Academic Science Journal

Adaptive tTrainer for Multi-layer Perceptron using African Vultures Optimization Algorithm

Tuqa Ali Mohamed and Muntadher Khamees Mustafa

Department of Computer Science, College of Sciences, Diyala University, Diyala, Iraq

Scicompms2108@uodiyala.edu.iq

Received: 7 September 2022 **Accepted: 15 January 2023**

DOI: https://doi.org/10.24237/ASJ.02.01.679B

Abstract

This paper utilized a newly proposed multi-layer perceptron (MLP) that has been trained using a meta-heuristic technique (algorithm) that was developed using the idea of the African Vultures Optimization Algorithm. The precision and consistency of the proposed method's convergence as performance metrics. The African Vultures Optimization Algorithm(AVOA) was recently proposed for use in training multi-layer perceptron (MLP), and it employs the five most common classification data sets currently available(XOR, balloon, breast cancer, heart, Iris) in the California University at Irvine UCI Repository .The newly Optimizers **(**AVOA**)** are being us for the first time as a Multi-Layer Perceptron (MLP) trainer, and its results are compared to those obtained using the more established gray wolf optimization (GWO), the whale optimization algorithm (WOA), and the sine cosine algorithm are examples of optimization techniques (SCA). Previously, AVOA was used to determine the best weights and biases for the optimal solution.

Keywords: Training MLP, ANN, WOA, SCA, GWO, and AVOA.

مدرب متكيف للشبكه متعدده الطبقات باستخدام خوارزمية تحسين النسور الأفريقية

تقى علي محمد و منتظر خميس مصطفى

قسم علوم الحاسوب - كلية العلوم - جامعة ديالى

الخلاصة

استخدمت هذه الورقة مقترحًا حديثًا متعدد الطبقات (MLP) تم تدريبه باستخدام تقنية الفوقيه للكشف عن مجريات الأمور (خوارزمية) تم تطويرها باستخدام فكرة خوارزمية تحسين النسور الأفريقية. مقاييس الأداء. تم اقتراح خوارزمية تحسين النسور الأفريقية (AVOA) مؤخرًا للاستخدام في تدريب MLP ، وهي تستخدم مجموعات بيانات التصنيف الخمس الأكثر شيوعًا المتاحة حاليًا(XOR ، بالون ، سرطان الثدي ، القلب ، القزحيه)في جامعة كاليفورنيا في إيرفين مستودع UCI. يتم استخدام Optimizers (AVOA) حديثًا لأول مرة كمدرب متعدد الطبقات (MLP) ، وتتم مقارنة نتائجها بالنتائج التي تم الحصول عليها باستخدام خوارزمية تحسين الذئب الرمادي الأكثر رسوخًا (GWO) ، خوارزمية تحسين الحوت (WOA)وخوارزمية جيب التمام هي أمثلة على تقنيات التحسين .(SCA) في السابق ، تم استخدام AVOA لتحديد أفضل الأوزان والتحيزات للحل الأمثل .

الكلمات المفتاحية: تدريب MLP ، الشبكه العصبية ، خوارزمية الحوت WOA، خوارزمية جيب تمام SCA،خوارزمية الذئب الرمادي GWO،خوارزميه تحسين النسور الافريقية AVOA.

Introduction

Neural networks (NN) are one of the most ground-breaking developments in the field of AI. By modeling neuronal activity in the human brain, [1] in the research literature [2], they typically attack classification problems. Many different kinds of NNs have been proposed [3,4] Self-referential the networks feed[5] the radial basis function (RBF) network,[6] the recurrent neural network, and the convolutional neural network. For instance, neural networks that "spike" are a type of self-organizing network. In feed forward NNs (FNN), data is sent in a single direction across the network. Recurrent NNs, on the other hand, as their name implies, allow for bi - directional neuronal communication. The final process involves spiking. NNs cause neuronal stimulation with spikes. The use of NNs in education is widespread. When a NN has the ability to learn, it means that it can improve itself through exposure to new information. Like real neurons, artificial neural networks (ANN) can learn from experience and improve with new information. [7] Both supervised and unsupervised learning methods are applied commonly in this setting. The first instance is when the NN is responding to feedback from the outside world (supervisor). On the other hand, in unsupervised learning, a NN makes changes to inputs (learns) without any additional external feedback [8].

The two most important reasons for doing this work are as follows:

Exploration and exploitation are two strong points in AVOA repertoire, which could propel it past the competition.

There is still a problem with multi-solution stochastic trainers getting stuck at the local optimum[9].

