International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 10, Issue 1, January - February 2021 ISSN 2278-6856

Optimizing Extreme Learning Machine using Whale Optimization Algorithm for Genes Classification

Mustafa Abdul Salam[1, *], Ahmed Taher Azar[2,3] and Rana Hussien[4]

[1]Artificial Intelligence Dept., Faculty of Computers and Artificial Intelligence, Benha University, Egypt

[2,]Scientific Computing Dept., Faculty of Computers and Artificial Intelligence, Benha University, Egypt

[3,]College of Computer & Information Sciences, Prince Sultan University, Riyadh, Kingdom of Saudi Arabia

[4]Scientific Computing Dept., Faculty of Computers and Artificial Intelligence, Benha University, Egypt

Abstract:***This paper presents the hybridization of Whale Optimization Algorithm (WOA) with Extreme Learning Machine (ELM) methodology for solving Gene classification problem. ELM is assumed to be a likely technique for prediction and classification problems. Despite its effectiveness, it needs a large number of nodes on a regular basis for the hidden layer. Using such a huge number of nodes within the hidden layer increases the ELM examination/assessment time. In addition, there is a little guarantee that the layout of weights and biases inside the hidden layer would be optimum. A recent swarm intelligence algorithm (WOA) mimics the conduct of the hunting party of humpback whales is proposed to optimize the ELM model. It is being used within the hidden layer to pick a smaller number of nodes to accelerate the execution of ELM. WOA chooses the optimal weights and bias of the hidden layer. Experimental results show that the proposed hybrid model (WOA-ELM) had better classification accuracy than the standard ELM and SVM.***

Keywords: Whale Optimization Algorithm, Bio-inspired Optimization, Extreme Learning Machine

1. INTRODUCTION

Single hidden layer feed-forward neural network (SLFN) is primarily considered as one of the most popular Learning Machine Models in Classification and Predication areas [1]. The Learning Algorithm is known to be at the core of the Neural Network. Traditional gradient-based machine learning approach such as Levenberg-Marquardt (LM) and Scaled-Conjugate-Gradient (SCG) complains Over-Fitting, Minima local, and long-term memory of waste [3][4]. Extreme Learning Machine (ELM) [5] has been presented to fix complex issues mentioned in gradient based machine learning algorithms. ELM is used as an SLFN learning solution. Although many real-life problems have been arisen, ELM has enormous precision and anticipation speed [6]. ELM hastily chooses weight of inputs and hidden layer bias instead of turning all Inner Parameters just as in Gradient Based Algorithms. ELM also helps provide an impartial description of the weight of the output. According to the erratic choice of weights of input and Hidden layer bias, gradient-based learning algorithms require fewer hidden neurons than ELM [7][8].

Bio-Inspired Algorithms have been used to refine the ELM in order to conquer its obstacles [9]. The Whale Optimizer (WO) algorithm has been used for collection weights of the input and biases to characterize ELM weights of the output. WOA-ELM performed a significant generalization of the simplified structure. WOA-ELM has been used to identify quality motions of patients with distinctive tumor forms and their advanced identification.
The new theoretical WOA ELM model integrating ELM with the Whale Optimizer Algorithm (WOA) is applied to the classification problems discussed in this paper. WOA is proposed to boost ELM inputs and latent bias weights [10].
The rest of this paper is structured as follows: Section 2 discusses the context details for the Extreme Learning Machine (ELM) Model and the Whale Optimizer Algorithm (WOA); Section 3 reflects on the potential approach and the implementation of classification problems; Section 4 discusses the experimental results, while Section 5 points out the core conclusions of the proposed model.

