

## Quality Analysis of Extreme Learning Machine based on Cuckoo Search and Invasive Weed Optimization

Nilesh Rathod<sup>1,\*</sup>, Sunil Wankhade<sup>2</sup>

<sup>1</sup>Department of Computer Engineering, Mct's Rajiv Gandhi Institute of Technology, Mumbai, India-400053

<sup>2</sup>Mct's Rajiv Gandhi Institute of Technology, Mumbai, India-400053

### Abstract

This paper explicates optimization driven Extreme Learning Machine (ELM) strategy developed with feed forward neural network (FFNN) for classification of data. In this paper we proposed hybrid optimization based Extreme Learning Machine which is based on modified Cuckoo Search and modified Invasive Weed Optimization. To optimize the input weights and hidden neurons we proposed CSIWO ELM method. With this intelligent hybrid optimization method, we find out optimal solution. We are measuring the quality of the solution through linear function-based analysis, objective function-based analysis, and optimization function-based analysis. The experimental results proposed in this paper show the feasibility and effectiveness of the developed CSIWO ELM method with encouraging performance compared with other ELM methods.

Received on 09 April 2022; accepted on 14 May 2022; published on 18 May 2022

**Keywords:** Extreme Machine Learning (ELM), Cuckoo search (CS), Invasive Weed Optimization, feed forward neural network (FFNN), Optimization.

Copyright © 2022 Nilesh Rathod *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/airo.v1i.383

### 1. Introduction

Extreme learning machines are feed-forward neural networks having a single layer or multiple layers of hidden nodes for classification, regression, clustering, sparse approximation, compression, and feature learning, where the hidden node parameters do not need to be modified. These hidden nodes might be assigned at random and never updated, or they can be inherited from their predecessors and never modified. In most cases, the weights of hidden nodes are usually learned in a single step which essentially results in a fast-learning scheme. These models are capable of producing good generalization performance and learning thousands of times quicker than backpropagation networks. These models can also outperform support vector machines in classification and regression applications, according to the research [1].

Extreme learning machine (ELM) [2] is an extremely efficient algorithm for training a single hidden layer

feed-forward network. The learning parameters of the hidden layer in the ELM algorithm network structure are randomly selected that only necessary to set the number of hidden layer network neurons. Finally, the output weight of the hidden layer is obtained by the least squares method, and no iteration is required for the network weight and offset in the process. Therefore, compared with the traditional neural network algorithm, the ELM algorithm has the advantages of fast training speed and strong generalization ability [3]. This unique characteristic of fast training mainly arises from two points. On one hand, before acquiring the training data, the parameters of latent layer are randomly generated and will be fixed without fine-tuning. On the other hand, instead of repeatedly applying chain rule for partial derivatives of model parameters [4], one can derive the output weights analytically. In theory, Huang et al. [5] have proven that the output weights solved from regularized least squares problem can maintain the universal approximation capability of connectionist model, even with the stochastic latent

\*Corresponding author. Email: [nilesh.rathod@mctrigit.ac.in](mailto:nilesh.rathod@mctrigit.ac.in)

layer weights. Therefore, ELM tends to have faster training process than that of back-propagation based neural networks and support vector machine [6], while achieving the good generalization. The ELM algorithm is an effective single-hidden layer feed forward neural networks learning algorithm. Traditional extreme learning machine (ELM) and Support Vector Machine (SVM) provides equal importance to all the samples leading to results biased towards the majority class. Many variants of ELM-like Weighted ELM (WELM), Boosting WELM (BWELM) etc. are designed to solve the class imbalance problem effectively [7].

Many real-life problems can be described as imbalanced classification problems, where the number of samples belonging to one of the classes is heavily outnumbered than the numbers in other classes. The samples with larger and smaller class proportion are referred to as the majority and the minority class respectively [7]. ELM randomly initializes the weights between the input and the hidden layer. These random weights are used to map the input data to the feature space. These weights remain unaltered amid the training phase. Due to this some of the samples usually at the classification boundary are misclassified in certain realizations. The methods available for the imbalanced classification problems [8] can be broadly categorized as the data level methods, the algorithmic level methods and the cost sensitive methods. The data level methods like oversampling and under sampling [9-10] alter the data space to reduce the impact of the class imbalance. The under-sampling method randomly selects a fraction of data from the majority class examples and balances the data distribution at the cost of information loss. For example, Easy Ensemble and Balance Cascade [11] algorithms employ under sampling for balancing the dataset. The oversampling method randomly duplicates the examples of the minority classes in order to increase the cardinality of the minority class. This may lead to over-fitting. The informed oversampling approach, on the other hand, generates synthetic minority class examples to balance the class distribution like synthetic minority over-sampling techniques (SMOTE) [12]. The algorithmic level methods [13-14] directly modify the classifier design to address the imbalanced learning. The cost-sensitive methods assign more penalty for misclassifying the minority class examples with respect to the majority class examples, that is, misclassification of minority class examples bears more cost [15].

The major bottleneck for ELM is random selection of hidden nodes and input weights that hamper the performance of ELM. To improve this, our model CSIWO ELM provides a method to optimize hidden neurons and input weights. The following results shows that the effectiveness of hybrid approach CSIWO ELM for improving data classification [16-18].

We hereby proposing the new classification algorithm CSIWO ELM which gives the better classification on heart disease dataset. This paper proposes the quality analysis of new developed algorithm. The following result shows that the developed algorithm gives the good classification result.

The detailed structure of the work is we are modifying the cuckoo search and invasive weed optimization to select the optimised input weight and hidden bias. And this selected input weight and hidden bias will send it to the ELM FFNN to improve the classification result. And in this paper we are analysing the quality of the classification result using Linear function based analysis, objective function based analysis and optimization function based analysis on Cleveland dataset, Switzerland dataset and Hungarian dataset. The major contribution of this paper are:

- An improved CSIWO ELM method introduced to optimize the hidden nodes and input weights.
- Forming of mathematical equation of the proposed method.
- Presented the result which shows effectiveness of the proposed model.
- Promoted Statical analysis of the proposed model which gives the validation of the result.
- Quality analysis of the proposed model.

## 2. Modified CSIWO ELM-based FFNN Approach for Data Classification

This section describes the developed CSIWO ELM-based FFNN for the classification of data. There are three phases which, involves Pre-Processing, Selection of Feature, and Classification of data. At first, the input data is subjected to pre-processing phase for missing value imputation and the transformation is done using exponential kernel. Then, the feature selection process is carried out using Jaro-Winkler distance for selecting the significant features for classifying the medical data. Then, the selected features are provided to FFNN for further classification. Here, the proposed modified CS-IWO is designed by combining CS and IWO algorithm. Thus, the optimal solution for performing the classification is obtained using FFNN. Figure 1 depicts the block diagram of developed modified CSIWO ELM-based FFNN for data classification. Consider the database used for obtaining the input data as,  $H$ . The database is represented as,

$$H = \{H_{mn}\}; (1 \leq m \leq A); (1 \leq n \leq B) \quad (1)$$

where, the total data and the number of features is denoted as,  $A$  and  $B$  respectively. The dataset is of the

size  $[A \times B]$ , The  $m$ th data in  $n$ th feature is denoted as,  $H_{mn}$ . The exponential kernel function is given by the below expression as,

$$K = \left( 1 - \frac{\|P - R\|}{2d^2} \right) \quad (2)$$

where, the smoothing factor and constant is represented by the term,  $d$  and  $R$ .