J. Hamidzadeh .etal.2012[10] in proposed system new classification method utilizing distance-based decision surface with nearest neighbor projection approach, called DDC. Kernel type of DDC has been extended to take into account the effective nonlinear structure of the data. DDC has some properties: (1) does not need conventional learning procedure (as k-NN algorithm), (2) does not need searching time to locate the k-nearest neighbors, and (3) does not need optimization process unlike some classification methods such as Support Vector Machine (SVM). In DDC, we compute the weighted average of distances to all the training samples. Unclassified sample will be classified as belonging to a class that has the minimum obtained distance. As a result, by such a rule we can derive a formula that can be used as the decision surface. We found DDC from viewpoint of accuracy behave between k-NN and SVM algorithms in the most situations. Moreover, we excluded training phase in DDC not To apply other distances or new distances based on how the data is spread out in order to make DDC work better. For sample data to have a good projection line, it might be in the shape of a sphere. Using techniques like kernel and my be. One option for getting decision surfaces is to reduce the amount of data.

s.mill.etal.2014[11] in proposed system the use of the recently developed Biogeography-Based Optimization (BBO) algorithm for training MLPs to reduce these problems. In order to investigate the efficiencies of BBO in training MLPs, five classification datasets, as well as six function approximation datasets are employed. The method Biogeography-Based Optimization (BBO) and data set balloon, iris, breast cancer, heart, sigmoid, cosine with one peak, sine with four peaks, sphere, Griewank, and Rosenbrock and the limitation not training BBO in other types of NNs, like recurrence, Kohonen ,or Radial basis function`(RBF) networks.

S.Lee and , J. Y. Choeh.2014[12] in proposed systemo develop models for predicting the helpfulness of reviews, providing a tool that finds the most helpful reviews of a given product. This study intends to propose HPNN (a helpfulness prediction model using a neural network), which uses a back-propagation multilayer perceptron neural network (BPN) model to predict the level of review helpfulness using the determinants of product data, the review characteristics, and the textual characteristics of reviews .the methode Helpfulness Prediction Model Using A Neural Network (HPNN) Focus only on these results by looking into what makes online customer reviews seem good or bad. In turn, helpful feedback can change how people feel about online shopping.

Seyedali Mirjalili 2015[13] in proposed system employs the recently proposed Grey Wolf Optimizer (GWO) for training Multi-Layer Perceptron (MLP) for the first time. Eight standard datasets including five classification and three function-approximation datasets are utilized to benchmark the performance of the proposed method. For verification, the results are compared with some of the most well-known evolutionary trainers: Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Ant Colony Optimization (ACO), Evolution Strategy (ES), and Population-based Incremental Learning (PBIL) GWO-based trainer was applied to five standard classification datasets (XOR, balloon, Iris, breast cancer, and heart) as well as three function-approximation datasets (sigmoid, cosine, and sine). GWO algorithm has not been used to find the best number of hidden nodes, layers, and other MLP structural parameters. There should be more research into how to fine-tune this algorithm.

H. Ramchoun .etal.2016 [14] in proposed system is developed to optimize the architecture of Artificial Neural Networks. The Genetic Algorithm is especially appropriate to obtain the optimal solution of the nonlinear problem. This method is tested to determine the optimal number of hidden layers and connection weights in the Multilayer Perceptron and the most favorable weights matrix after training. We have proposed a new modeling for the multilayer Perceptron architecture optimization problem as a mixedinteger problem with constraints. Depending on the Iris data, the results obtained demonstrates the good generalization of neural networks architectures. The method EBP Not much training on real problems like diabetes, thyroid, and cancer from other databases

I. Aljarah .etal.2016 [15] in proposed system a new training algorithm based on the recently proposed whale optimization algorithm (WOA). It has been proved that this algorithm is able to solve a wide range of optimization problems and outperform the current algorithms. This motivated our attempts to benchmark its performance in training feed forward neural networks. WOA in training MLPs. The high local optima avoidance and fast convergence speed were the main motivations to apply the WOA to the problem of training MLPs. The problem of training MLPs was first formulated as a minimization problem. It is suggested that other kinds of ANNs be trained with WOA. It is worth thinking about how WOA-trained MLP can be used to solve classification problems in engineering. Using the WOA-trained MLP to solve function approximation datasets can also be a useful contribution.