2. RELATED WORKS

Shu et al. [15] suggested an ELM hybrid model based on PSO as PSO Algorithm improved the efficiency of the conventional ELM. Although HuaLing et al. [16] increased the ELM convergence efficiency by combining ELM and enhanced PSO, Parv et al. [17] used an evolutionary approach to build ELM ensembles to control the preference of simple learners in order to have an optimal solution. Wu et al. [18] identified the ELM Genetic Ensemble. Sundararajanc et al. [19] used a genuine new genetic algorithm called 'RCGA-ELM' to pick the best neurons in the hidden layer. He concluded that ELM model's input weights and bias values contribute to better results. Zhao et al. [20] introduced a genetic ELM that is based on the economic distribution of the power grid. Abdul Salam et al. [21] refined the ELM model and improved performance relative to the classic ELM model in the stock market forecast and proposed a new Flower Pollination Algorithm (FPA). Emary et al. [22] highlighted that, in this model, the dragonfly optimizer introduced a hybrid dragonfly algorithm with an extreme predictive learning system that enhanced the ELM model's efficiency in the predictive sector. Parv et al. [23] implemented a technique for the selection of functions based on the Moth-Flame Optimization (MFO) algorithm. The results revealed that the MFO algorithm has relatively outperformed the approaches. Aljara et al. [24] reported that, in most datasets, the proposed WOA-based training algorithm is able to exceed current algorithms; not only in terms of precision, but even in terms of convergence, the findings were greater. Abd El Aziz et al. [25] revealed that in almost all images, in terms of PSNR and SSIM, WOA and MFO algorithms are better than other algorithms at the three-level threshold, regardless of how much WOA is better than MFO. For a greater range of values of limits, Khaled ben Oualid et al. [26] referred that the WOA algorithm appears to be very efficient both in terms of its incremental convergence to the global optimum and in terms of its substantial and precise loss reduction. Mohapatra et al. [27] revealed that ELM-based classifiers display improved performance when projecting higher dimensional space features. However, ELM is combined with CS and ICS in order to achieve a more precise and efficient grouping. The findings suggested that ICSELM is very efficient in minimizing the issue of poor conditions and that this leads to improved outcomes relative to basic ELM, OSELM and CSEL. Deng et al. [28] showed that only if the kernel is purely positive can RKELM approximate some non-linear function with zero error. By missing iterative steps or cursing the scale of the kernel matrix, it is easy to achieve substantial cost savings in the RKELM training process. Large-scale studies on various benchmarks showed that RKELM can produce competitive and stable outcomes at a rapid pace of learning. Wang et al. [9] addressed the effectiveness of the ELM and a revised EELM algorithm was proposed. The proposed algorithm includes an optimum set of input weights and assumptions before output weights are determined, preferably ensuring the maximum column rank of H. This increases the learning rate (precision tests, prediction precision, learning time) and the strength of the networks to a certain degree. Liu et al. [30] suggested the multiple extension of the kernel Extreme Learning Machine to enable heterogeneous databases to be handled through kernel tuning and convergence. Huang et al. [31] shows that the splitting hyperplane continues to pass through the root of the space function of the ELM, resulting in less space. Restrictions on optimization and increased generalization performance Than SVM. It is also noted that there is a generalization of the performance of the ELM that is less susceptible to learning parameters, known as the number of hidden nodes. Zong et al. [32] sowed that the weighted ELM is indicated to be capable of balanced data generalization.Machine learning (ML) methods such as artificial neural networks (ANNs), support vector machines (SVMs), extreme learning machine (ELM), are considered the most commonly used ML models in classification, regression. Also were used in natural language processing (NLP), and text mining such as social media sentiment analysis. But these methods may suffer from local minima and overfitting problems due to using local optimization training algorithms such as gradient descent algorithm in ANN [33]. Swarm Intelligence algorithms such as particle swarm optimization (PSO), follower pollination algorithm (FPA), ant colony optimization (ACO), and artificial bee colony (ABC), can solve the problems or drawbacks of machine learning models such as ANN, SVM, and ELM methods [34,35]. Using swarm intelligence or meta-heuristic algorithms in optimizing and training classical machine learning models can enhance the accuracy and generalization ability of these methods [36-41].

The proposed approach maintains the following benefits of the original ELM: (1) it is practically simple and practical to implement; (2) a wide variety of mapping functions or kernels can be used with the proposed framework; (3) the proposed method can be used specifically for operations of multiclass classification; (4) following integration with the weighting process, the weighted ELM is capable of handling data with imbalanced class distribution while retaining a good output of well-balanced data as an unweighted ELM; (5) the weighted ELM can be generalized to cost-sensitive learning by assigning different weights according to user needs for each case.

3. PRELIMINARIES

A. EXTREME LEARNING MACHINE (ELM)

In 2004, a recent adaptation of the neural system called the extreme learning machine (ELM) was introduced by Huang et al [5]. This unused learning method is structured by a single layer feed-forward, where input weights are randomly picked, and output weights are logically represented using the Moore Penrose Generalized Inverse [14].This definition of Extreme Learning Machines is outlandish from the perspective of the traditional neural network to the understanding that ELMs do not learn input weights that interface the inputs with the hidden layer [22]. While Huang et al. have shown that the least amount of training error could have been accomplished by ELMs, while at the same time having the least weight accuracy. This means the ELMs have a much higher record in generalization. The entire algorithm for the extreme learning machine is illustrated below.

Assume that, given a single secret ELM layer, the output function is functional of $i-th$ hidden node is

$h_i(x)=G(a_i, b_{i,x})$ (1)

where a_i and b_i are the parameters of the $i-th$ hidden node.

For SLFNs with L secret nodes, the ELM's output characteristic **is**

$$f_L(x) = \sum_{i=1}^{L} \beta_i \, h_i(x) \quad (2)$$

where β_i is the output weight if the $i-th$ hidden node.

