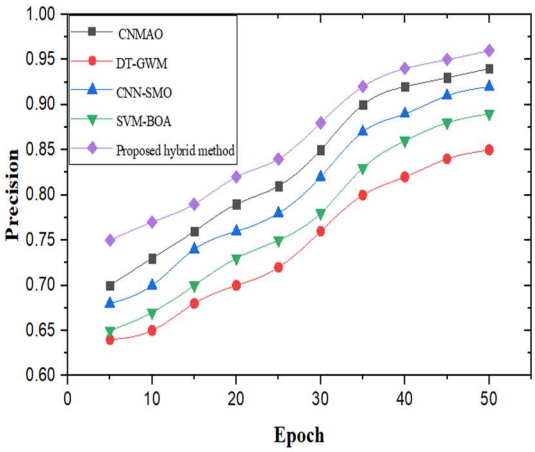


Figure 4: Epoch Vs Precision For CNMAO, DT-GWM, CNN-SMO, SVM-BOA, And Proposed Hybrid Method

Figure 4 shows the precision values of five different classification methods, namely CNMAO, DT-GWM, CNN-SMO, SVM-BOA, and the proposed Hybrid (CS-DBN) Method, over the course of 50 epochs. The precision values for each method are shown in the y-axis, while the x-axis represents the epochs.As can be observed from the graph, all methods generally show an increase in precision as the number of epochs increases. The proposed Hybrid (CS-DBN) Method consistently shows the highest precision values across all epochs, with a precision of 0.75 at epoch 5, increasing steadily to a peak value of 0.96 at epoch 50.CNMAO and CNN-SMO show similar precision values, with both methods starting at around 0.68-0.70 precision at epoch 5, and gradually increasing to reach 0.94-0.92 precision at epoch 50. The SVM-BOA and DT-GWM start with lower precision values than the other methods, with SVM-BOA starting at 0.65 and DT-GWM starting at 0.64 precision at epoch 5. However, both methods show a steady increase in precision as the number of epochs increases, with SVM-BOA reaching a precision of 0.89 at epoch 50, and DT-GWM reaching a precision of 0.85 at epoch 50.Overall, the graph suggests that the proposed Hybrid (CS-DBN) Method shows the highest precision values across all epochs, followed by CNMAO and CNN-SMO, while SVM-BOA and DT-GWM start with lower precision values but show steady improvement as the number of epochs increases.

Recall: Recall can be used to evaluate the proportion of true positives (correctly identified instances of a particular heart disease) among all instances that actually belong to that particular heart disease class.

Recall=True positive/True positive+False negative

Table 3: Comparison Of Existing And Proposed Method Based On Recall

Epochs	Recall				
	CNMAO	DT-GWM	CNN-SMO	SVM-BOA	Proposed hybrid method
5	0.65	0.54	0.62	0.57	0.72
10	0.68	0.56	0.65	0.60	0.75
15	0.72	0.62	0.65	0.68	0.78
20	0.75	0.65	0.68	0.71	0.80
25	0.79	0.69	0.72	0.75	0.83
30	0.82	0.72	0.76	0.78	0.86
35	0.85	0.75	0.79	0.81	0.88
40	0.89	0.79	0.84	0.85	0.92
45	0.92	0.82	0.88	0.89	0.94
50	0.93	0.84	0.89	0.91	0.95