M. Hesami .etal.2019 [16] in proposed system The aim of this study was modeling and optimizing in vitro sterilization of chrysanthemum, as a case study, through Multilayer Perceptron- Non-dominated Sorting Genetic Algorithm-II (MLP-NSGAII). MLP was used for modeling two outputs including contamination frequency (CF), and explant viability (EV) based on seven variables including $HgCl_2$, $Ca(ClO)_2$, Nano-silver, H_2O_2, NaOCl, $AgNO_3$, and immersion times. Subsequently, models were linked to NSGAII for optimizing the process, and the importance of each input was evaluated by sensitivity analysis. Results showed all of the R^2 of training and testing data were over 94%. According to MLP-NSGAII, optimal CF (0%), and EV (99.98%) can be obtained from 1.62% NaOCl at 13.96 min immersion time and method GA (MLP-NSGAII). MLP (MLP-NSGAII). MLP not focus on comparing and evaluating different multi-objective optimization algorithms in different areas of plant science, especially in plant tissue culture areas.

A. A Heidari Hesami .etal.2019 [17] in proposed systema new hybrid stochastic training algorithm using the recently proposed grasshopper optimization algorithm (GOA) for multilayer perceptrons (MLPs) neural networks. The GOA algorithm is an emerging technique with a high potential in tackling optimization problems based on its flexible and adaptive searching mechanisms. It can demonstrate a satisfactory performance by escaping from local optima and balancing the exploration and exploitation trends. The proposed GOAMLP model is then applied to five important datasets: breast cancer, parkinson, diabetes, coronary heart disease, and orthopedic patients .the method GOA GOA can be used with other kinds of NNs and big data sets.

S. Samadianfard.etal.2020 [18] in proposed system artificial intelligence algorithms is selecting the finest weights in the layers of neural networks that must permit the extraction of the relevant features within the input information for creating an accurate model. Constructing the best predictive model demands input data, which is considered as a crucial and useful tool for calculation of wind energy potential. In the present study, the utility of a reliable and robust method for predicting the wind speed for ten locations is revealed, where the wind speed amount of the target location was forecasted using input data of neighboring reference locations. In the current study by using the MLP, MLP-WOA, and MLP-GA models where the Whale Optimization and genetic algorithms combined with standalone MLP for each of the ten target stations, daily wind speed values are predicted. Furthermore, another climate or atmospheric information is not used for wind speed prediction with this method. To evaluate the performance of MLP-WOA, Several statistical indices were used the method IRIMO The data used is not described

I. Al-Badarneh, .etal.2020 [19] in proposed system an approach for training the MLP using three evolutionary algorithms for imbalanced classification. The proposed models are GWO-MLP, PSO-MLP, and SSA-MLP. The proposed approaches adopted three different fitness functions; accuracy, f1-score, and g-mean, and are evaluated using ten imbalanced datasets. The average results of 30 independent runs, the best results, and the standard deviations were calculated for each metric. The results showed that there is no clear superiority for one method over the other. However, the experiments showed that there is an obvious advantage of using g-mean and f-score fitness functions over the classification accuracy rate when the dataset is imbalanced. Neuro-evolutionary models with g-mean fitness function are recommended when it is preferable to increase the recall of both classes (e.g. major and minor). Whereas, using the f-measure of the minor class as fitness function is preferable when the minor class is more important Not looking into how the proposed meta-heuristic approach for optimizing Convolutional Neural Networks (CNN) could be used for complex tasks like classifying images and texts. Declaration.

F. Ece.etal.2020 [20] in proposed system the results, the advantage of the single models is their lower processing time, but the lowest accuracy can be the most important limitation and disadvantage of the single models compared with the hybrid ones. This was also claimed by several researches. In the case of using the hybrid models, the advantages of MLP–PSO such as higher accuracy and lower processing time overtake the MLP–GA. 5. Conclusions In this paper, modeling was performed by MLP–GA and MLP–PSO in two scenarios including with Tanh (x) and with the Gaussian function as default as the output function in thirteen categories. Research outcomes were evaluated using RMSE and correlation coefficient values to compare the accuracy and performance of the developed models in training and testing steps. Based on the results, using Tanh (x) as the output function improved the accuracy of models significantly. MLP–PSO with population size 125 followed by MLP–GA with population size 50 provided higher accuracy in the testing step by RMSE 0.732583 and 0.733063, MAPE of 28.16%, 29.09% and correlation coefficient 0.694 and 0.695, respectively. As is clear, the only advantage of the single MLP is its lower processing time but the important disadvantage can be claimed the lower accuracy compared with the hybrid models. Not address the stock market's beatings, that the stock market's fluctuations will be successfully addressed. So, the return difference is a problem with the current study.

