A Deep Learning H$_2$O Framework for Emergency Prediction in Biomedical Big Data

AHMED SAAD ELSAYAD, ALI I. EL DESOUKY, MOFREH MOHAMED SALEM, AND MAHMOUD BADAWY

Department of Computer Engineering and Systems, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt
Corresponding author: Mahmoud Badawy (engbadawy@mans.edu.eg)

This work was supported in part by the U.S. Department of Commerce under Grant BS123456.

ABSTRACT Recently, the design and implementation of new healthcare systems have gained an interest in both industry and academia. The amalgamation between the Internet of Things, cloud, edge computing, and big data helps the proliferation of new scenarios for smart medical services and applications. Deep learning is currently paying a lot of attention for its utilization with big healthcare data. To this end, the main objective of this study is to propose a Deep Learning H$_2$O (DLH$_2$O) framework for improving the performance and selection of the optimal features to predict emergency cases. The proposed DLH$_2$O framework consists of data preprocessing layer, feature selection layer and deep learning layer. The DLH$_2$O framework aims to find the optimal subset of features and minimize the error of the classification through a proposed new variant of the Whale Optimization Algorithm (WOA) called ACP-WOA. The proposed changes have been done on the following parameters $a$, $a_2$, $A$, and $C$ which should affect both exploration and exploitation of WOA. The experiments conducted in order to test the validity of DLH$_2$O Framework. In regard to the datasets, five experiments are performed for this purpose. The results demonstrate the superiority of ACP-WOA compared to the other state-of-the-art meta-heuristic algorithms in terms of time, error, and scalability. The proposed ACP-WOA is also tested on CEC2017 benchmark functions and proves its superiority over WOA in terms of accuracy.

INDEX TERMS Deep learning, H$_2$O framework, healthcare, neural networks, optimization, whale optimizer.

I. INTRODUCTION Throughout the world, there has been a gradual increase in life expectancy leading to a dramatic increase in the number of elderly people [1]. Recently, there has been extensive research indicate that approximately 89% of the elderly are likely to live independently, and 80% of older people over the age of 65 suffer from at least one chronic disease, making it difficult to take care of themselves [2]. Emergency cases emerge very quickly and need immediate intervention and medical support. Faster response in such cases can be life-saving. Technology can play a major role in such cases [3].

Health care systems are an important field in which the Internet of Things (IoT) promises substantial changes and has a major impact [4]. IoT has now become one of the 21$^{st}$ century’s most effective connectivity paradigms. IoT expands the Internet definition as all things in our everyday life are part of the Internet due to their communications and computational capacities [5], [6]. The Body Sensor Network (BSN) is one of the most important technologies used in the modern healthcare system based on IoT. It is essentially a collection of low-power and lightweight wireless sensor nodes used to monitor the functions of the human body and its surroundings [7]. The synergy between healthcare and technology has taken a big leap across the world. For example, the IoT and Big Data Analytics are increasingly gaining popularity for the next generation of eHealth and mHealth services [27].

In the world of the IoT, the things and persons in charge are at the center of a complex network of device-based interactions. It’s no big surprise that IoT has a complex ecosystem. The interconnected devices produce large amounts of data and they require extreme-scale parallel computing systems to perform processing [8]. The convergence of IoT and computing technologies such as cloud, edge, and fog computing has become necessary. Modern healthcare system is also closely connected with research in big data. However, research on
modern healthcare system has not considered the issue of real-time requirements. Training data mining and machine learning algorithms from the generated data is a key challenge because this process can take a very long time [9]. The main challenge here is the time and the accuracy of the learning process. The accuracy should be maximized while the time should be minimized so this could be seen as an optimization problem.

Evolutionary Algorithms (EAs) have been proposed for resolving optimization problems of big data which include a lot of variables and require to be investigated in a short period of time; however, most of them lack scalability problems to mitigate big data problems [10]. Nature-inspired meta-heuristic algorithms imitate biological or physical phenomena when solving optimization problems. They can be classified into three groups: physics, swarm, and evolution based methods. Physics-based techniques imitate the physical rules in the universe. Swarm-based methods imitate the behavior of animals. The nature-inspired methods include evolution-based techniques that are derived from the laws of natural evolution [11].