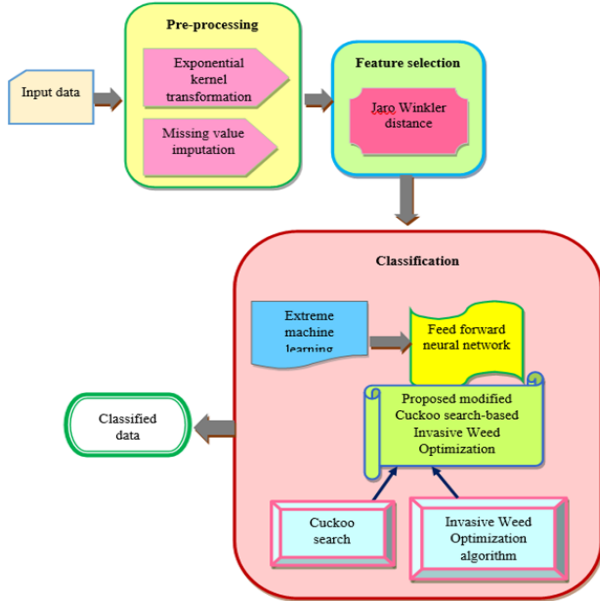


Figure.1. Block diagram of developed data classification using developed modified CSIWO ELM-based FFNN

The exponential kernel transformation improves the data classification process by improving the computation speed and simplifying the process of computation [4-7]. The output data from the pre-processing process is denoted as,  $P_H$ . The similarity between the two features  $h_1$  and  $h_2$  Jaro-Winkler distance is expressed as,

$$p = f(h_1, h_2) + (\beta \cdot \lambda (1 - f(h_1, h_2))) \quad (3)$$

$$f(h_1, h_2) = \begin{cases} 0 & ; \text{if } s = 0 \\ \frac{1}{3} \left( \frac{s}{|k_1|} + \frac{s}{|k_2|} + \frac{s-w}{s} \right) & ; \text{Otherwise} \end{cases} \quad (4)$$

where, Jaro similarity metric and matching sequences are represented as,  $f$  and  $S$ , respectively. The number of transposition and common prefix length is denoted as,  $W$  and  $\beta$ , respectively. The length of two sequences are given as,  $k_1$  and  $k_2$ , respectively. The scaling factor is represented as,  $\lambda$ . The suitable features for the data selection process are compared using the Jaro-Winkler distance. The output data obtained from the feature selection process is represented as,  $W = \{W_1, W_2, \dots, W_l\}$ . For further processing, the selected features are provided to the data classification module.

## 2.1. Data classification using modified CSIWO ELM-based FFNN

This section explains the developed modified CSIWO ELM-based FFNN method used for the classification of data. The FFNN is trained using the CSIWO ELM for the classification of data. The input to the FFNN classifier is the output data obtained from the feature selection process. The developed CSIWO ELM method is the integration of ELM with hybrid optimization algorithm. The CSIWO is developed by integrating the CS algorithm in IWO. The CS algorithm helped in optimization of the hidden neurons and input weights of ELM. The pairwise competition within the particles is randomly selected from the swarm in the CS. The generation, whereas the loser of the competition is updated generation, whereas the loser of the competition is updated and transferred to the next generation. The CS provided better the problems in high dimension [19-21]. The IWO depends on the colonization behaviour is invasive weeds. The IWO provided efficient solution against the different values of the parameter. Hence, the CS and IWO algorithms are integrated to provide better solution to the optimization problem. The architecture and training process of FFNN are described below. Let us consider the dataset  $H$  with  $Y$  samples is given as,  $(W_m, Y_m)$ , where,  $m = 1, 2, \dots, N$ , the input and target vector is represented as,  $W_m$  and  $Y_m$ , respectively. The FFNN contains activation function  $q(W)$  and hidden layer with  $E$  neurons. The input and target vector for training FFNN is represented as,  $W_m = [W_{m1}, W_{m2}, \dots, W_{my}]^T$  and  $Y_m = [Y_{m1}, Y_{m2}, \dots, Y_{mx}]^T$ . The FFNN output is represented as,

$$\bar{L}_j = \sum_{j=1}^E \rho_j \cdot q(a_j \cdot W_m + e_j), j = 1, \dots, N \quad (5)$$

where, the weight vector connecting the input with the output node is represented as,  $a_j = [a_{j1}, a_{j2}, \dots, a_{jy}]^T$ . In  $j$ th hidden node, the bias is represented as,  $e_j$ . The hidden and output nodes are connected with the output weights using the set of values,

such as,  $\rho_j = [\rho_{j1}, \rho_{j2}, \dots, \rho_{jx}]^T$ . The  $N$  equation are represented as,  $Q\rho = T$ , where

$$Q(a_1, a_2, \dots, a_E, e_1, e_2, \dots, e_E, W_1, W_2, \dots, W_N) = \begin{pmatrix} q(a_1 \cdot W_1 + e_1) \dots q(a_E \cdot W_1 + e_E) \\ \vdots \\ q(a_1 \cdot W_N + e) \dots q(a_E \cdot W_N + e_E) \end{pmatrix}_{N \times E} \quad (6)$$

$$\rho = [\rho_1^T \dots \rho_E^T]_{E \times x}^T \quad (7)$$

$$T = [u_1^T, \dots, u_N^T]_{N \times x}^T \quad (8)$$

The minimum norm least square solution is determined for obtaining the output weight matrix,  $\rho$ . The minimum norm least square solution is given as,

$$\bar{\rho} = \min_{\rho} \|Q\rho = T\| = Q^+ T \quad (9)$$

$$\bar{\rho} = (Q^T Q)^{-1} Q^T T \quad (10)$$

where, the output matrix of hidden layer is represented as,  $Q$  and MP pseudo inverse of  $H_{is}$  is denoted as,  $Q^+$ . The output weight is determined by the regularized ELM for ensuring invertibility using the equation as,

$$\bar{\rho} = (\bar{C}I + Q^T Q)^{-1} Q^T T \quad (11)$$

where, the identity matrix and regularization matrix are characteristics of specific species. represented as,  $I$  and  $\bar{C}$ , respectively. The FFNN classifier is trained for the classification of data by the optimal solution provided by the developed modified CSIWO algorithm.

### Training process of FFNN using introduced modified CSIWO ELM.

This The training process of FFNN based on devised modified CSIWO approach is explicated in this section, and its algorithmic steps are represented as below, [22-25]

#### 1. Modified CS algorithm

The modified CS approach is explicated in below sections.