The feed-forward neural network with the multilayer perceptron

As was discussed FNNs are NNs with unidirectional neuronal connections [21] .This NN stacks in numerous layers. First is input, then output. Hidden layers are between input and output. One-hidden-layer FNN or MLP [22]. MLPs output inputs, weights, and biases. [13]First, we compute the input weight summing by (1).

$$S_j = \sum_{i=1}^{n}(Wij * Xi) - \theta j, \quad j = 1, 2, \ldots h \qquad (1)$$

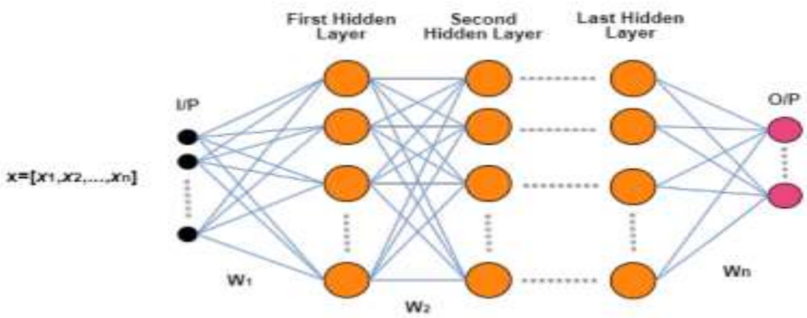
Where n is the total number of input nodes, W ij represents the weight of the link from the I the input node to the j the hidden node, j is the bias (threshold) of the j the hidden node, and Xi is the I the input. n is the total number of input nodes. The following calculation is used to determine the output of each concealed node: (2)

$$Sj = \text{sigmoid } sj = \frac{1}{(1 + exp - (-sj))} \quad , j = 1, 2,...h \qquad (2)$$

Second the calculated results from the hidden nodes (hidden layer) are used to define the final results in the following ways:

$$O_k = h \sum_{j=1}^{h} (W_{jk} .S_j) - \theta_k, \quad k = 1, 2,......,m \qquad (3)$$

$$O_k = \text{sigmoid } (O_k) = \frac{1}{(1 + \exp(-ok))}, \quad k = 1, 2,.....,m \qquad (4)$$



General MLP Architecture [28].

AVOA "African Vultures Optimization Algorithm"

The AVOA is designed to mimic the hunting and navigational habits of African vultures, which it takes its cues from. The technique begins with a group of vultures and examines the best and second-best options available.. The method consists of four steps[23]:

(i) selecting the top bird from among a group of vultures

(ii) the percentage of vultures that are starving to death.

(iii) exploration

(iv) Exploitation .Whether vultures are in the exploration or exploitation phase can be seen by how fast they are going hungry. Two different strategies for exploration are implemented, with the strategy selected by the parameter 'P1'. During the exploitation phase, two distinct escape and siege-fight strategies are implemented based on the parameter 'P2'. Due to its lower computational complexity, the AVOA has been established to execute faster than other algorithms. The development of AVOA's work as represented in point [24] :

- A habitat may have as many as N vultures. Meta-heuristic algorithms, like AVOA, have a fixed population size that is depending on the problem that researchers are trying to solve.
- In a natural ecosystem, a large number of vultures are can physically divided into two distinct groups. Before categorization of the vultures, the algorithm computes the fitness function of all possible solutions to the problem (the initial population). The phrase "first and best vulture" refers to the finest possible reaction, whereas the phrase "second-best vulture" stands for the second-best possible response.
- In each performance, the other vultures are replaced or relocated by members of the other population .Using this technique; vultures can be classified according to their primary natural role, which is to gather food in large groups. Inability to find and eat food is a problem for each group of vultures. Because vultures have a natural urge to eat, they can evade the hunger trap. Under the assumption that the weakest and hungriest in The crowd is the worst possible answer; the vultures are attempting to maintain a safe distance from the worst possible solution while coming up with the best possible solution. Vultures work hard to climb to the top of the AVOA food chain, but there are only two options that stand out as the most powerful and effective ones.
- When solving complex Due to the inherent nature of optimization issues, there is no assurance that the predictions for the global optimum in the final population will be accurate once the exploration phase has been completed.
- As a result, it causes the best local site to be found too soon. As computers have gotten better at solving complex optimization problems, it has become easier to get out of local optimal spots. During the last rounds of the AVOA algorithms, the exploitation phase and the exploration activities take place. Figure shows the different steps of an algorithm
- R(i) = { BestVulture1if pi = L1 BestVulture2if pi = L2 (1) [24]