The H2O framework is a basic framework for multi-layer neural networks that can be used to perform Deep Learning (DL) tasks. Deep Learning Architectures (DLA) are hierarchical feature extraction models, usually including multiple levels of non-linearity. DL models are able to learn useful raw data representation and demonstrate high performance on complex data such as images, voice and text [12].

The main objective of this study is to predict the emergency cases in intensive care units. Accordingly, this study introduces a Deep Learning H2O (DLH2O) Framework that can deal with the big data generated from the BSN and provides a new variant of the Whale Optimization Algorithm (WOA) called AC-Parametric WOA (ACP-WOA). The ACP-WOA extracts the most important features that will be used in the learning process. The new variant of the WOA is compared to the original WOA and to the other state-of-the-art meta-heuristic algorithms. The proposed algorithm is tested on CEC2017 benchmark functions and proves its superiority.

The paper is organized as follows: In Section 2, The related work is reviewed. In Section 3, The WOA is discussed. In Section 4, the proposed DLH2O framework is described. The proposed ACP-WOA is described in Section 5. In Section 6, the experiments are presented and the results are analyzed. In Section 7, the paper is concluded.

II. RELATED WORK
In recent years, there is a massive growth of data resulting in the creation of big data. Big data is characterized by large volume and velocity; thus it needs new high-performance processing. Classical distributed computing models include MPI (Message Passing Interface), MapReduce, and Dryad. One of the most widespread open-source implementations of MapReduce is Apache Hadoop [23]. Despite its great popularity, MapReduce (including Hadoop) doesn’t scale well when dealing with online and iterative processes. Apache Spark is considered as an alternative to Hadoop as it has the ability to perform distributed computing faster using in-memory primitives [24].

Alfian et al. [28] propose a personalized healthcare monitoring system for diabetic patients by utilizing Bluetooth Low Energy (BLE) -based sensors and real-time data processing. While Elhoseny et al. [29], proposed cloud- IoT in integrated industry 4.0 for HCS. The proposed architecture is executed using three different algorithms, Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Parallel Particle Swarm Optimization (PPSO). The proposed fitness function is composed of three essential attributes which are CPU utilization, turn-around time, and waiting time. Ijaz et al. [30], proposed HPM by combining DBSCAN-based outlier detection, SMOTE, and RF classifier. The proposed model is believed to help users to find the danger of diabetes and hypertension at the initial phase.

The H2O framework provides modern Deep Learning. Unlike the neural networks of the past, it performs very well in a variety of different problems which makes it the best algorithm to choose in such problems. The H2O framework is big data-friendly. Fast in-memory distributed parallel processing abilities of H2O can be used for better predictions [25]. H2O software depends on R, java, and python to enhance machine learning for Big Data. For deployment production, a developer doesn’t need to fear about the difference in the production and development environment. Once H2O models are created, they can be used as a java object. H2O algorithms can be employed by numerous users from business analysts to developers acquainted with a lot of programming languages. Billions of data rows in-memory can be handled by H2O by applying the techniques of in-memory compression, even with a small cluster. H2O software includes nearly all machine learning algorithms, like Naive Bayes, generalized linear modeling (such as logistic regression, linear regression.), time series, gradient boosting, k-means clustering, principal components analysis, random forest, and finally deep learning.

Modern DL differs from past neural networks as it provides scalability, training stability, and generalization with big data. Because it performs well in a number of different problems, DL becomes the choice that provides the highest accuracy. DL has many theoretical frameworks, however this study concentrates on the feed-forward design used by H2O. The neuron, a biologically imitated model of the human one, is the fundamental unit in the model. Regarding humans, the neurons send output signals with different strength travel through the synaptic junctions, then combined as input to activate connected [26].