#### Algorithm 1. Pseudocode of modified CS method

S.No.	Pseudocode of modified CS technique
1	Objective function $f(G), G = (G_1, \dots, G_c)^E$
2	Initialize the population of $j$ host nests $G_h (h = 1, 2, \dots, j)$
3	while ( $b <$ Stopping condition )
4	Randomly get a cuckoo (say $g$ ) by Levy flights using equation (16)
5	Compute the fitness value using Fitness function expression (13)
6	Select the nest amongst $j$ (say $u$ ) arbitrarily
7	if ( $F_g > F_u$ )
8	Replace $u$ by new solution
9	end
10	Abandon a fraction of worse nests and generate new ones at new location by Levy flights
11	Keep the optimal solution
12	Rank the solutions and find current optimum
13	end while

The algorithm 1 shows that modified approach of cuckoo search, this algorithm present the how modification will improve the performance of the model.

following are the steps in CS algorithm.

#### i) Initialization :

The population of CS approach are initialized arbitrarily. The elements, such as input weights and hidden biases are optimized in a particle. Moreover, candidate ELM is specified through the hidden biases as well as input and the elements are initialized within the range of  $[-1, 1]$ . The solution  $G(\delta)$  at iteration  $\delta$  is illustrated as,

$$G(\delta) = \{J_1(\delta), \dots, J_v(\delta)\}, J_h(\delta) = \{T_{h1}, \dots, T_{hk}\} \quad (12)$$

where,  $J$  denotes the feasible solution  $k$  is attained amongst solution set. The population of  $V$  particles are randomly initiated and updated.

#### ii) Fitness function computation :

The fitness measure is estimated for finding optimal solution and it should be minimal, which is calculated by following expression.

$$\chi = \min(1 - (LF + B + \psi + \mu)) \quad (13)$$

where,  $LF$  indicates loss function,  $\chi$  implies fitness function,  $\psi$  denotes Mean Square Error (MSE),  $B$  signifies accuracy, and  $\mu$  refers entropy. The accuracy is a closeness measure, which is calculated by,

$$B = \frac{M + N}{M + O + N + P} \quad (14)$$

where,  $M$  is true positive,  $N$  indicates true negative,  $O$  symbolizes false positive and  $P$  specifies false negative. In least square solution is given as, addition, MSE is calculated by,

$$\psi_c = \frac{1}{q} \sum_{j=1}^q (E_q^* - E_q)^2 \quad (15)$$

where,  $\psi_c$  represents fitness function,  $q$  indicates total indicates total amount of samples,  $E_q^*$  refers target output and  $E_q$  denotes the final output of FFNN.

#### iii) Computation of updation equation

Here, the CS and IIWO is modified for introducing modified CSIWO model. Usually, CS approach is the breed characteristics of specific species. The velocity updation equation in CS method is expressed in below expression. Here, the step size  $\kappa$  is made iterative and it is specified as,

$$G_h^{\delta+1} = G_h^{\delta} + \kappa^{\delta+1} \oplus \text{Levy}(\eta) \quad (16)$$

$$\kappa^{\delta+1} = \max\left(\frac{\kappa^{\delta}}{w}\right) \quad (17)$$

where,  $G_h^\delta$  denotes solution at current iteration, symbolizes entry wise multiplication. The distance  $W$  given by,

$$w = \left\| r_{m_x} - r_{m_y} \right\| \quad (18)$$

where,  $r_{m_x}$  is fitness of solution  $m_x$ , which has best value, and  $r_{m_y}$  denotes fitness of solution  $m_y$ , that has second best value. Additionally, Levy flights are significant to afford random walk, when their random steps are drawn from Levy distribution for huge steps.

$$Levy \sim t = b^{-\eta} (1 < \eta \leq 3) \quad (19)$$

The above equation gives the levy flight equation.

#### iv) Re-evaluation of fitness function :

The fitness measure is computed for each iteration, and the fitness measure with maximal value is taken as optimal solution.

#### v) Termination :

The above processes are repeated until the best optimal solution is obtained.

### 2. Modified IWO algorithm

The general IWO model is population driven optimization approach, which is motivated by weed colonies characteristics. Here, instead of IWO method, improved IWO technique is employed.

#### i) Initialization :

At first, population of plants are formulated by means of chaotic mapping, which is specified as,

$$G = \{G_1, G_2, \dots, G_e, \dots, G_d\} \quad (20)$$

where,  $d$  signifies total amount of solution and  $G_e$  denotes  $e$  th solution.

#### ii) Calculation of fitness function :

The fitness measure is estimated for identifying optimum solution. The fitness function with minimal value is taken as best possible solution, which is already expressed above in fitness function equation.

#### iii) Update the solution :

After the computation of fitness function, solution updation is performed using modified IWO technique. The standard equation of Improved IWOA for obtaining best position is

$$G_h^{\delta+1} = \gamma(\delta)G_h^\delta + (G_{\text{best}} - G_h^\delta) \quad (21)$$

where,  $G_h^{\delta+1}$  refers new weed in  $h^{\text{th}}$  iteration,  $G_{\text{best}}$  signifies best weed in entire population and  $\gamma(\delta)$  symbolizes standard deviation.

$$\gamma(\delta) = \left( \frac{L - \delta}{L} \right)^{k^{\delta+1}} (\gamma_{\text{initial}} - \gamma_{\text{final}}) + \gamma_{\text{final}} \times g(\delta) \quad (22)$$

Here,  $g(\delta)$  is equal to chaotic mapping in  $\delta$  th iteration.

Here, the modulation index  $k$  is modified as,

$$k^{l+1} = \max^* \left( \frac{k^l}{w} \right) \quad (23)$$

where, max indicates maximal value of  $k$  found till modulation index,  $k^l$  represents current iteration, and  $w$  specifies distance, which is already represented in denoted in above distance equation.

#### iv) Verify feasibility of solution:

The optimum solution is computed using the fitness measure and if a new solution is better than previous solution, then update a value with new one.

#### v) Termination:

The above processes are continued until the better solution is obtained.

### Algorithm 2. Pseudocode of modified IWO method

S.No.	Pseudocode of modified IWO technique
1	Input: Population of plants
2	Output: Best solution
3	Start
4	Initialize the population of plant
5	Estimate the fitness function based on fitness function equation (13)
6	while ( $h < \text{max imum amount of weed}$ )
7	Update the solution by means of expression
8	end while
9	Verify feasibility of solution
10	Replace the optimal solution
11	End

The algorithm 2 shows the modified approach of invasive weed optimization. This modification helps in improving model performance [36-40].

## 3. Result and Discussion

This section describes the results and discussion of proposed modified CSIWO ELM-based FFNN with respect to evaluation metrics by varying the training data percentage.

### 3.1. Experimental setup

The experimentation of proposed modified CSIWO ELM based FFNN approach is done in PYTHON tool. The dataset utilized for demonstration of proposed modified CSIWO ELM-based FFNN is heart disease dataset, which consists of Cleveland, Switzerland, and Hungarian dataset from UCI machine learning repository. This dataset is a type of multivariate dataset, and it is mainly used for the purpose of classification tasks. This dataset consists of 76 attributes out of which

only 14 attributes are used. The presence of heart disease is represented utilizing the integer from 0 to 4. The number of attributes existed in proposed modified CSIWO ELM-based FFNN approach is 303.

### 3.2. Comparative analysis

This section explains the comparative analysis of proposed chaotic IWO+FFNN-tangential based on three functions, namely linear, objective, and optimization using three datasets. Moreover, the techniques employed for comparison are Cuckoo search + FFNN, Chaotic IWO + FFNN, IWO-CS ELM-exponential, CSIWO ELM-based FFNN-tangential, and Modified CS + FFNN-exponential.