$\boldsymbol{h(x)} = [\boldsymbol{G}(\boldsymbol{h_i}(\boldsymbol{x}), \ldots, \boldsymbol{h_L}(\boldsymbol{x}))]$. This is ELM's secret mapping of layer output. Provided N preparation samples, the hidden layer output matrix H of the ELM is given as:

$$\mathrm{H} = \begin{bmatrix} \mathrm{h}(\boldsymbol{x_1}) \\ \cdot \\ \cdot \\ \boldsymbol{h}(\boldsymbol{x_N}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}(\boldsymbol{a_1}, \boldsymbol{b_1}, \boldsymbol{x_1}) & \cdots & \boldsymbol{G}(\boldsymbol{a_L}, \boldsymbol{b_{1L}}, \boldsymbol{x_1}) \\ \vdots & \vdots & \vdots \\ \boldsymbol{G}(\boldsymbol{a_1}, \boldsymbol{b_1}, \boldsymbol{x_N}) & \cdots & \boldsymbol{G}(\boldsymbol{a_L}, \boldsymbol{b_L}, \boldsymbol{x_N}) \end{bmatrix} \quad (3)$$

T is the training data target matrix: $\mathrm{T} = \begin{bmatrix} \boldsymbol{t_1} \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{t_N} \end{bmatrix}$ (4)

ELM is a set of regularization neural networks, but with non-tuned invisible layer mappings, its objective function is (produced by either random hidden nodes, kernels or otherimplementations).

$$\textbf{Minimize}: \|\beta\|_p^{\sigma_1} + C\|\boldsymbol{H\beta} - \boldsymbol{T}\|_q^{\sigma_2}, \quad (5)$$

Where $\boldsymbol{\sigma_1} > \mathbf{0}, \boldsymbol{\sigma_2} > \mathbf{0}, \boldsymbol{p}, \boldsymbol{q} = \mathbf{0}, \frac{1}{2}, \mathbf{1}, \mathbf{2}, \ldots, +\infty$.Different combinations of $\sigma_1, \sigma_2, pandq$can be used and lead to different predication, grouping, sparse coding, compression, function learning, and clustering learning algorithms.

The simplest ELM training algorithm trains a model of the form (for single hidden layer sigmoid neural networks) as a special example.:

$\hat{\boldsymbol{Y}} = \boldsymbol{W_2}\boldsymbol{\sigma}(\boldsymbol{W_1 x})$ (6)

If w_1 is the input-to-hidden layer weight matrix, σ is an activation function, and $w_2$is the hidden-to-output-layer weight matrix. The algorithm continues like this:

Fill $\boldsymbol{W_1}$with random values (e.g., Gaussian random noise) estimate $\boldsymbol{W_2}$by least-squares fit to a matrix of response variables Y, calculated using the pseudoinverse.$^{+}$,given a design matrix X:

$$\boldsymbol{W_2} = \boldsymbol{\sigma}(\boldsymbol{W_1 X})^{+Y} \quad (7)$$

B. WHALE OPTIMIZATION ALGORITHM(WOA)

On a population basis, WOA may be a new algorithm implemented in 2016 by Mirjalili and Lewis [10]. WOA stems from the humpback whales' social behavior. Comparable to other population-based algorithms, To modernize and optimize the role of candidate arrangements at each stage, WOA uses three rules, namely prey, bypass and prey[13][24], using a sequence of irregular candidate solutions (population) and jobs. It can be discussed as follows:

B.1 Bubble Net Attacking Method.

Two strategies are designed to explain the net bubble behavior of the Humpback whales mathematically, referred to as the Bubble Net Attacking Activity Monitoring Phase [11]. The following two strategies will be described:

B.1.1 Surrounding prey

Humpback whales locate and provide the survivor with a postponement. WOA Algorithm rates; prevailing fittest search Agent position to be the victim's destination or near to the ideal location, and other search Agents may attempt to resume their work against the most excellent look agent. The path shall be conveyed as follows in the following circumstances:

$$\vec{D} = |\vec{C}.\overrightarrow{X^*}(t) - \vec{X}(t)|, \quad (8)$$

$$\vec{X}(t+1) = \overrightarrow{X^*}(t) - \vec{A}\,.\vec{D} \quad (9)$$

The position vector of the best solution obtained to date by loop t is the position vector of each expert if the current iteration is shown by t.,|| It is the absolute sum and is the multiplication of element by element. This determines the coefficient of the vector as follows:

$$\vec{A} = 2\vec{a}.r - \vec{a}, \quad (10)$$

$$\vec{C} = 2r \quad (11)$$

Calculate each solution's fitness**,** where the iteration series is linearly decreased from 2 to 0 and r is a random integer [0,1].

By Shrinking Encircling Tool, this type of whaling is repeated by decreasing the case estimate (10). Note that Fluctuation extends further decreased by to put it another way, maybe random calculation within the duration [-a, a] where it decreases from 2 to during iterations. Proposing random values for the unused area of the search agent in [−1,1] could be located anywhere between the first area of the agent and the area of the current best agent.

B.1.2. Position Spiral Updating

The difference is determined between the whale at (X, Y) and the abuser at (X*, Y*). At this point, after the Helix-shaped forming of the Humpback whale, a winding condition is formed between the position of the whale and the victim to be taken as follows:

$$\vec{X}(t+1) = e^{bk}.cos(2\pi k).\ \overrightarrow{D^*} + \overrightarrow{X^*}(t), \quad (12)$$

$$\overrightarrow{D^*} = |\overrightarrow{X^*}(t) - \vec{X}(t)|, \quad (13)$$

To retrieve the pattern of the logarithmic spiral, if **B** can be a constant value, and k can be an arbitrary number within the range [−1, 1]. This way, in the middle of optimization, WOA thinks about changing whaling. There is a 50 percent chance of choosing between the lease component and the contract portion, and the contract components are as follows:

$$\vec{X}(t+1) = \begin{cases} \overrightarrow{X^*} - \vec{A}.\vec{D} \\ e^{bk}.\cos(2\pi k).\overrightarrow{D^*} + \overrightarrow{X^*}(t), & if\ p \geq 0.5, \end{cases} \quad (14)$$

Where p is a random number within the (0,1).