There has been extensive research on biomedical big data sets. Mirjalili and Lewis [11] proved the superiority of the WOA algorithm in comparison to other meta-heuristic algorithms and conventional techniques. Nakamura et al. [13] conducted experiments to compare several meta-heuristic algorithms such as (BAT – FFA – PSO) but didn’t include WOA in this comparison. Sharawi et al. [14] introduced a
feature selection approach using the WOA and proved that WOA has the ability to find the best features with maximum accuracy. Hassan et al. [15] showed that a hybrid algorithm of WOA and Naïve Bayes (NB) saves storage space and accelerates the classification process. In the literature, several methods on enhancing WOA algorithms have appeared. These methods can be classified into hybridization, improvement, and variation of the traditional whale algorithm [16]. Gharachorloo and Gholizadeh et al. [16] have proposed a new hybrid PSO-WOA approach for some unconstrained benchmark research functions. Hybrid PSO-WOA is the outcome of using PSO for the phase of exploitation. The PSO weakness is its inability to cover limited search space with the presence of constant inertia weight while solving higher-order or complex design problems. WOA is used for the exploration process due to the use of the logarithmic spiral feature, which ultimately covers a vast area of unknown search space. Therefore, WOA guides the particles to a maximum value more efficiently and reduces the processing time [17].

In regard to improving the WOA, the Levy flight trajectory-based WOA (LWOA) has been proposed. The LWOA makes WOA's action more efficient and quicker and avoids premature convergence. The Levy flight trajectory can be used to improve population diversity and perform the ability to jump from the local optima. This method obtains a better trade-off between the exploration and the exploitation of the WOA [18], Elhosseini et al. [19] have introduced a new variant of WOA which aims to tune a low number of control parameters. A and C impact on exploration and exploitation. This new variant of the WOA is applied to benchmark functions and compared to some of the well-known state-of-the-art algorithms. The results show the superiority of their proposed WOA than the other compared techniques.

III. WHALE OPTIMIZATION ALGORITHM (WOA)

The Whale Optimization Algorithm (WOA) is a heuristic search algorithm that was proposed in 2016 by Seyedali Mirjalili and Andrew Lewis [16]. WOA is proposed for continuous optimization problems WOA is considered as a Wrapper based Feature Selection algorithm that proved that it’s better in performance as compared to some existing algorithms. WOA is inspired by nature came from the mimic hunting behavior of the humpback whales. Each solution in WOA is considered to be a whale. Each whale tries to find prey. There are two ways in which whales use for searching the prey location and attack them. First, whale encircles the prey, and secondly, they create a bubble net around the prey to start to eat them. According to optimization, exploration is preferring when a whale looking for the prey, and the exploitation when attacking.

A. INSPIRATION

The whales are beautiful creatures; they are considered to be the largest animals in the world. An adult whale is 30 meters long and weighs 180 tons. Whales are generally considered predators. The whales never sleep because they need to breathe from the surface of the oceans. Only half of the whale’s brain sleeps. What is unique about the whales is that they are known to be very intelligent and emotional creatures [16].

Hof and Gucht [20] think that there are specific cells in certain regions of whale brains, such as human brains. Such regions are referred to as spindle cells. These regions are responsible for human action, sentiment, and social behavior. In other words, spindle cells differentiate humans from other organisms. The main cause of whales’ smartness is that the whales have more cells than an adult human. Whales can think, understand, judge, communicate, and even get as emotional as a person, but with a lower level of smartness. It has been noted that whales (mostly killer whales) are also capable of developing their own dialect.

One of the most exciting behaviors of whales is that they are usually seen together in groups. However, they can live alone. Some of the kinds (such as killer whales) are capable of living their whole life in a family [11]. The humpback whales are characterized by their distinct hunting methods. They create distinctive bubbles along a ‘9’-shaped path or circle. This foraging behavior is named the bubble-net feeding method [21].

B. A MATHEMATICAL-BASED MODEL AND OPTIMIZATION ALGORITHM

The model of the WOA involves three phases. The first phase mimics the search for prey (the exploration phase), the second phase is the prey encircling, and the third phase represents the bubble-net feeding method of humpback whales (the exploitation phase). The pseudocode of the WOA is shown in algorithm 1.