#### Linear function-based analysis:

We can analyse the model by first representing it as a linear function and then interpreting the components of the function. As long as we can figure out, the initial value and the rate of change of a linear function. Initial value is a term that is typically used in applications of functions. It can be represented as the starting point of the relationship we are describing with a function. In the case of linear functions, the initial value is typically the  $y$ -intercept. This analysis gives us the quality in the solution[41-42].

This section illustrates the analysis of proposed modified chaotic IWO+FFNN-tangential based on linear function in terms of performance measures using three datasets. Linear activation function is also referred as identity function in which the activation is directly proportional to the input. Typically, linear function is constrained with two major limitations. There is no possibility to employ backpropagation since the derivative of the function is always a stable value and it has no similarity to the input. The second limitation is that all layers in the neural network will aggregate into one layer if a linear activation function is utilized. Specifically, a linear function converts the neural network into a single layer. Here, linear activation is analyzed based on two types, such as exponential and tangential function[43].

#### (i) Analysis using Cleveland dataset

Figure 2 represents the analysis of proposed modified CSIWO ELM-based FFNN in terms of accuracy. When the training data = 90%, the accuracy obtained by proposed modified CSIWO ELM-based FFNN is 0.947, whereas the existing approaches attained the accuracy of 0.883 for cuckoo search + FFNN, 0.898 for chaotic IWO + FFNN, 0.901 for modified CS + FFNN, 0.913 for modified chaotic IWO + FFNN, 0.929 for IWO-CS ELM, and 0.938 for CSIWO ELM-based FFNN. However, the proposed scheme shows the performance enhancement of proposed approach with that of conventional techniques, such as cuckoo search + FFNN is 6.715%, chaotic IWO + FFNN is 5.175%, modified

CS + FFNN is 4.808%, modified chaotic IWO + FFNN is 3.601%, IWO-CS ELM is 1.854%, and CSIWO ELM-based FFNN is 0.925%.

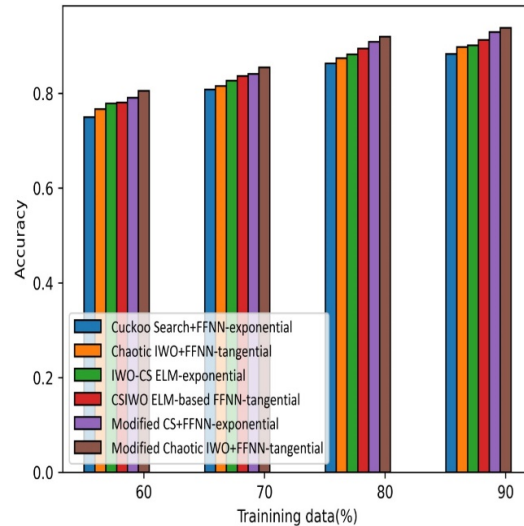


Figure 2. Analysis based on linear function using accuracy on Cleveland

Fig.3.depicts the comparative analysis of developed modified CSIWO ELM-based FFNN using precision. When 90% training data is considered, the precision acquired by proposed modified CSIWO ELM-based FFNN is 0.902, while the existing methods achieved the precision of 0.837 for cuckoo search + FFNN, 0.851 for chaotic IWO + FFNN, 0.862 for modified CS + FFNN, 0.872 for modified chaotic IWO + FFNN, 0.881 for IWO-CS ELM, and 0.889 for CSIWO ELM-based FFNN. However, the proposed scheme shows the performance enhancement of proposed approach with that of conventional techniques, such as cuckoo search + FFNN is 7.348%, chaotic IWO + FFNN is 5.794%, modified CS + FFNN is 4.567%, modified chaotic IWO + FFNN is 3.397%, IWO-CS ELM is 2.402%, and CSIWO ELM-based FFNN is 1.495%

In the same way we can calculate and draw the sensitivity and specificity based on the linear function of Cleveland dataset, the values are mentioned in the Table1.

#### (ii) Analysis using Switzerland dataset

Figure 4 shows the comparative analysis of proposed modified CSIWO ELM-based FFNN in terms of accuracy using Switzerland dataset. By considering the training data percentage as 90%, accuracy obtained by proposed modified CSIWO ELM-based FFNN is 0.957. Results the performance improvement of proposed scheme with that of existing methods, like cuckoo search + FFNN is 6.498%, chaotic IWO + FFNN is 5.399%, modified CS + FFNN is 3.981%, modified

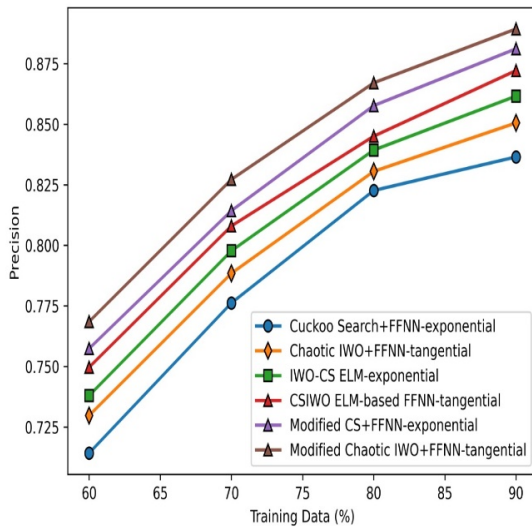


Figure 3. Analysis based on linear function using Precision on Cleveland

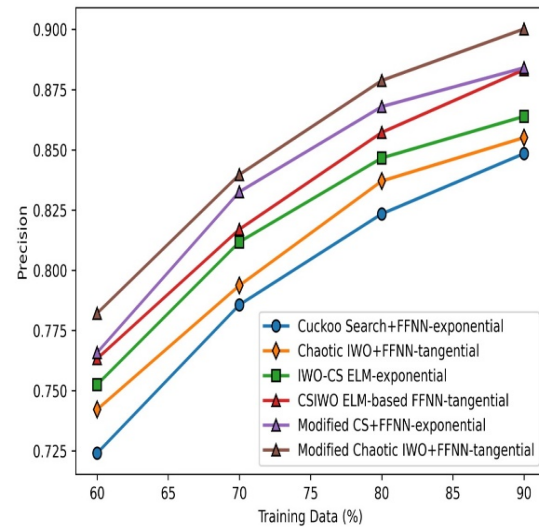


Figure 5. Analysis based on linear function using Precision on Switzerland

chaotic IWO + FFNN is 3.178%, IWO-CS ELM is 2.425%, and CSIWO ELM-based FFNN is 1.267%. However, the classical schemes achieved the accuracy of 0.895 for cuckoo search + FFNN, 0.906 for chaotic IWO + FFNN, 0.919 for modified CS + FFNN, 0.927 for modified chaotic IWO + FFNN, 0.934 for IWO-CS ELM, and 0.945 for CSIWO ELM-based FFNN.

data considered, the precision acquired by proposed modified CSIWO ELMbased FFNN is 0.9083, while the existing methods achieved the precision of 0.8484 for cuckoo search + FFNN, 0.8550 for chaotic IWO + FFNN, 0.8638 for modified CS + FFNN, 0.8832 for modified chaotic IWO + FFNN, 0.8840 for IWOCs ELM, and 0.9002 for CSIWO ELM-based FFNN. However, the proposed scheme shows the performance enhancement of proposed approach with that of conventional techniques, such as cuckoo search + FFNN is 6.592%, chaotic IWO + FFNN is 5.863%, modified CS + FFNN is 4.893%, modified chaotic IWO + FFNN is 2.761%, IWO-CS ELM is 2.670%, and CSIWO ELM-based FFNN is 0.896%. Similarly, we have calculated the sensitivity and specificity of the Switzerland dataset. And accordingly, we have drawn the accuracy, sensitivity, specificity and precision graph for all the dataset i.e., Cleveland, Switzerland and Hungarian for Linear function-based analysis.