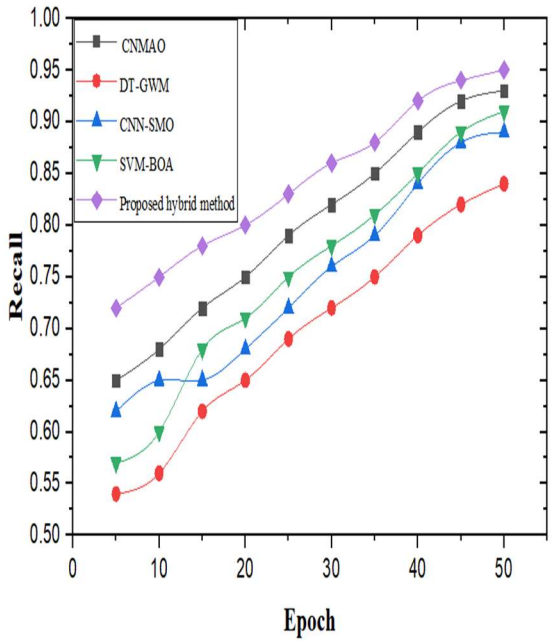


Figure 5: Recall Vs. Epochs Comparison For Different ML Algorithms

Figure 5 illustates that the performance of different methods in terms of their recall values across different epochs, ranging from 5 to 50. The x-axis represents the number of epochs, while the y-axis represents the recall values. From the graph, it is observed that the recall values generally improve as the number of epochs increases for all methods. The proposed Hybrid (CS-DBN) Method consistently outperforms the other methods in terms of recall values across all epochs, with a steady increase from 0.72 at epoch 5 to 0.95 at epoch 50. The CNMAO method also shows a good improvement in recall values, starting at 0.65 and increasing to 0.93 at epoch 50. The other three methods, DT-GWM, CNN-SMO, and SVM-BOA, show lower recall values compared to the other two methods, with SVM-BOA showing the lowest recall values.The figure 5 suggests that the proposed hybrid(CS-DBN) method is the most effective for heart disease classification among the five methods.

F1 Score: The F1 score can be used to measure the overall balance between precision and recall for all classes of heart disease.

F1Score=

2*(Precision*Recall)/(Precision+Recall)

Table 4: F1 Score Comparison Of Machine Learning Models Over Epochs

Epochs	F1 Score				
	CNMAO	DT-GWM	CNN-SMO	SVM-BOA	Proposed hybrid (CS-DBN) method
5	0.67	0.58	0.64	0.60	0.73
10	0.70	0.60	0.67	0.63	0.75
15	0.73	0.64	0.69	0.68	0.78
20	0.76	0.67	0.71	0.71	0.80
25	0.79	0.70	0.74	0.75	0.83
30	0.83	0.73	0.78	0.78	0.86
35	0.87	0.77	0.82	0.81	0.89
40	0.90	0.80	0.86	0.85	0.92
45	0.92	0.82	0.89	0.88	0.94
50	0.93	0.84	0.90	0.89	0.95

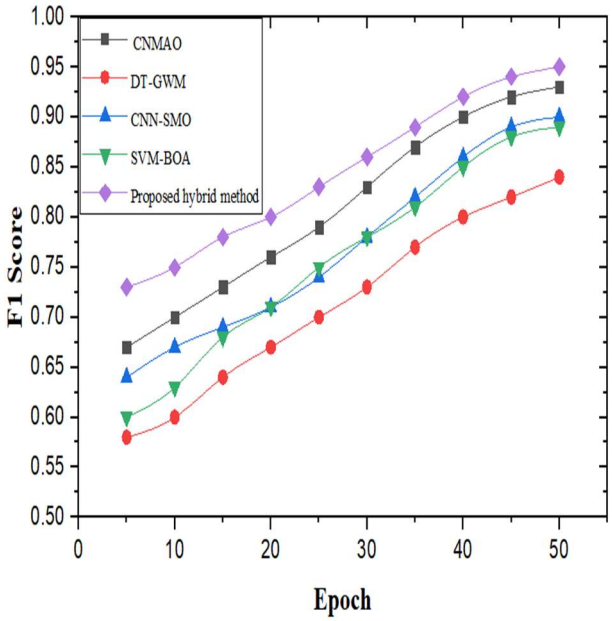


Figure 6. F1 Score Analysis Of Various Methods Figure 6 Shows The F1 Score Achieved By Five Different Machine Learning Methods

The figure shows that the F1 score achieved for the methods CNMAO, DT-GWM, CNN-SMO, SVM-BOA, and proposed hybrid (CS-DBN) for different epochs ranging from 5 to 50. As the epochs increase from 5 to 50, all the methods show an improvement in their F1 score. The proposed hybrid (CS-DBN) method outperforms all the other methods consistently across all epochs, with an F1 score of 0.73 at epoch 5 increasing to 0.95 at epoch 50. CNMAO, DT-GWM, CNN-SMO, and SVM-BOA also show improvements in

their F1 scores as the epochs increase, but they are consistently outperformed by the proposed hybrid (CS-DBN) method.Overall, the figure shows that the proposed hybrid(CS-DBN) method achieves the highest F1 score, indicating that it provides the best balance between precision and recall and is the most accurate method for this task.