1) PREY ENCIRCLING

In this phase, WOA begins with an initial best candidate solution which is the target prey or is close to the best. The remaining search agents update their locations regarding the best search agent consequently. This is explained by equations 1 and 2 as follows:

\[
\hat{D} = |\hat{C}.\hat{X}^*(t) - \hat{X}(t)| \quad (1)
\]

\[
\hat{X}(t + 1) = \hat{X}^*(t) - \hat{D}\hat{A} \quad (2)
\]

where \(t\) is the iteration number while \(\hat{A}\), and \(\hat{C}\) are coefficient vectors. The \(\hat{X}\) is the location vector and \(\hat{X}^*\) is the location vector of the optimal obtained solution. Finally, \(\hat{X}^*\) should be iteratively updated. The vectors \(\hat{A}\), and \(\hat{C}\) are measured by equations 3 and 4.

\[
\hat{A} = 2\hat{a}\hat{r} - \hat{a} \quad (3)
\]

\[
\hat{C} = 2\hat{r} \quad (4)
\]

where \(\hat{r}\) is defined as vector in \([0, 1]\) and \(\hat{a}\) is linearly decreased from 2 to 0 over the number of iterations. This modeling permits any search agent to update its location in the area of the present best solution and mimics prey encircling.
Algorithm 1: The Whale Optimization Algorithm

1. Initialize the whales population \( X_i (i = 1, 2, \ldots, n) \);
2. Calculate the fitness of each search agent;
3. \( X^* \) = the best search agent;
4. while \( (t < \text{The maximum number of iterations}) \) do
   
   for each agent in search agents do
   
   5. Update \( a, A, C, l, \) and \( p \);
   6. if \( p < 0.5 \) then
      
      7. if \( |A| < l \) then
         
         8. Update position of current search agent by the Eq.(1);
      
      else
         
         9. Select a random search agent \( (X_{rand}) \);
      
         10. Update position of current search agent by the Eq.(8);
      
   end
   
   else
   
   11. Update position of current search agent by the Eq.(5);
   
end
   
12. Check if any search agent goes beyond search space and amend it;
13. Calculate the fitness of each search agent;
14. Update \( X^* \) if there is a better solution;
15. \( t = t + 1 \);
16. end

The search agents can move in hyper-cubes around the optimal solution attained.

2) BUBBLE-NET ATTACKING METHOD (EXPLOITATION PHASE)

This phase involves two methods as follows:

1) Shrinking encircling mechanism: The value of \( \bar{a} \) is decreased by equation (3) and the range of \( \bar{A} \) is decreased by \( \bar{a} \). Consequently, \( \bar{a} \) is randomly placed in \([-\bar{a}, \bar{a}]\). Where \( a \) changes from 0 to 2 over the course of iterations. Putting random value for \( \bar{A} \) in \([-1, 1]\), the new position of the search agent can be determined anywhere between the original position of the agent and the current best position. Figure 1 displays the possible positions from \((X, Y)\) towards \((X^*, Y^*)\) that can be attained by \(0 \leq A \leq 1\) in a two-dimensional space.

2) Spiral updating position: as explained in figure 2 the distance between the prey and the positions of the whale is measured, after that an equation of spiral is generated between prey and whale positions to imitate the helix shape movement of whales. This can be modeled by equations 5 and 6.

\[
\bar{X}(t+1) = \bar{D} \cdot e^{bi} \cos 2\pi l + \bar{X}^*(t) \quad (5)
\]

\[
\bar{D}(t) = |\bar{X}^*(t) - \bar{X}(t)| \quad (6)
\]

Equation 6 displays the distance of the \(i^{th}\) whale to the prey (optimal solution attained), \( b \) represents a constant to define the logarithmic spiral shape and \( l \) is a random number in between \([-1, 1]\). The whale swims around its prey within a shrinking circle in a spiral-shaped path. An assumption of 50% possibility is applied to select between the two modes and find the whale’s next position by equation 7 as follows [11]:

\[
\bar{X}(t+1) = \begin{cases} 
\bar{X}^*(t) - \bar{D} \cdot \bar{A} & \text{if} \ p < 0.5 \\
\bar{X}^*(t) - \bar{D} \cdot \cos(2\pi l) & \text{if} \ p \geq 0.5 
\end{cases} \quad (7)
\]

where \( p \) is defined as a number in \([0, 1]\).

3) EXPLORATION PHASE

Whale attains a worldwide optimization in this phase. Humpback whales search for its prey randomly as shown in figure 3. The \( \bar{D} \) is randomly chosen between \([-1, 1]\) to move the search agent away from the reference whale. The updated location of a search agent is found by randomly choosing an agent that permits the whale algorithm to carry out a global search. The exploration mechanism is mathematically described by equations 8 and 9 as follows:

\[
\bar{D} = |\bar{C} \cdot X_{rand} - \bar{X}| \quad (8)
\]

\[
\bar{X}(t+1) = X_{rand} - A \cdot D \quad (9)
\]
where $X_{\text{rand}}^*$ represents a random position for a random whale that is chosen from the present population.