B.2 Search for Prey

The same method is based on the various features of the vector that can be seen at the point of the prey's test, which is known to be the prosecution stage. Depending on the area in which the victim is identified, whales conduct random scans. WOA uses a random value vector that is stronger or less than 1. In this manner, to support quest agents navigate away from the neighborhood of the Whale [26]. In conjunction with a randomly selected search agent, the position of the search agent is increased at the inquiry level instead of the finest stage of search agent manipulation). In order to do a global search and solve the complete local problem [11], this approach makes a variation in the WOA algorithm. The science demonstration will be seen as follows:

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - \vec{A}.\vec{D} \quad (15)$$

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_{rand}} - \vec{X}\right|, \quad (16)$$

where $\overrightarrow{X_{rand}}$ The irregular position vector (a random whale) selected from the present population is the irregular position vector.

Algorithm1: Whale Optimization Algorithm (WOA)

Generate Initial Population X_i (i = 1, 2, …, n)
Calculate the fitness of each solution
X^*= the best search agent
while (t < Max_Iteration)
    for each solution
        Update a, A, C, l, and p
        if 1 (p < 0.5)
            if 2 (|A| < +1)
                Update the existing solution's location by Eq. (9)
            else if 2 (|A| > +1)
Choosing a random search agent
Update the new search agent's location with an Eq. (15)
end if 2
else if 1 (p ≥ 0.5)
Update the latest search location for an Eq. (12)
        end if 1
    end for
Verify that every solution reaches the search space and adjust it
Calculate each solution's fitness
Update X^* if a better solution occurs t = t + 1
end while
return X^*

---

4. THE PROPOSED WOA-ELM MODEL

Extreme learning machines (ELMs) have the edge of low training time, while preserving palatable grouping and preaching, providing that a sufficient number of hidden nodes are chosen within the show. Inside hidden layer, although the optimal hidden layer weighting is not feasible, the vast number of nodes slows down the execution of the ELM experiments. As a result, fewer nodes within the hidden layer must be used in this hybrid model to speed up the ELM's execution while preserving an optimum choice of hidden layer weights and predispositions. It is also used with the same approach when setting the weights and predispositions of the output layer. To catch the hidden layer weights and preferences that maximize the overall execution of the ELM, the WO algorithm is best used [42]. WOA algorithm is employed to overcome the traditional ELM drawbacks, and automatically select the best weights and biases values to overcome the overfitting and local minima problems found in classical ELM model.
The proposed WOA-ELM model is described in algorithm2.

Algorithm 2: Hybrid Whale Optimization and Extreme Learning Machine Algorithm (WOA-ELM)

Inputs:

- N: the maximum number of Agents
- T: Number of Iterations
- X^*: the best search agent
- P : The current search position
- T rn.: Training Data set
- V ld.: Validation Data set

Outputs:

- $f$best: Optimal hidden weights and biases
- $f(f$best): sum square error for the NN over the validation set $f$best

Initialize:

- Generate X_i Initial Population (i = 1, 2, …, n)
- Calculate each solution's fitness
- X^*= the ideal agent for a search

while (t < Max_Iteration)
    for each solution
        Update **a**, **A** by eq 9, **C** by eq 10, **l,** and **p**
        if 1 (p < 0.5)
            if 2 (|A| < +1)
                Update the position of the current solution by Eq. (9)
            else if 2 (|A| > +1)
                Choosing a random search agent
                Update the new search agent's location
end if 2
else if 1 (p ≥ 0.5)
Update the latest search location by the Eq. (12)
        end if 1
    end for
Verify that every solution reaches the search space and adjust it
Calculate each solution's fitness
Update X^* if a better solution occurs t = t + 1
end while
return X^*

Algorithm	Parameters	Value
WOA	Number of Search agent	20
	Iterations Number.	100
ELM	Number of Input Nodes	20
	Number of Hidden Nodes	20
	Activation Fun.	Sigmoid
	Number of iterations	100
SVM	Number of Input Nodes	20
	Number of iterations	50

5. EXPERIMENTAL RESULTS

A. Dataset Description

With the classification information package of the UCI machine-learning store used in the experiments and comparisons [12], the presentation provided by WOA-ELM was prepared and checked. A number of highlights and incidents were chosen to provide the set of information as agents with various kinds of issues to be addressed in the strategy illustrated. To ensure that optimization algorithms are applied in colossal search spaces as seen in table (1), we picked a set of high-dimensional results separately. The approach of gathering the information is divided into a method of cross-validation assessment. K-1 folds are used for preparation and adoption of K-fold cross-validation and the remainder of the folding is used for testing purposes. L times have been renovated with this technique. Subsequently, K*L times with the collection of personal details would be checked by the maximizing consumer. The material is analyzed in the same way for planning, receipt and review. To prepare the classifier used for optimization and final review, the planning part is used. The validation component was used during the optimization process to validate the execution of the classifier. The evaluation section is used to determine the last highlights chosen by the prepared classifier.