- Eq. (1) calculates the likelihood of picking the chosen vultures to steer the other vultures toward one of the best options in each group. The variables that will be measured prior to the Search operation, where both parameters have values between 0 and 1 and their sum equals 1. In order to select the optimal answer for each group using Eq, the likelihood of selecting the best option is increased utilizing the roulette wheel (2).

$$Pi=\frac{Fi}{\sum_{i=1}^{n} Fi} \qquad (2) \quad [24]$$

In AVOA, an intensification will be enhanced if the -numeric parameter is close to the value 1, and if the -numeric parameter is close to the value 0. If the -numeric parameter is near to the number 1, in addition,

As the -numeric parameter gets closer to the value 0, AVOA becomes more diverse.

Phase Two: The rate at which vultures are starving

When full, vultures have high energy levels and can fly farther in their search for food. However, when hungry, vultures lack the stamina needed to fly as far and must instead compete for food with stronger vultures. Hungry vultures also tend to become violent.

To simulate this behavior mathematically

t = h × (sin w (π 2 × iteration i max iterations) + cos (π 2 × iteration i max iterations) − 1) (3) [24]

The AVOA algorithm

The (category -1-)

8- for (each vulture (p_i))

9- Select R (i) using Eq.

10- Update the F using Eq.

11- if (f ≥ 1) then

12- if (p_1≥ rand r_1) then

13- Update location of vulture

14- else

15- Update location of vulture using Eq2.

16- if (f < 1)then

17- if ($f \geq 0.5$) then

18- if ($p_2 \geq$ rand r_2) then

19- Update location of vulture

20- else

21- Update location of vulture

22- else

23- if ($p_3 \geq$ rand r_3) then

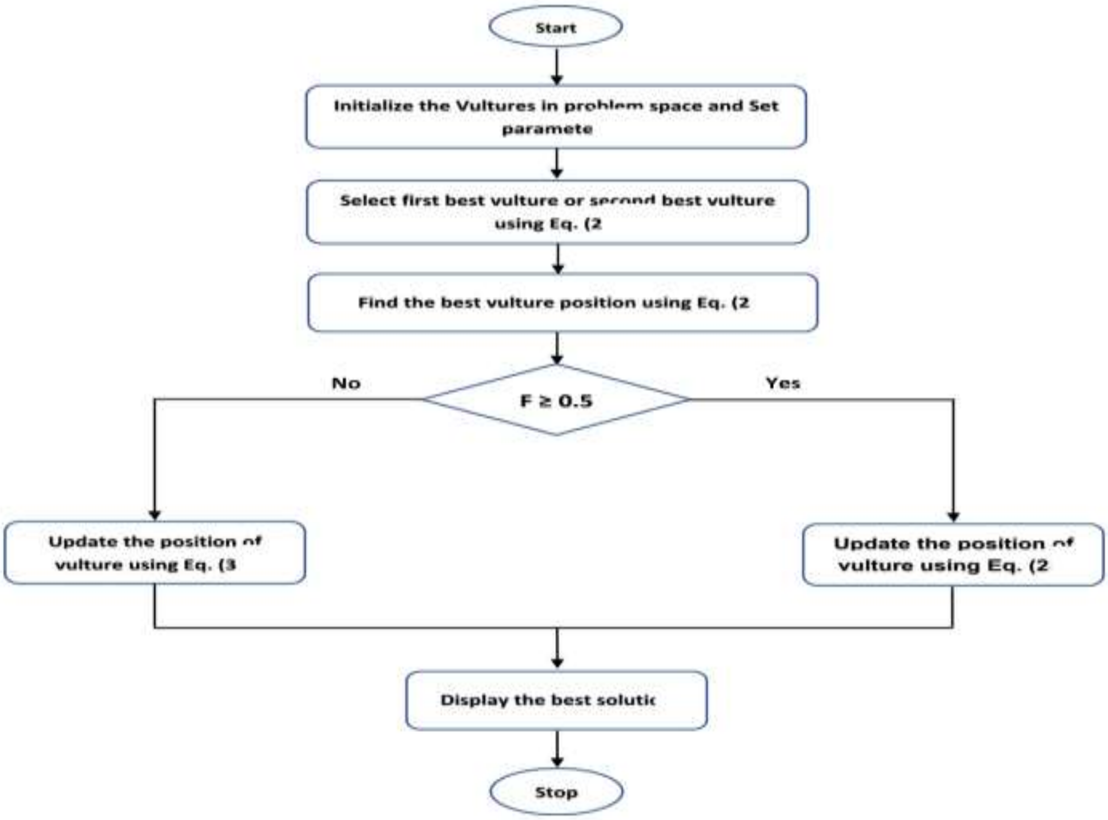
24- Update location of vulture

25- else

26- Update location of vulture using Eq3.