IV. THE PROPOSED DEEP LEARNING $H_2O$ (DLH$_2$O) FRAMEWORK

The proposed DLH$_2$O framework is used to optimize the multiclass classification problem for the intensive care unit. The H$_2$O, as shown in figure 4, consists of three layers. The first layer is the pre-processing layer which is mainly responsible for data integration and data cleaning. In this layer, the dataset is also divided into training, validation and testing subsets. The second layer is the feature selection layer where the proposed ACP-WOA is used to select the best features that would be used in the third layer which is the DL layer. The DL layer uses the best-selected features along with the best configuration of the neural network to train the neural network. In the next subsection, the layers of the proposed DLH$_2$O framework will be explained in detail.

A. THE PREPROCESSING LAYER

Mainly, data is incomplete, inconsistent, and may contain many errors. This layer is responsible for handling these issues. The data in hand is divided into three datasets. The first dataset is for a hypertensive patient. The second is for a hypotensive patient. These datasets are taken for a period of one year long for each patient. At the first phase of preprocessing, normalization is done on some columns in these datasets. Each value in these columns is rescaled. The second phase of preprocessing is to remove useless columns such as the time column. At the third and the last phase of the preprocessing, the labels for each entry in the dataset are modified from string values to numeric values. These classes are (normal, warning, alert, emergency) and converted to 1, 2, 3, 4 respectively.

B. THE FEATURE SELECTION LAYER

In this layer, the proposed ACP-WOA is used in the feature selection process. This is done in an innovative way. Simply converting the columns into zeros and ones, a binary representation, helps greatly not only in simplifying the problem but also in making it a one-dimensional problem. The proposed algorithm is then used to determine a value from 0 to $2^{NTF} - 1$, where $NTF$ is the number of total features. The generated value will be converted to a sequence of zeros and ones in a binary format. The zero marks columns or features that will not be taken into consideration in the learning process which will be done later by the neural network. On the other hand, one means that the corresponding column will be added to the list of the selected features. These selected features are then used in training the neural net to reach the highest accurate model based on our fitness function described in equation 10.

$$\text{FitnessFunction} = \frac{1}{M} \sum_{k=1}^{M} MCE(K)$$

where $MCE$ is the Mean Per-Class Error.

The learning process takes place for $M$ trials and the average of the $MCE$ in these trials is taken as a fitness function. The main goal here is to minimize the resulted value.

C. THE DEEP LEARNING LAYER

The proposed ACP-WOA algorithm, the WOA algorithm, the PSO algorithm, the BAT algorithm, and the FireFly algorithm are investigated in this layer. The objective is to assess the performance of the proposed ACP-WOA.

THE CLINICAL DATASET

These datasets belong to old patients complaining from blood pressure disorders and they have been taken from PhysioNet MIMIC-II [22].

V. THE PROPOSED ACP-WOA

The whale algorithm is a very powerful meta-heuristic algorithm that works in two phases exploring and exploitation.
In these two phases, some parameters are used to configure the algorithm and some of them affect the performance dramatically. Most of the whale variants can achieve better quality. However, it is found that researchers have paid less attention to the standard Whale’s simplicity. In this section, the main work is focused on parameters tuning in the standard WOA that are responsible for its two main phases. The proposed changes have been done on the following parameters $a$, $a_2$, $A$, and $C$ which should affect both exploration and exploitation of WOA. These changes are described in the following equations.

$$a = 2 - t \times \frac{2}{Max_{\text{iter}}^2}$$  \hspace{1cm} (11)
$$a_2 = -1 - t \times \frac{1}{Max_{\text{iter}}^2}$$  \hspace{1cm} (12)
$$A = -2 \times a \times \sin r 1 - a$$  \hspace{1cm} (13)
$$C = -2 \times \sin r 2$$  \hspace{1cm} (14)

As noticed, the parameters $a$ and $a_2$ are time variants which decrease more slowly with the change of the denominator to the square of the maximum iteration numbers which narrows the range of change. While the changing of $A$ and $C$ is converted to sinusoidal change hence the fluctuation of the change is more efficient than normal randomization. The proposed modified whale algorithm is introduced in Algorithm 2 below.