Dataset Name	No. of Features	No. of Samples
Gene expression cancer RNA-Seq Data Set	16382	801

Table 1: Data Set used in the Experiment

B. Parameters Settings

For the proposed and compared versions, five hundred cycles have been planned. The ELM has 20 input layer nodes. It has five hidden nodes, but more hidden nodes are expected than algorithms of classical inclination. In the output layer, it has one node. The WOA and ELM algorithm parameter settings are summarized in the table 2.

Table 2:Parameters Settings

C. Performance Evaluation Criteria

In compliance with six assessment requirements, the proposed and compared models were tested. The accuracy of the classification was checked by these criteria. The evaluation criteria must be determined as follows:

- **Accuracy:** It is one metric for the measurement of models of classification. Occasionally, reliability is the distribution of predictions that our data has gotten right. Formally, specificity must be taken after definition:

$$\text{Accuracy} = \frac{Numberofcorrectpredictions}{Totalnumberofpredictions} \quad (17)$$

In addition, accuracy can be determined in terms of positives and negatives for double classification as taken after:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (18)$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

- **Precision:** Is the proportion of correctly predicted positive expectations of all expected positive perceptions.
  - **Precision** $= \frac{TP}{TP+TF}$ (19)
  - **micro**: by counting the cumulative true positive, false negative and false positive ratios internationally.
  - **macro**: the metrics for each label are estimated and the non-weighted mean is found for each label. The imbalance of a symbol does not take this into consideration.
  - **weighted**: To find their average weighted help, measure the metrics for each mark (the number of true instances for each label). To compensate for the imbalance of the mark, this changes the 'macro'; it can result in an F-score that is not between accuracy and recall.
- **Recall** (Sensitivity): Is the proportion of correctly predicted optimistic impressions to other expectations of real life-yes.

- **Recall** = $\frac{TP}{TP+FN}$ (20)

- **F1 Score:** This is the weighted normal of exactness and analysis. This score therefore takes into account both untrue positive and untrue negatives. Instinctively, it's not as easy to get it as accuracy, but F1 is typically more useful than accuracy, particularly if you have an uneven transmission of lessons. Precision functions well where untrue positive and untrue negatives have taken a proportional toll.

WOA-ELM	ELM	SVM
25.088	0.247	0.049

- **F1 Score** = $\frac{2*(Recall * Precision)}{(Recall + Precision)}$ (21)

-

6. Simulation Results

The hybrid WOA-ELM model has a better accuracy than standard ELM and SVM as shown on Table 3 and Figure1.

Table 3: Accuracy of different optimization algorithms

WOA-ELM	ELM	SVM
86.094	73.913	61.491

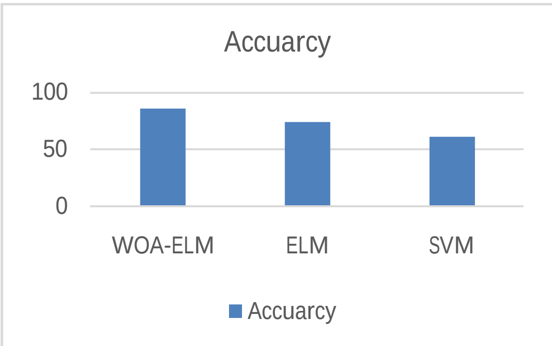
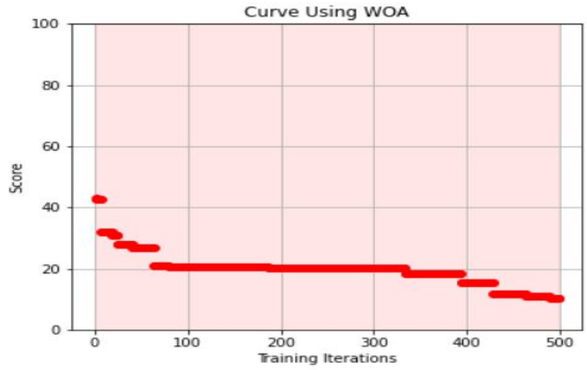


Figure 1: Accuracy Comparison between WOA-ELM and compared models

WOA-ELM achieves fast convergence to global minimum as shown Figure 2.



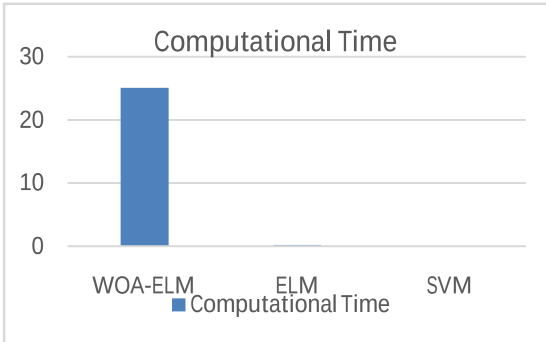


Figure 2: Learning Curve of WOA-ELM.