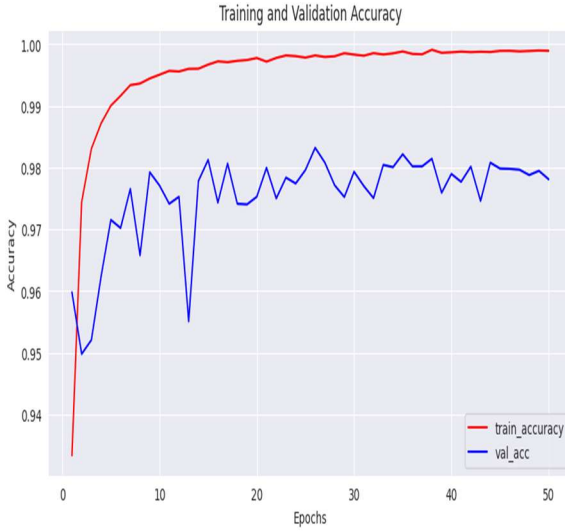


Figure 7: Training And Validation Accuracy Graph

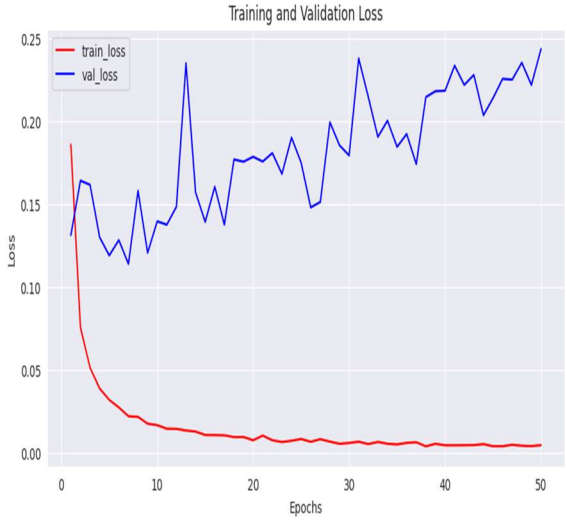


Figure 8: Training And Validation Loss Graph

The above Figures 7 and 8 shows the proposed methods training and validation accuracy and loss graphs. The horizontal axis refers epochs from 0 to 50 while the vertical axis holds the loss and accuracy values. From the observance of figure 7, the training accuracy is higher than the validation accuracy, but both the curves follows upward when the epochs counts are increased. At epoch 13, the sudden fall is occur for validation curve afterthat it follows, upward trend as higher. At the end, the proposed hybrid (CS-DBN) has training accuracy and validation accuracy as 0.99% and 0.97% respectively. The training and validation loss figure 8 shows the loss (error) of the model on the training data during the training process respectively. As the model learns from the training data, the training loss typically decreases, indicating that the model is getting better at minimizing the difference between its predicted output and the true output.If the training accuracy is increasing and reaches a high level, it indicates that the proposed hybrid (CS-DBN) model is learning well and can make accurate predictions on the training data. If the training loss is decreasing and reaches a low level, it indicates that the model is learning well and can make accurate predictions on the training data.

7. CONCLUSION:

In this paper the proposed hybrid (CS-DBN) model for predicting and analyzing the accuracy of the heart diseases. The tabulation for the data set and the analyzing of the hamming distance is done based on the data cleaning and the feature selection process. As a per the experimental result the paper conclude that the proposed hybrid(CS-DBN) model achieved with higher accuracy for predicting the heart diseases.

REFERENCES:

[1]
[2] Boukhatem, C., Youssef, H. Y., & Nassif, A. B. (2022, February). Heart disease prediction using machine learning. In *2022 Advances in Science and Engineering Technology International Conferences (ASET)* (pp. 1-6). IEEE.
[3] Riyaz, L., Butt, M. A., Zaman, M., &Ayob, O. (2022). Heart disease prediction using machine learning techniques: a quantitative review. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 3* (pp. 81-94). Springer Singapore.
[4] Sun, H., & Pan, J. (2023). Heart Disease Prediction Using Machine Learning Algorithms with Self-Measurable Physical Condition Indicators. *Journal of Data Analysis and Information Processing*, *11*(1), 1-10.
[5] Nandakumar, P., & Narayan, S. (2022). Cardiac disease detection using cuckoo search enabled deep belief network. *Intelligent Systems with Applications*, *16*, 200131.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[6] Yadav, A., Gediya, L., & Kazi, A. (2021). Heart disease prediction using machine learning. *International Research Journal of Engineering and Technology (IRJET*, *8*(09).