**VI. EXPERIMENTAL RESULTS**

The following experiments are performed on VMware machine running Linux OS with 2.8 GB of RAM, 2 cores of processor. The number of features is 11 so the range of numbers generated by the optimizing algorithm is from 0 to 2047. The trials are done 10 times and the average was taken as a fitness score. A neural network with a single hidden layer with 10 neurons is used. The experiments are conducted in order to test the validity of DLH$_2$O Framework. In regard to the datasets, five experiments are performed for this purpose. Three public datasets with 35233 samples: Hypertensive patient, Hypotensive patient, and Normotensive patient are used in these experiments. The number of features is 11 and the number of classes is 4.

The datasets have been partitioned into three disjoint subsets $Z_1$, $Z_2$, $Z_3$ which correspond to the training, validation and test sets, respectively. We have tested the proposed ACP-WOA against WOA, PSO, FFA, and BAT. The final subset of features will be the one that minimizes the error over $Z_3$. The related discussion will be presented in the next subsections.

**A. PERFORMANCE ANALYSIS OF THE DLH$_2$O FRAMEWORK**

1. **EXPERIMENT 1**

This experiment is performed to select the best number of neurons used in a single layer artificial neural network. 20 iterations were done. The results obtained are shown in Figure 5. The figure shows that with more neurons the error decreased. The number of neurons in the hidden layer is set to 10.

2. **EXPERIMENT 2**

The proposed optimizer is tested against the other optimizers to check its performance. This process is done over 50 iterations to reach a stable and good solution. The results of these
The multiple trials of training the neural network with different subset of features for dataset "Hypertensive patient" as selected by optimization algorithms.

The ACP-WOA and WOA outperform the other algorithms for the hypertensive patient and normotensive patient datasets. In regard to the hypotensive patient dataset, the ACP-WOA and WOA have achieved the second-best.

### EXPERIMENT 3

This experiment is done to compare the number of features selected by each optimizer. The results are shown in figure 10 and Table 2. The ACP-WOA algorithm selects an acceptable number of features and gets away from the extremes, unlike BAT which has selected all eleven features and PSO which has selected four features.

#### TABLE 1. A comparison between optimizers based on Mean_per_class_error at the last iteration.

|         | MWOA   | WOA    | FFA    | BAT    | PSO    |
|---------|--------|--------|--------|--------|--------|
| P1      | 0.182222 | 0.180499 | 0.208855 | 0.269089 | 0.217239 |
| P2      | 0.287942 | 0.286190 | 0.295339 | 0.284053 | 0.286964 |
| P3      | 0.215205 | 0.215485 | 0.255281 | 0.282261 | 0.229441 |

#### EXPERIMENT 4

The experiment is done to compare the time taken for each algorithm to run. The results are displayed in figure 11 and Table 3. Time is so critical in this application as this is an emergency case prediction. As well, the error must be taken into consideration because the algorithm may produce very
good predictions but in too much time. On the other hand, the results may be produced too quickly but with poor performance and wrong predictions. The ACP-WOA has a unique combination between time and error and has shown its superiority over the other state-of-the-art algorithms. Although the ACP-WOA is ranked the second-best in time, it is considered to be the first in error and time combination.

5) EXPERIMENT 5
The experiment is done to check the scalability in DLH₂O Framework using different optimization techniques. The results are displayed in table 4 and figures 12, 13, 14, 15, and 16. The ACP-WOA, WOA and firefly are scalable but PSO and BAT aren’t. This means that the former group can be trained with more data and generate more accurate results.
changing the search agent numbers to 100, 50 and 25. F2 is deleted so this function is skipped in tests.