Although the hybrid WOA-ELM model has a better accuracy and convergence, it has the worst computational time compared to the standard ELM and SVM as shown on Table 4 and Figure 3.

Table 4: Computational Time of different optimization algorithms

Figure 3: Computational Time for WOA-ELM

The hybrid WOA-ELM model has a better precision than standard ELM and SVM as shown on Table 5 and Figure 4.

Table 5: Precision of different optimization algorithms

Precision	WOA-ELM	ELM	SVM
Micro avg	0.86	0.74	0.75
Macro avg	0.72	0.52	0.72
Weighted avg	0.82	0.67	0.80

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 10, Issue 1, January - February 2021 ISSN 2278-6856

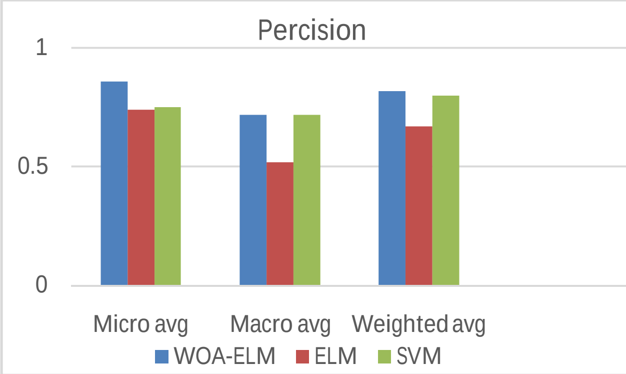


Figure 4: Precision of different optimization algorithms

The hybrid WOA-ELM model has a better recall than standard ELM and SVM as shown on Table 6 and Figure 5.

Table 6: Recall of different optimization algorithms

Recall	WOA-ELM	ELM	SVM
Micro avg	0.86	0.74	0.81
Macro avg	0.68	0.59	0.60
Weighted avg	0.79	0.74	0.76

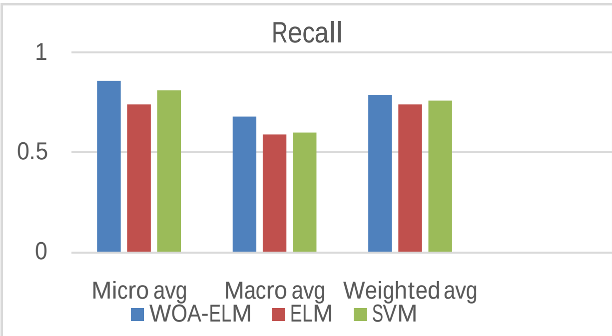


Figure 5: Recall of different Optimization Algorithms

The hybrid WOA-ELM model has a better F1Score than standard ELM and SVM as shown on Table 7 and Figure 6.

Table 7: F1Score of different optimization algorithms

F1Score	WOA-ELM	ELM	SVM
Micro avg	0.86	0.74	0.80
Macro avg	0.76	0.55	0.76
Weighted avg	0.86	0.70	0.84

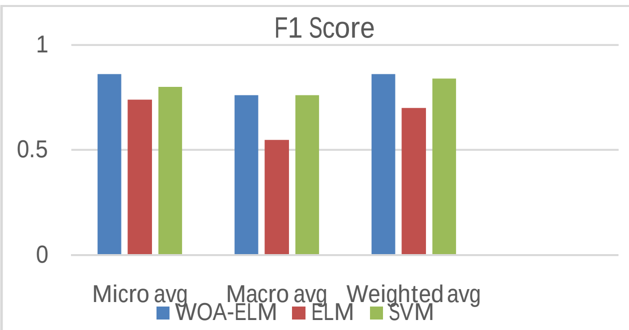


Figure 6: F1 Score of different Optimization Algorithms

6 CONCLUSION AND FUTURE WORK

This paper proposes an effective bio-inspired whale algorithm to optimize Extreme Learning Machine (ELM). Whale Optimization Algorithm (WOA)was best used to select input weights and hidden layer biases to build a more efficient, instead of random, network structure as seen in standard ELM models. The proposed WOA-ELM model was applied to gene classification data from the UCI repository. This data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extraction of gene expressions of patients having different types of tumor: BRCA, KIRC, COAD, LUAD and PRAD. Few iterations can be expected from the proposed model convergence to a global minimum. The proposed model overcame the over-fitting problem found in the standard ELM model. The displayed WOA-ELM parameters are limited and can be quickly modified. The suggested model obtained the lowest error value for all the compared assessment parameters.

The research concluded that WOA was extremely successful when optimizing the ELM model, and more research efforts should be made in this interesting field.

References

[1] Y. Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, and Lendasse Op-elm: Optimally pruned extreme learning machine. IEEE Transactions on Neural Networks, 21(1):158–162, 2010.

[2] G.B.Huang, Q.Y. Zhu, and C.K. Siew Extreme learning machine: theory and applications. Neurocomputing, 70(1):489–501, 2006.

[3] F. Han and D.S. Huang Improved extreme learning machine for function approximation by encoding a priori information. Neurocomputing, 69(1):2369–2373, 2006.