Return $p_{best\ vultures}$

Pseudo The AVOA algorithm [25]



AVOA flow chart [26]

AVOA -based MLP trainer

While training an MLP with meta-heuristics, the representation of the problem is the first and most crucial step [11]. Therefore, it is important to frame the MLP training problem in a way that whether or not Meta heuristics are acceptable. As introduced,, the weights and biases are the most crucial settings when training an MLP. A trainer's task is to find the optimal combination of the weights and the biases that leads to the best classification, close approximation, and classification rate forecasting are all possibilities. Therefore, the weights and biases are the independent variables. Since the AVOA algorithm expects the variables to be presented as a vector, the MLP's variables are presented in this format. : $\overline{V} = \{ \overline{W}, \overline{\theta} \}=\{W1,1, W1,2, \cdots , Wn, h, \theta1, \theta2, \cdots , \theta h\}$ (4.1) [13]

Wij j is the bias (threshold) of the I the hidden node, and n is the number of input nodes. Next, specify the AVOA objective function. Training an MLP aims to increase its classification, approximation, or prediction accuracy. Sample training and testing. MLPs are evaluated using MSE. . In this metric, the MLP is fed a predefined set of training samples, and the difference between the desired and actual output is measured using the following formula.

$$MSE = \sum_{i=1}^{m} (O_i^k - d_i^k)^2 \quad (4.2)\ [13]$$

Where d_i^k is the output that should be produced by the i the input unit using the k the training sample, while O_i^k is the output that actually gets produced .Where the value of m represents the number of outputs. The capacity of an MLP to generalize from examples provided during training is directly related to how well it performs. As a consequence of this, the performance of the MLP is evaluated by calculating the typical mean square error across all training samples.:

$$\overrightarrow{MSE} = \sum_{k=1}^{s} \frac{\sum_{i=1}^{m} (O_i^k - d_i^k)^2}{s} \quad (4.3)\ [13]$$

Where S stands for training samples, m for outputs, and d_i^k for decimal k. I is the output that should be produced by the I the input unit when trained with the k the sample, and O_i^k is the output that actually gets produced. After all, both the variable and the average MSE for the AVOA algorithm may be utilized to frame the MLP training issue in the following manner: To bring down :$F(\bar{V})$ = Mean Squared Error (4.4) [13]

Impact of biases and weights Assumed Training Samples Mean Squared Error Weights and biases are provided by AVOA to MLP, and in return, AVOA is given the overall average MSE for all training sample. Given that the weights and the biases tend to converge on the best MLPs that have been obtained to this point, there is a good chance that the MLP would get better with each iteration. The stochastic nature of AVOA means that it cannot be relied upon to always return the best MLP for a given dataset. However, as the MLPs are evolved using the best MLPs obtained so far, the population-wide average MSE decreases over time. Basically, if you run the AVOA algorithm enough times, it will converge on a solution that is superior to the initial solutions that were generated at random. In the following, we look into why the AVOA algorithm is so effective when training on the MLP in real world.

Results and Discussion

In this subsection, The AVOA-based MLP trainer that we've suggested here is trained with the use of five different standard classification datasets that can be found in the (UCI) Machine Learning Repository [27]. (balloon, heart ,XOR, breast cancer ,iris) table 1.

Table 1: Results from experiments on the XOR dataset

DATASETS	M.L.P. IN STRUCTURE	NUMBER OF THE ATTRIBUTES
Balloon	4. 4.9	4
Heart	22.45.1	22
3-bitsXOR	3.7.1	3
Breast cancer	9.19.1	9
Iris	4.9.3	4

For this dataset, the MLP's output must be the Boolean XOR of the input. Table results show that AVOA -MLP, GWO-MLP, and WOA-MLP all achieve a perfect classification rate100 % in table 2.

Table 2: Experimental research results from the balloon dataset

ALGORITHM	CLASSIFICATION RATE	MSE (AVE ±STD)
AVOA –MLP	100 %	0.3297± 0.2889
GWO-MLP	100 %	0.009410 ±0.29500
WOA-MLP	100 %	0.0006524 ± 0.00049
SCA –MLP	62.5%	0.118739 ±0.011574

The Balloon dataset consists of two classes, eighteen training/test samples, and four attributes. This dataset's trainers have 55 dimensions. The findings are tabulated below. The table results demonstrate that across all algorithms, classification accuracy is 100% in table 3.