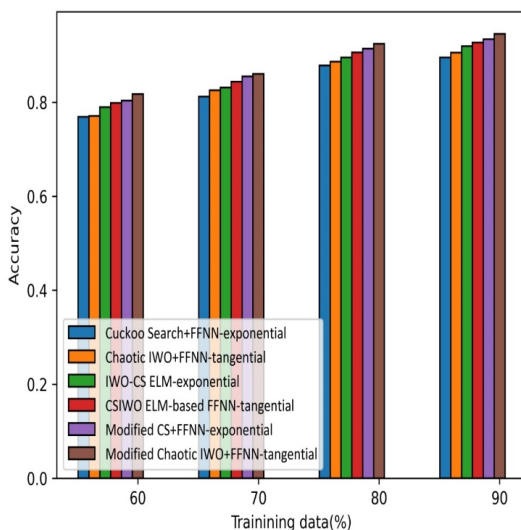


Figure 4. Analysis based on linear function using accuracy on Switzerland

**Objective function-based analysis:**

Figure 5 depicts the precision-based analysis of developed modified CSIWO ELM-based FFNN for varying percentage of training data. For 90% training

The objective function may involve plugging the candidate solution into a model and evaluating it against a portion of the training dataset, and the cost may be an error score, often called the loss of the model. In order to find the optimal solution, we need some way of measuring the quality of any solution. This function, taking data and model parameters as arguments, can be evaluated to return a number. Any given problem contains some parameters which can be changed; our goal is to find values for these parameters which maximize this number. In this paper we are analysing quality of the solution using objective function based on the Cleveland, Switzerland, and Hungarian dataset.

The following effective result shows that our proposed modified CSIWO ELM based FFNN selects the optimal parameter and evaluate our model solution against the training dataset.

#### (i) Analysis using Cleveland dataset

Figure 6 represents the analysis of proposed modified CSIWO ELM-based FFNN in terms of accuracy. When the training data = 90%, the accuracy obtained by proposed modified CSIWO ELM-based FFNN is 0.945, whereas the existing approaches attained the accuracy of 0.885 for cuckoo search + FFNN, 0.896 for chaotic IWO + FFNN, 0.904 for modified CS + FFNN, 0.910 for modified chaotic IWO + FFNN, 0.933 for IWO-CS ELM, and 0.936 for CSIWO ELMbased FFNN. However, the proposed scheme shows the performance enhancement of proposed approach with that of conventional techniques, such as cuckoo search + FFNN is 6.285%, chaotic IWO + FFNN is 5.188%, modified CS + FFNN is 4.296%, modified chaotic IWO + FFNN is 3.656%, IWO-CS ELM is 1.274%, and CSIWO ELM-based FFNN is 0.855%.

Figure 7 illustrates the analysis of proposed modified CSIWO ELM-based FFNN with respect to precision. With 90% of the training data taken into consideration, the precision attained by developed modified CSIWO ELMbased FFNN is 0.906, while the existing methods attained the precision of 0.839 for cuckoo search + FFNN, 0.847 for chaotic IWO + FFNN, 0.864 for modified CS + FFNN, 0.869 for modified chaotic IWO + FFNN, 0.878 for IWO-CS ELM, and 0.888 for CSIWO ELM-based FFNN. Also, the proposed scheme showed performance improvement with that of conventional techniques, like cuckoo search + FFNN is 7.37%, chaotic IWO + FFNN is 6.41%, modified CS + FFNN is 4.61%, modified chaotic IWO + FFNN is 4.07%, IWO-CS ELM is 3.02%, and CSIWO ELM-based FFNN is 1.96%.

Similarly, we have calculated the sensitivity and specificity of the Switzerland as well as Hungarian dataset. And accordingly, we have drawn the accuracy, sensitivity, specificity and precision graph for all the dataset i.e., Cleveland, Switzerland and Hungarian for Objective function-based analysis. All calculated values is reflected in Table1.

#### Optimization function-based analysis:

Optimization function is the of finding a set of inputs to an objective function that results in a maximum function evaluation. An optimization algorithm is used to find the values of the parameters that minimize the error of the function when used to map inputs to outputs. We are performing the optimization function-based analysis on Cleveland, Switzerland, and Hungarian dataset to check the set of input that can perfectly fit to objective function to maximize function evaluation. The proposed modified CSIWO ELM based FFNN selects the optimal parameters which minimize the error and gives greater results on all datasets.

#### (i) Analysis using Cleveland dataset

Figure 8 represents the analysis of proposed modified CSIWO ELM-based FFNN in terms of accuracy. When the training data = 90%, the accuracy obtained by proposed modified CSIWO ELM-based FFNN is 0.914, whereas the existing approaches attained the accuracy of 0.861 for cuckoo search + FFNN, 0.867 for chaotic IWO + FFNN, 0.876 for modified CS + FFNN, 0.887 for modified chaotic IWO + FFNN, 0.899 for IWO-CS ELM, and 0.909 for CSIWO ELMbased FFNN. However, the proposed scheme shows the performance enhancement of proposed approach with that of conventional techniques, such as cuckoo search + FFNN is 5.827%, chaotic IWO + FFNN is 5.207%, modified CS + FFNN is 4.204%, modified chaotic IWO + FFNN is 2.977%, IWO-CS ELM is 1.697%, and CSIWO ELM-based FFNN is 0.553%.

Analysis of the proposed modified CSIWO ELM-based FFNN with Cleveland dataset is depicted in figure 9 using precision. The value of precision of the proposed modified CSIWO ELM-based FFNN method when 0% of training data taken into account is 0.899, while the precision of the other method are 0.839, 0.851, 0.854, 0.863, 0.880 and 0.885 for the cuckoo search + FFNN, chaotic IWO + FFNN, modified CS + FFNN, modified chaotic IWO + FFNN, IWO-CS ELM, and CSIWO ELM-based FFNN respectively. The proposed method is also shown to produce an improvement in performance when compared to the other techniques such as of 6.65% with cuckoo search + FFNN, 5.23% with chaotic IWO + FFNN, 4.97% with modified CS + FFNN, 3.93% with modified chaotic IWO + FFNN, 2.02% with IWO-CS ELM, and 1.51% with CSIWO ELM-based FFNN.

Similarly, we have calculated the sensitivity and specificity of the Cleveland dataset. And accordingly, we have drawn the accuracy, sensitivity, specificity and precision graph for all the dataset i.e. Cleveland, Switzerland and Hungarian.