[7] Digumarthi, J., Gayathri, V. M., &Pitchai, R. (2022, April). Early Prediction of Cardiac Arrhythmia using Novel Bio-inspired Algorithms. In *2022 8th International Conference on Smart Structures and Systems (ICSSS)* (pp. 01-04). IEEE.

[8] Diwakar, M., Tripathi, A., Joshi, K., Memoria, M., & Singh, P. (2021). Latest trends on heart disease prediction using machine learning and image fusion. *Materials Today: Proceedings*, *37*, 3213-3218.

[9] Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Computers in Biology and Medicine*, *136*, 104672.

[10] Cheung, C. Y., Xu, D., Cheng, C. Y., Sabanayagam, C., Tham, Y. C., Yu, M., ... & Wong, T. Y. (2021). A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nature biomedical engineering*, *5*(6), 498-508.

[11] Ashfaq, A., Sant'Anna, A., Lingman, M., &Nowaczyk, S. (2019). Readmission prediction using deep learning on electronic health records. *Journal of biomedical informatics*, *97*, 103256.

[12] Al-Makhadmeh, Z., &Tolba, A. (2019). Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: A classification approach. *Measurement*, *147*, 106815.

[13] Naz, H., & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. *Journal of Diabetes & Metabolic Disorders*, *19*, 391-403.

[14] Wong, K. K., Fortino, G., & Abbott, D. (2020). Deep learning-based cardiovascular image diagnosis: a promising challenge. *Future Generation Computer Systems*, *110*, 802-811.

[15] Hong, S., Zhou, Y., Shang, J., Xiao, C., & Sun, J. (2020). Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in biology and medicine*, *122*, 103801.

[16] Basir, M. A., Hussin, M. S., & Yusof, Y. (2020). Optimization of Multi-objective ENORA and NSGA-II based on Bio-Inspired Algorithms for Classification Problem. *International Journal*, *9*(1.3).

[17] Kotiyal, V., Singh, A., Sharma, S., Nagar, J., & Lee, C. C. (2021). ECS-NL: An enhanced cuckoo search algorithm for node localisation in wireless sensor networks. *Sensors*, *21*(11), 3576.

[18] Basir, M. A., Hussin, M. S., & Yusof, Y. (2020). Integrated bio-search approaches with multi-objective algorithms for optimization and classification problem. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *18*(5), 2421-2431.

[19] Baburaj, E. (2022). Comparative analysis of bio-inspired optimization algorithms in neural network-based data mining classification. *International Journal of Swarm Intelligence Research (IJSIR)*, *13*(1), 1-25.

[20] Barbosa, C. E. M., & Vasconcelos, G. C. (2019, June). Analysis of Cuckoo Search Efficiency. In *2019 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1351-1358). IEEE.

[21] Larabi-Marie-Sainte, S., Alskireen, R., &Alhalawani, S. (2021). Emerging applications of bio-inspired algorithms in image segmentation. *Electronics*, *10*(24), 3116.

[22] Yahaya, L., Oye, N. D., & Garba, E. J. (2020). A comprehensive review on heart disease prediction using data mining and machine learning techniques. *American Journal of Artificial Intelligence*, *4*(1), 20-29.

[23] Sharma, V., Yadav, S., & Gupta, M. (2020, December). Heart disease prediction using machine learning techniques. In *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)* (pp. 177-181). IEEE.

[24] Tarawneh, M., &Embarak, O. (2019). Hybrid approach for heart disease prediction using data mining techniques. In *Advances in Internet, Data and Web Technologies: The 7th International Conference on Emerging Internet, Data and Web Technologies (EIDWT-2019)* (pp. 447-454). Springer International Publishing.

[25] Sun, W., Zhang, P., Wang, Z., & Li, D. (2021). Prediction of cardiovascular diseases based on machine learning. *ASP Transactions on Internet of Things*, *1*(1), 30-35.