[4] K.K.A. Ghany, H.A. Hefny, A.E. Hassanien, and M.F. Tolba ,Kekre's transform for protecting fingerprint template.13th International Con- ference on Hybrid Intelligent Systems, pages 186–191, 2013.

[5] G.B. Huang, Q.Y. Zhu, and C.K.Siew, Extreme learning machine: a new learning scheme of feedforward neural networks. In International Joint Conference on Neural Networks (IJCNN 2004), pages 985–990. Budapest, Hungary, 2004.

[6] X. Li, H. Xie, R. Wang, Y. Cai, J. Cao, F. Wang, H. Min, and X. Deng, Empirical analysis: stock market prediction via extreme learning machine. Neural Computing and Applications, 1(3):1âA˘ S¸, 12, 2014.

[7] Q.Y. Zhu, A.K. Qin, P.N. Suganthan, and G.B. Huang, Evolutionary extreme learning machine, pattern recognition. Pattern Recognition, 38(10):1759–1763, 2005.

[8] G.P. Zhao, Z.Q. Hen, Miao, and Z.H. Man, on improving the conditioning of extreme learning machine: a linear case. In 7th International Conference on Information, Communications and Signal Processing (ICICS2009), page 1âA˘ S¸ 5. Maucu, China, 2009.

[9] Y. Wang, F. Cao, & Y. Yuan, A study on effectiveness of extreme learning machine. Neurocomputing, 74(16): 2483-2490, (2011).

[10] M.M. Mafarja, M. Majdi, and SeyedaliMirjalili, Hybrid whale optimization algorithm with simulated annealing for feature selection. Neurocomputing 260:302-312, 2017.

[11] M.F. Isham, M. Firdaus, Optimized ELM based on Whale Optimization Algorithm for gearbox diagnosis, MATEC Web of Conferences. EDP Sciences 255, 2019.

[12] K. Bache and M. Lichman, Uci machine learning repository, 2013.

[13] S. Mirjalili, A. Lewis, The whale optimization algorithm, Advances in engineering software 95: 51-67, 2016.

[14] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine, theory and applications, Neurocomputing 70.1-3: 489-501, 2006.

[15] Y. Shu Y. Xu. Evolutionary extreme learning machine-based on particle swarm optimization. International Symposium on Neural Networks, 3971(1):644–652, 2006.

[16] Q. HuaLing F. Han, H. FenYao. An improved evolutionary extreme learning machine based on particle swarm optimization. Neurocomputing, 116(1):87–93, 2013

[17] M. Alhamdoosh D. Wang. Evolutionary extreme learning machine ensembles with size control. Neurocomputing, 102(1):98–110, 2013.

[18] Z. Wu J. Yang X. Xue, M. Yao. Genetic ensemble of extreme learning machine. Neurocomputing, 129(1):175–184, 2014.

[19] N. Sundararajanc S. Suresha, S. Saraswathib. Performance enhancement of extreme learning machine for multi-category sparse data classification problems. Engineering Applications of Artificial Intelligence, 23(7):1149–1157, 2010.

[20] J. Zhao Z. Dong H. Yang, J. Yia. Extreme learning machine based genetic algorithm and its application in power system economic dispatch. Neurocomputing, 102(7):154–162, 2013.

[21] O. Hegazy, O.S. Soliman, M. Abdul Salam. Fpa-elm model for stock market prediction. International Journal of Advanced Research in Computer Science and Software Engineering, 5(2):1050–1063, 2015.

[22] E. Emary K.K.A. Ghany B. Parv M.A. Salam, H.M. Zawbaa. A hybrid dragonfly algorithm with extreme learning machine for prediction. International Symposium on INnovations in Intelligent SysTems and Applications (INISTA), pages 1–6, 2016.

[23] B. Parv M. Sharawi H.M. Zawbaa, E. Emary. Feature selection approach based on moth-flame optimization algorithm. IEEE Congress on Evolutionary Computation (CEC), pages 4612–4617, 2016.

[24] I. Aljarah, H. Faris & S. Mirjalili, optimizing connection weights in neural networks using the whale optimization algorithm. Soft Computing, 22(1):1-15, (2018).

[25] M. Abd El Aziz, A. A. Ewees& A. E. Hassanien, Whale optimization algorithm and moth-flame optimization for multilevel thresholding image segmentation. Expert Systems with Applications, 83:242-256, 2017

[26] K. ben oualidMedani, S. Sayah, A. Bekrar, Whale optimization algorithm based optimal reactive power dispatch: A case study of the Algerian power system. Electric Power Systems Research, 163:696-705,2018.

[27] P. Mohapatra, S. Chakravarty, & P. K. Dash,An improved cuckoo search based extreme learning machine for medical data classification. Swarm and Evolutionary Computation, 24: 25-49, (2015).

[28] E. adavandi, S ostafayi, &P. Soltani, A grey wolf optimizer-based neural network coupled with response surface method for modeling the strength of siro-spun yarn in spinning mills. Applied Soft Computing, 72, 1-13, 2018

[29] W. Y. Deng, Y. S. Ong, & Q. H. A. Zheng, fast reduced kernel extreme learning machine. Neural Networks, 76:29-38, (2016).