Similarly, we have calculated the sensitivity and specificity of the Switzerland as well as Hungarian dataset. And accordingly, we have drawn the accuracy, sensitivity, specificity and precision graph for all the dataset i.e., Cleveland, Switzerland and Hungarian for Optimization function-based analysis. All calculated values is reflected in table 1.

## 4. Comparative discussion

Table1 portrays the comparative discussion of proposed modified CSIWO ELM-based FFNN with respect to evaluation metrics using three datasets, namely Cleveland dataset, Switzerland dataset, and Hungarian dataset. From the discussion, it is very vivid that the proposed modified CSIWO ELM-based FFNN has achieved a maximum accuracy, sensitivity and



**Table 1.** Comparative discussion

Subjects	Met-rics/Methods	Cuckoo search + FFNN-exponential	Chaotic IWO + FFNN-tangential	IWO-CS ELM-exponential	CSIWO ELM-based FFNN-tangential	modified CS + FFNN-exponential	Proposed Modified chaotic IWO + FFNN-tangential
Cleveland dataset (Linear)	Accuracy	0.883	0.898	0.901	0.913	0.929	0.938
	Sensitivity	0.843	0.857	0.868	0.879	0.888	0.896
	Specificity	0.870	0.877	0.883	0.893	0.902	0.913
Switzerland dataset (Linear)	Accuracy	0.895	0.906	0.919	0.927	0.934	0.945
	Sensitivity	0.873	0.883	0.898	0.905	0.916	0.925
	Specificity	0.850	0.864	0.877	0.887	0.893	0.901
Hungarian dataset (Linear)	Accuracy	0.872	0.880	0.897	0.907	0.916	0.921
	Sensitivity	0.839	0.849	0.851	0.869	0.871	0.887
	Specificity	0.850	0.864	0.877	0.887	0.893	0.901
Cleveland dataset (objective)	Accuracy	0.885	0.896	0.904	0.910	0.933	0.936
	Sensitivity	0.846	0.855	0.871	0.876	0.885	0.895
	Specificity	0.871	0.876	0.884	0.892	0.900	0.910
Switzerland dataset (objective)	Accuracy	0.896	0.902	0.920	0.926	0.936	0.943
	Sensitivity	0.857	0.860	0.875	0.891	0.896	0.905
	Specificity	0.874	0.880	0.898	0.903	0.918	0.927
Hungarian dataset (objective)	Accuracy	0.875	0.879	0.899	0.902	0.920	0.920
	Sensitivity	0.840	0.846	0.852	0.866	0.875	0.883
	Specificity	0.854	0.861	0.879	0.884	0.892	0.901
Cleveland dataset	Accuracy	0.861	0.867	0.876	0.887	0.899	0.909
	Sensitivity	0.846	0.858	0.861	0.870	0.887	0.892
	Specificity	0.860	0.870	0.884	0.895	0.902	0.911

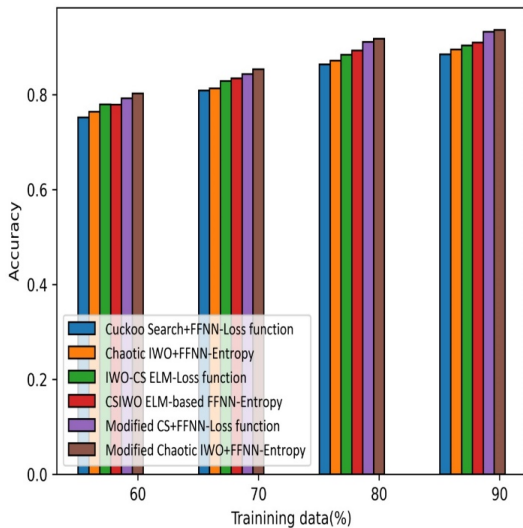


Figure 6. Analysis based on objective function using accuracy on Cleveland.

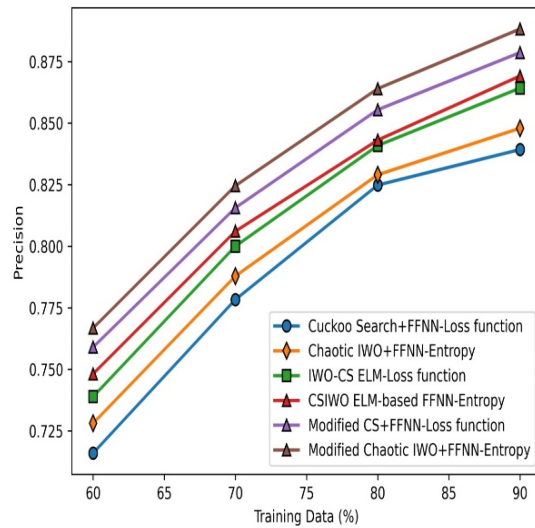


Figure 7. Analysis based on objective function using Precision on Cleveland.

specificity of 0.957, 0.915, and 0.939 for linear function using Switzerland dataset. Similarly, while considering objective function, the proposed modified CSIWO

ELM-based FFNN has obtained maximum accuracy, maximum sensitivity, and maximum specificity with the values of 0.959, 0.920, and 0.940, respectively. While considering the optimization-based function,

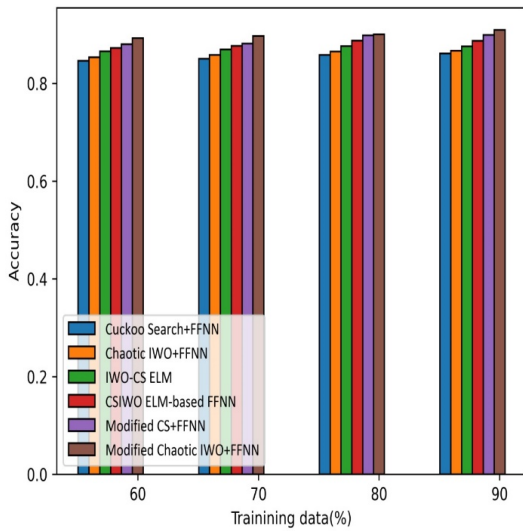


Figure 8. Analysis based on optimization function using accuracy on Cleveland

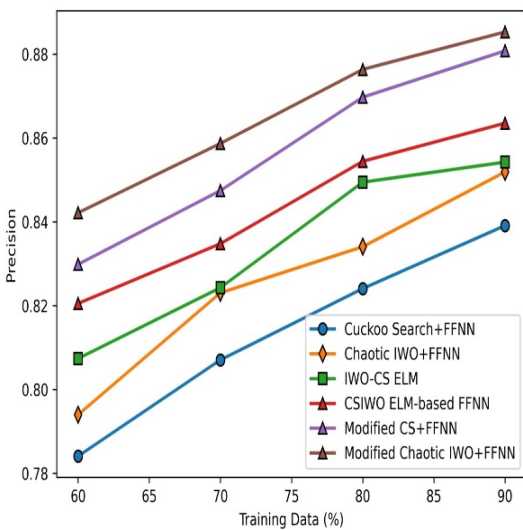


Figure 9. Analysis based on optimization function using Precision on Cleveland

the proposed modified CSIWO ELM-based FFNN has obtained maximum accuracy of 0.951, maximum sensitivity of 0.958, and maximum specificity of 0.950 [32-33].