Table 3: Experimental results for the iris dataset

Algorithm	Classification rate	MSE (AVE ±STD)
AVOA –MLP	100 %	2.4314e-07±7.1404e-07
GWO-MLP	100 %	9.38e-15 ±2.81e-14
WOA-MLP	100 %	4.61e-24 ±7.52e-23
SCA –MLP	100 %	0.000585 ±0.000749

The Iris dataset is 3 Classes, 150 samples for training and testing, and four qualities As a result, the MLP structure for resolving this dataset, and there are seventy-five variables involved in the issue. The outcomes of training several different algorithms are shown in the table below. In comparison to other algorithms, the results provided in Table 4 the demonstrate that the AVOA-MLP algorithm achieves the greatest classification rate of 92%.

Table 4: Experiments were done on breast cancer data.

ALGORITHM	CLASSIFICATION RATE	MSE (AVE ±STD)
AVOA –MLP	92 %	0.1160 ± 0.0541
GWO-MLP	88%	0.089912 ±0.123638
WOA-MLP	100 %	0.009410±0.029500
SCA –MLP	27%	0.084050 ± 0.035945

The breast cancer dataset contains two classes and nine attributes across 599 training samples. Trainers solve this dataset 10 times, and the aggregated results are tabulated below. The table results show that compared to other algorithms, AVOA -MLP has the highest classification rate at 99% in table 5.

Table 5: Experiments were done on the heart data set

ALGORITHM	CLASSIFICATION RATE	MSE (AVE ±STD)
AVOA –MLP	99 %	0.0012 ±7.4498e-05
GWO-MLP	98%	0.003026 ±0.001500
WOA-MLP	98%	0.003026 ±0.001500
SCA –MLP	85 %	0.089912 ±0.123638

The algorithms' final success came with the heart dataset, a classification problem with twenty-two features, eighty training samples, one hundred eighty-seven test samples, and two classes. These algorithms are training MLPs structure. The tabulated findings are presented below. As can be seen in the table below, AVOA -MLP algorithms achieve the highest classification rate of any algorithm tested, at 90% in table 6.

Table 6

ALGORITHM	CLASSIFICATION RATE	MSE (AVE ±STD)
AVOA –MLP	90 %	0.1302 ± 0.0225
GWO-MLP	88.75%	0.089912 ±0.123638
WOA-MLP	37.5%	0.084050 ± 0.035945
SCA –MLP	75 %	0.122600 ±0.007700

Conclusions

Recently, the AVOA algorithm was proposed, and its first use as an MLP trainer can be found in this paper. This research was driven by a curiosity about the extensive capabilities. The high level of, this algorithm for exploration and the exploitation. The AVOA algorithms were first to be presented with the problem of training an MLP. After determining the algorithm's optimal weights and biases, they put it to use. Five common classifications in datasets (Iris, breast cancer XOR, balloon, and heart) datasets were used to test the proposed AVOA -based trainer. Results from the AVOA -MLP algorithm were compared to those from five more stochastic optimization simulators. The results showed the proposed MLP training method worked. The

AVOA-MLP avoids local optima a high degree, enhancing the possibility of finding good MLP weights and biases. Because the AVOA-MLP trainer is so popular, its recommended weights and biases are accurate. This paper identifies strong and poor algorithms and discusses their causes. The AVOA algorithm can determine the ideal number of hidden nodes and layers in MLPs. Tuning this algorithm needs more investigation.in future works, AVOA can be applied to other types of the NNs or large-scale datasets. In addition, researchers can investigate the performance of used MLP-based approaches and the possibility of using other algorithms in training MLP.

References

1. Bhesdadiya, R., Jangir, P., Jangir, N., Trivedi, I. N., & Ladumor, D. (2016).Training multi-layer perceptron in neural network using whale optimization algorithm. Indian J Sci Technol, 9(19), 28-36.
2. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophysics 5:115–133
3. Bebis G, Georgiopoulos M (1994) Feed-forward neural networks. Potentials, IEEE 13:27–31
4. Kohonen, T. (1990). The self-organizing map. Proceedings of the IEEE, 78(9), 1464-1480.
5. Park J, Sandberg IW (1993) Approximation and radial-basisfunction networks. Neural Comput 5:305–3
6. Ghosh-Dastidar, S., & Adeli, H. (2009). Spiking Neural Networks. International journal of neural systems, 19 4, 295-308 .
7. Hertz J (1991) Introduction to the theory of neural computation. Basic Books 1
8. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
9. A. Botalb, M. Moinuddin, U. M. Al-Saggaf, and S. S. A. Ali, “Contrasting Convolutional Neural Network (CNN) with Multi-Layer Perceptron (MLP) for Big Data Analysis,” Int. Conf. Intell. Adv. Syst. ICIAS 2018, no. August, pp. 1–5, 2018,