[30] X. Liu., L. Wang, G.B. Huang, J. hang, & J. Yin, Multiple kernel extreme learning machine. Neurocomputing, 149:253-264, 2015

[31] G. B Huang, X. Ding, & H. Zhou, Optimization method based extreme learning machine for classification. Neurocomputing, 74(1-3):155-163, 2010.

[32] W. Zong, G.B. Huang & Y. Chen, Weighted extreme learning machine for imbalance learning. Neurocomputing, 101: 229-242, 2013.

[33] O. Hegazy, O.S. Soliman, and M. Abdul Salam, "A Machine Learning Model for Stock Market Prediction", International Journal of Computer Science and Telecommunications, Vol. (4), Issue (12),pp. 17-23, December 2013.

[34] O. Hegazy, O.S. Soliman, and M. Abdul Salam, "LSSVM-ABC Algorithm for Stock Price Prediction", International Journal of Computer Trends and Technology (IJCTT), Vol. (7), Issue (2),pp. 81-92, Jan 2014.

[35] O. Hegazy, O.S. Soliman, and M. Abdul Salam, "Optimizing LS-SVM using Modified Cuckoo Search algorithm (MCS) for Stock Price Prediction", International Journal of Advanced Research in Computer Science and Management Studies, Vol. (3), Issue (2),pp. 204-224, February 2015.

[36] O. Hegazy, O.S. Soliman, and M. Abdul Salam, "Comparative Study between FPA, BA, MCS, ABC, and PSO Algorithms in Training and Optimizing of LS-SVM for Stock Market Prediction", International Journal of Advanced Computer Research Vol.(5), Issue (18),pp.35- 45, March-2015.

[37] H.Abdul-Kader, MA. Salam. Evaluation of Differential Evolution and Particle Swarm Optimization Algorithms at Training of Neural Network for Stock Prediction. Int. Arab. J. e Technol. 2012;2(3):145-51.

[38] R. Salem, M. Abdul Salam, H. Abdelkader and A. Awad Mohamed, "An Artificial Bee Colony Algorithm for Data Replication Optimization in Cloud Environments," in IEEE Access, vol. 8, pp. 51841-51852, 2020, doi: 10.1109/ACCESS.2019.2957436.

[39] EM Badr, M. Abdul Salam, M Ali, H Ahmed, Social Media Sentiment Analysis using Machine Learning and Optimization Techniques, International Journal of Computer Applications (0975 – 8887) Volume 178 – No. 41, August 2019.

[40] MA. Salam, M. Ali, Optimizing Extreme Learning Machine using GWO Algorithm for Sentiment Analysis. International Journal of Computer Applications.;2020, 975:8887.

[41] O.S. Soliman, and M. A. Salam, "A Hybrid BA-LS-SVM Model and Financial Technical Indicators for Weekly Stock Price and Trend Prediction", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), vol. 4, issue 4, pp. 764-781, 2014.

[42] S. Mirjalili, A. Lewis, The whale optimization algorithm, Adv. Eng. Softw. 95: 51–67, 2016.

AUTHORS

Mustafa Abdul Salamwas born on November 1, 1981 in Sharkia, Egypt. He received the B.Sc. from Faculty of Computers &Informatics, Zigzag University, Egypt in 2003 with grade very good with honor, and obtains master degree in information system from faculty of computers and information, menufia university, Egypt in 2009 specializing in Hybrid Machine Learning and Bio Inspired Optimization algorithms. He obtained his Ph.D. degree in information system from faculty of computers and information, Cairo University, Egypt. He is currently an assistant professor in AI department, Faculty of Computers and AI, Benha University, Egypt. He has worked on a number of research topics. Mustafa has contributed more than 25+ technical papers in the areas of neural networks, support vector machines, optimization, time series prediction, extreme learning machine, hybrid AI models in international journals, international conferences, local journals and local conferences. His majors are Machine Learning, Big Data, Stream Data Mining, and Deep Learning.

Ahmad Taher Azar (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from the Faculty of Engineering, Cairo University, Egypt, in 2006 and 2009, respectively. He is a Research Professor with Prince Sultan University, Riyadh, Saudi Arabia. He is also an Associate Professor with the Faculty of Computers and Artificial Intelligence, Benha University, Egypt. He worked in the areas of control theory and its applications, process control, chaos control and synchronization, nonlinear control, robust control, and computational intelligence. He workied as an Associate Editor for the IEEE Transactions on Neural Networks and Learning Systems, from 2013 to 2017, and ISA Transactions (Elsevier), from 2018 to 2020. He is the Editor-in-Chief of the International Journal of System Dynamics Applications (IJSDA) (IGI Global, USA) and the International Journal of Intelligent Engineering Informatics (IJIEI) (Inderscience Publishers, Olney, U.K.). He is currently an Associate Editor of the IEEE Systems Journal. (Based on document published on 7 July 2020).

Rana Usama Hussien received B.Sc. degree in Computer and information Science from Ain Shams University in 2013. She is working on M.Sc. in Computers and Artificial intelligence, Benha University, Egypt.