#### 4.1. Statistical analysis

Pair-wise statistical tests are performed to assess the algorithmic performance by combining various algorithms. Here, statistical test is performed on the accuracy, sensitivity, and specificity values.

Table 2 depicts the t values of different methods based on accuracy. For the hypothesis testing, t value should be less than 0.1. In table 2, the proposed CSIWO ELM-based FFNN discards null hypothesis in most of the combinations by reaching the value of 0.0227. The table depicts that the statistical test almost always returns lower t values for the proposed CSIWO ELM-based FFNN than for other algorithms and more often rejects the null hypothesis. Table 3 represents the pair-wise statistical test of algorithms based on sensitivity. The methods considered for the comparison are analyzed with the proposed CSIWO ELMbased FFNN to find the performance deviation of the methods using statistical test. In Table 3, the proposed CSIWO ELMbased FFNN discards null hypothesis by reaching the t value 0.0254 as compared with the existing methods. Table 4 represents the pair-wise statistical test of algorithms based on specificity. In Table 4 , the proposed CSIWO ELM-based FFNN discards null hypothesis by reaching the t value 0.0203 as compared with the existing methods.

### 5. Conclusion and Future Scope

In this paper, a novelty ELM method is named Hybrid Intelligent Feed Forward Neural Network Extreme Learning Machine based on modified Cuckoo Search and modified Invasive weed optimization (CSIWO – ELM) method is proposed. The main features of this model to optimize the hidden neurons and input weights. The modified CS and IWO selects the input weights and hidden neurons which effectively gives the greater result according to the proposed model. The proposed CSIWO – ELM model analysed using evaluation metrics for three functions, such as linear-based function, objective-based function, and optimization-based function using three datasets. Moreover, the developed modified CSIWO ELM based FFNN has achieved maximum accuracy of 0.959, maximum sensitivity of 0.920, and maximum specificity of 0.940 for Switzerland dataset based on objective function. Also statical test is performed on the model which also shows the performance of the proposed model.

The future improvisation of developed modified CSIWO ELM-based FFNN would be the inclusion of more advanced classifiers with some other hybrid optimization for accurate classification of data. Also, it challenging to see the use of ELM in the deep learning technique.

**Acknowledgement.** We acknowledge Principal and Management of Mcts Rajiv Gandhi Institute of Technology Mumbai, India for constant support. Also wanted to thanks to Dr. Sunil Wankhade for motivation.

**Table 2.** Pair-wise statistical test of methods on Accuracy

Subjects	Cuckoo search + FFNN-exponential	Chaotic IWO + FFNN-tangential	IWO-CS ELM-exponential	CSIWO ELM-based FFNN-tangential	modified CS + FFNN-exponential	Modified chaotic IWO + FFNN-tangential
Cuckoo search + FFNN-exponential	-	0.1527	0.6156	0.6062	0.2846	0.0362
Chaotic IWO + FFNN-tangential	0.1527	-	0.6283	0.6325	0.3648	0.0227
IWO-CS ELM-exponential	0.6156	0.6283	-	0.6138	0.2496	0.0325
CSIWO ELM-based FFNN-tangential	0.6062	0.6325	0.6138	-	0.3721	0.0318
modified CS + FFNN-exponential	0.2846	0.3648	0.2496	0.3721	-	0.0256
Modified chaotic IWO+FFNN-tangential	0.0362	0.0227	0.0325	0.0318	0.0256	-

**Table 3.** Pair-wise statistical test of methods on sensitivity

Subjects	Cuckoo search + FFNN-exponential	Chaotic IWO + FFNN-tangential	IWO-CS ELM-exponential	CSIWO ELM-based FFNN-tangential	modified CS + FFNN-exponential	Modified chaotic IWO + FFNN-tangential
Cuckoo search + FFNN-exponential	-	0.173	0.6427	0.6381	0.5432	0.0315
Chaotic IWO + FFNN-tangential	0.173	-	0.2272	0.5471	0.4573	0.0417
IWO-CS ELM-exponential	0.6427	0.2272	-	0.6267	0.5436	0.0318
CSIWO ELM-based FFNN-tangential	0.6381	0.5471	0.6267	-	0.4358	0.0321
modified CS + FFNN-exponential	0.5432	0.4573	0.5436	0.4358	-	0.0254
Modified chaotic IWO+FFNN-tangential	0.0315	0.0417	0.0318	0.0321	0.0254	-

**Table 4.** Pair-wise statistical test of methods on specificity

Subjects	Cuckoo search + FFNN-exponential	Chaotic IWO + FFNN-tangential	IWO-CS ELM-exponential	CSIWO ELM-based FFNN-tangential	modified CS + FFNN-exponential	Modified chaotic IWO + FFNN-tangential
Cuckoo search + FFNN-exponential	-	0.1960	0.5426	0.5367	0.4531	0.0217
Chaotic IWO + FFNN-tangential	0.1960	-	0.2269	0.5382	0.4285	0.2427
IWO-CS ELM-exponential	0.5426	0.2269	-	0.5193	0.4268	0.0214
CSIWO ELM-based FFNN-tangential	0.5367	0.5382	0.5193	-	0.3580	0.0209
modified CS + FFNN-exponential	0.4531	0.4285	0.4268	0.358	-	0.0203
Modified chaotic IWO+FFNN-tangential	0.0217	0.2427	0.0214	0.0209	0.0203	-

## 6. References

- [1] Eshay, M., Faris, H. and Obeid, N, "Improving extreme learning machine by competitive swarm optimization and its application for medical diagnosis problems," Expert Systems with Applications, (2018), vol.104, 134-152.
- [2] Huang G-B, Zhu Q-Y, Siew C-K., "Extreme learning machine: theory and applications", Neurocomputing, (2006), vol.70, no.1-3, pp.489-501
- [3] Cheng Y, Zhao D, Wang Y, Pei G., "Multi-label learning with kernel extreme learning machine autoencoder", Knowledge-Based Systems, (2019), vol.178, pp.1-0.
- [4] Werbos PJ., "Generalization of backpropagation with application to a recurrent gas market model", Neural networks, (1988), vol.1, no.4, pp.339-56.
- [5] Huang GB, Chen L, Siew CK., "Universal approximation using incremental constructive feedforward networks with random hidden nodes", Neural Networks, (2006), vol.17, no.4, pp.879-92.
- [6] Huang GB., "An insight into extreme learning machines: random neurons, random features and kernels", Cognitive Computation, (2014), vol.6, no.3, pp.376-90.
- [7] Raghuwanshi BS, Shukla S., "Class imbalance learning using UnderBagging based kernelized extreme learning machine", Neuro computing, (2019), vol.329, pp.172-87.
- [8] Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F., "A review on ensembles for the class imbalance problem: Bagging, boosting and hybrid-based approaches",

- IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), (2012), vol.42, pp.63-484.
- [9] He, H., Garcia, E. A., "Learning from imbalanced data", IEEE Transactions on Knowledge and Data Engineering, (2009), vol.21, pp.263-1284.
- [10] Liu, X. Y., Wu, J., Zhou, Z. H., "Exploratory under-sampling for class imbalance learning", IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics), (2009), vol.39, pp.539-550.
- [11] Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P., "SMOTE: Synthetic minority over-sampling technique", Journal of Artificial Intelligence Research, (2002), vol.16, pp.321-357.
- [12] Cieslak, D. A., Hoens, T. R., Chawla, N. V., Kegelmeyer, W. P., "Hellinger distance decision trees are robust and skew-insensitive", Data Mining and Knowledge Discovery, (2012), vol.24, pp.136-158.
- [13] Tang, Y., Zhang, Y., Chawla, N. V., Krasser, S., "SVMs modeling for highly imbalanced classification", IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics), (2009), vol.39, pp.281-288.
- kernel space of support vector machines", IEEE Transactions on Neural Networks Learning Systems, (2018), pp.1-12.
- [15] Zong, W., Huang, G. -B., Chen, Y., "Weighted extreme learning machine for imbalance learning", Neurocomputing, (2013), vol.101, pp.229-242.
- [16] Ertuğrul, Ö.F. and Kaya, Y., "A detailed analysis on extreme learning machine and novel approaches based on ELM," American Journal of computer science and engineering, (2014), vol.1, no.5, 43-50.
- [17] Shukla S, Raghuvanshi BS., "Online sequential classspecific extreme learning machine for binary imbalanced learning", Neural Networks, (2019), vol.119, 235-48.
- [18] Dai H, Cao J, Wang T, Deng M, Yang Z., "Multilayer one-class extreme learning machine", Neural Networks, (2019), vol.115, 11-22.
- [19] Li H, Yang X, Li Y, Hao LY, Zhang TL., "Evolutionary extreme learning machine with sparse cost matrix for imbalanced learning", ISA transactions, (2020), vol.100, 198-209.
- [20] Cheng Y, Zhao D, Wang Y, Pei G.. "Multi-label learning with kernel extreme learning machine autoencoder", Knowledge-Based Systems, (2019), vol.178, 1-0.
- [21] Cai Z, Gu J, Luo J, Zhang Q, Chen H, Pan Z, Li Y, Li C., "Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf optimization strategy", Expert Systems with Applications, vol.138, (2019) 12814.
- [22] Werbos PJ., "Generalization of backpropagation with application to a recurrent gas market model", Neural networks, (1988), vol.1, no.4, 339-56.
- [23] Huang G, Huang GB, Song S, You K., "Trends in extreme learning machines: A review", Neural Networks, (2015), vol.61, 32-48.
- [24] Kasun LL, Zhou H, Huang GB, Vong CM., "Representational learning with extreme learning machine for big data", IEEE intelligent systems, (2013), vol.28, no.6, 31-4.
- [25] Wang, T., Cao, J., Lai, X., Chen, B., "Deep weighted extreme learning machine", Cognitive Computation, (2018), vol.10, no.6, 890-907.
- [26] Tang, J., Deng, C., Huang, G., "Extreme learning machine for multilayer perceptron", IEEE Transactions on Neural Networks and Learning Systems, (2016), vol.27, no.4, pp.809-821.
- [27] Wong CM, Vong CM, Wong PK, Cao J., "Kernel-based multilayer extreme learning machines for representation learning", IEEE transactions on neural networks and learning systems, (2016), vol.29, no.3, 757-62.
- [28] Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F., "A review on ensembles for the class imbalance problem: Bagging, boosting and hybrid-based approaches", IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), (2012), vol.42, 63-484.
- [14] Mathew, J., Pang, C. K., Luo, M., Leong, W. H., "Classification of imbalanced data by oversampling in [29] He, H., Garcia, E. A., "Learning from imbalanced data", IEEE Transactions on Knowledge and Data Engineering, (2009), vol.21, 263-1284.
- [30] Yang, X.S. and Deb, S., "Engineering optimization by cuckoo search", International Journal of Mathematical Modelling and Numerical Optimization, (2010), vol.1, no.4, 330-343.
- [31] Karimkashi, S. and Kishk, A.A., "Invasive weed optimization and its features in electromagnetics", IEEE transactions on antennas and propagation, (2010), vol.58, no.4, 1269-1278.
- [32] Heart Disease Data Set taken from, "<https://archive.ics.uci.edu/ml/datasets/heart+disease/>" (2020).
- [33] Yang, Dingding, Liming Tang, Shiqiang Chen, and Jun Li. "Image Segmentation Based on Exponential Kernel Function." 14th International Symposium on Pervasive Systems, Algorithms and Networks, (2017) 293-298.
- [34] Cai, W., Yang, J., Yu, Y., Song, Y., Zhou, T. and Qin, J., PSO-ELM: A hybrid learning model for short-term traffic flow forecasting. IEEE access, (2020), 8, 6505-6514.
- [35] Krishnan, G.S. and Kamath, S., A novel GA-ELM model for patient-specific mortality prediction over large-scale lab event data. Applied Soft Computing, (2019), 80, 525-533.
- [36] Nilesh Rathod Sunil Wankhade. "Review of Optimization in Improving Extreme Learning Machine" EAI Endorsed Transactions on Industrial Networks and Intelligent Systems (2021), Vol.8, no.28, DOI: 10.4108/eai.17-92021.170960.
- [37] S. Gaikwad, S. Patel and A. Shetty, "Brain Tumor Detection: An Application Based on Machine Learning," 2021 2nd International Conference for Emerging Technology (INCET), (2021), pp. 1-4, doi: 10.1109/INCET51464.2021.9456347.
- [38] UCI heart disease data set, retrieved from <http://archive.ics.uci.edu/ml/datasets/heart+disease> (2018, September 26), Accessed 20th Jun (2018).
- [39] Spyridon Mourtas, Vasilios Katsikis and Chrysostomos Kasimis " Feedback Control Systems Stabilization Using a Bio-inspired Neural Network" EAI Endorsed Transactions on AI and Robotics (2022), Vol.1, pp.1-13, DOI:<https://doi.org/10.4108/airo.v1i.17>.
- [40] Moshayedi, A. J., Roy, A. S., Kolahdooz, A., Shuxin, Y. "Deep Learning Application Pros And Cons Over Algorithm", EAI Endorsed Transactions on AI and Robotics (2022), Vol.1, PP. 1–13. <https://doi.org/10.4108/airo.v1i.19>.

- [41] Moshayedi, A. J., Shuvam Roy , A., Sambo, S. K. ., Zhong, Y. ., Liao, L. , "Review On: The Service Robot Mathematical Model" . EAI Endorsed Transactions on AI and Robotics (2022), Vol.1, pp. 1–19. <https://doi.org/10.4108/airo.v1i.20>
- [42] Khan, A. T., Cao, X., Li, Z., Li, S. . "Evolutionary Computation Based Real-time Robot Arm Path-planning Using Beetle Antennae Search", EAI Endorsed Transactions on AI and Robotics (2022), Vol.1, pp. 1–10. <https://doi.org/10.4108/airo.v1i.6>
- [43] Nilesh Rathod, Sunil Wankhade,"Optimizing neural network based on cuckoo search and invasive weed optimization using extreme learning machine approach",Neuroscience Informatics,(2022), Vol.2, Issue 3, 100075,ISSN 2772-5286. <https://doi.org/10.1016/j.neuri.2022.100075>.