10. J. Hamidzadeh, R. Monsefi, and H. Sadoghi Yazdi, “DDC: Distance-based decision classifier,” Neural Comput. Appl., vol. 21, no. 7, pp. 1697–1707, 2012,
11. S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Let a biogeography-based optimizer train your Multi-Layer Perceptron,” Inf. Sci. (Ny)., vol. 269, pp. 188–209, 2014,
12. S. Lee and J. Y. Choeh, “Predicting the helpfulness of online reviews using multilayer perceptron neural networks,” Expert Syst. Appl., vol. 41, no. 6, pp. 3041–3046, 2014,
13. S. Mirjalili, “How effective is the Grey Wolf optimizer in training multi-layer perceptrons,” Appl. Intell., vol. 43, no. 1, pp. 150–161, 2015, doi: 10.1007/s10489-014-0645-7.
14. H. Ramchoun, M. Amine, J. Idrissi, Y. Ghanou, and M. Ettaouil, “Multilayer Perceptron: Architecture Optimization and Training,” Int. J. Interact. Multimed. Artif.
15. I. Aljarah, H. Faris, and S. Mirjalili, “Optimizing connection weights in neural networks using the whale optimization algorithm,” Soft Comput., vol. 22, no. 1, pp. 1–15, 2018,
16. M. Hesami, R. Naderi, and M. Tohidfar, “Modeling and optimizing In vitro sterilization of chrysanthemum via multilayer perceptron-non-dominated sorting genetic algorithm-II (MLP-NSGAII),” Front. Plant Sci., vol. 10, no. March, pp. 1–13, 2019,
17. A. Heidari, H. Faris, I. Aljarah, and S. Mirjalili, “An efficient hybrid multilayer perceptron neural network with grasshopper optimization,” Soft Comput., vol. 23, no. 17, pp. 7941–7958, 2019.
18. S. Samadianfard et al., “Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm,” Energy Reports, vol. 6, pp. 1147–1159, 2020.
19. I. Al-Badarneh, M. Habib, I. Aljarah, and H. Faris, “Neuro-evolutionary models for imbalanced classification problems,” J. King Saud Univ. - Comput. Inf. Sci., no. xxxx, 2020
20. M. Stock and P. Index, “Training Multilayer Perceptron with Genetic Algorithms and Particle Swarm Optimization for,” mdpi, vol. 5, 2020.

21. Abdollahzadeh, B., Soleimanian Gharehchopogh, F., & Mirjalili, S. (2021). Artificial gorilla troops optimizer: A new nature-inspired metaheuristic algorithm for global optimization problems. International Journal of Intelligent Systems, 36(10), 5887-5958.
22. Bebis G, Georgiopoulos M (1994) Feed-forward neural networks. Potentials, IEEE 13:27–31
23. B. Abdollahzadeh, F. S. Gharehchopogh, and S. Mirjalili, "African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems," Comput. Ind. Eng., vol. 158, no. January, p. 107408, 2021
24. J. Fan, Y. Li, and T. Wang, "An improved African vultures optimization algorithm based on tent chaotic mapping and time-varying mechanism," PLoS One, vol. 16, no. 11, p. e0260725, 2021.
25. D. F. Campos, A. Matos, and A. M. Pinto, "An adaptive velocity obstacle avoidance algorithm for autonomous surface vehicles," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 8089–8096.
26. Yakout, A. H., Kotb, H., AboRas, K. M., & Hasanien, H. M. (2022). Comparison among different recent metaheuristic algorithms for parameters estimation of solid oxide fuel cell: Steady-state and dynamic models. Alexandria Engineering Journal, 61(11), 8507-8523.
27. Blake, C., Keogh, E. and Merz, C.J. (1998) UCI repository of machine learning databases. University of California, Irvine. http://www.ics.uci.edu/~mlearn/MLRepository.html
28. A. Botalb, M. Moinuddin, U. M. Al-Saggaf, and S. S. A. Ali, "Contrasting Convolutional Neural Network (CNN) with Multi-Layer Perceptron (MLP) for Big Data Analysis," Int. Conf. Intell. Adv. Syst. ICIAS 2018, no. August, pp. 1–5, 2018.