VII. CONCLUSION

The healthcare environment is considered to be ‘information rich’ yet ‘knowledge poor’. Healthcare systems contain a huge data available within. However, there is a shortage of effective analysis tools to find hidden relationships in data. A deep learning framework for the prediction of emergency cases of eldest people is introduced. This DLH$_2$O framework consists of three layers pre-processing layer, feature selection layer and deep learning layer. A new variant of the Whale Optimization Algorithm (WOA) called AC-Parametric WOA (ACP-WOA) is introduced. Five tests have been done to validate DLH$_2$O Framework using three public datasets. ACP-WOA has been compared against WOA, PSO, FFA and BAT. ACP-WOA has achieved an acceptable time, error, scalability and accuracy. The proposed algorithm proved to be more scalable and faster than the original whale optimizer. Also ACP-WOA is more accurate than most of the compared algorithms. As future work, new metaheuristic algorithms will be checked against our proposed whale algorithm. Also, the effect of increasing the complexity of datasets used in the deep learning is still to be investigated.
### TABLE 5. The comparison between WOA and MWOA on CEC2017 functions.

| Function | Algorithm | SA=100  | SA=50   | SA=25   |
|----------|-----------|---------|---------|---------|
| F1       | WOA       | 2.68E+10| 3.76E+10| 4.97E+10|
|          | MWOA      | 1.66E+10| 2.64E+10| 3.78E+10|
| F2       | WOA       | Skipped | Skipped | Skipped |
|          | MWOA      | Skipped | Skipped | Skipped |
| F3       | WOA       | 2.79E+05| 2.80E+05| 2.76E+05|
|          | MWOA      | 2.58E+05| 2.84E+05| 2.73E+05|
| F4       | WOA       | 6.18E+03| 9.68E+03| 1.32E+04|
|          | MWOA      | 3.87E+03| 6.24E+03| 9.52E+03|
| F5       | WOA       | 9.06E+02| 9.48E+02| 9.77E+02|
|          | MWOA      | 8.97E+02| 9.25E+02| 9.56E+02|
| F6       | WOA       | 6.88E+02| 6.93E+02| 7.00E+02|
|          | MWOA      | 6.87E+02| 6.90E+02| 7.00E+02|
| F7       | WOA       | 1.39E+03| 1.43E+03| 1.47E+03|
|          | MWOA      | 1.34E+03| 1.40E+03| 1.46E+03|
| F8       | WOA       | 1.13E+03| 1.16E+03| 1.18E+03|
|          | MWOA      | 1.11E+03| 1.15E+03| 1.17E+03|
| F9       | WOA       | 1.43E+04| 1.64E+04| 1.76E+04|
|          | MWOA      | 1.37E+04| 1.53E+04| 1.96E+04|
| F10      | WOA       | 8.27E+03| 8.72E+03| 9.18E+03|
|          | MWOA      | 8.21E+03| 8.59E+03| 8.98E+03|
| F11      | WOA       | 2.09E+04| 2.59E+04| 3.53E+04|
|          | MWOA      | 1.73E+04| 2.19E+04| 3.12E+04|
| F12      | WOA       | 2.64E+09| 5.28E+09| 9.39E+09|
|          | MWOA      | 1.39E+09| 2.83E+09| 5.76E+09|
| F13      | WOA       | 3.13E+08| 9.98E+08| 2.68E+09|
|          | MWOA      | 1.10E+08| 3.63E+08| 1.35E+09|
| F14      | WOA       | 3.10E+06| 4.67E+06| 1.04E+07|
|          | MWOA      | 2.90E+06| 4.28E+06| 6.50E+06|
| F15      | WOA       | 5.63E+07| 1.92E+08| 5.67E+08|
|          | MWOA      | 1.68E+07| 7.18E+07| 2.89E+08|
| F16      | WOA       | 4.50E+03| 5.09E+03| 5.93E+03|
|          | MWOA      | 4.45E+03| 4.92E+03| 5.35E+03|
| F17      | WOA       | 3.04E+03| 3.14E+03| 4.37E+03|
|          | MWOA      | 2.88E+03| 3.04E+03| 5.19E+03|
| F18      | WOA       | 2.81E+07| 3.95E+07| 8.30E+07|
|          | MWOA      | 1.91E+07| 3.49E+07| 6.88E+07|
| F19      | WOA       | 1.30E+08| 2.95E+08| 7.35E+08|
|          | MWOA      | 5.76E+07| 1.15E+08| 3.77E+08|
| F20      | WOA       | 3.01E+03| 3.14E+03| 3.23E+03|
|          | MWOA      | 2.97E+03| 3.07E+03| 3.16E+03|
| F21      | WOA       | 2.68E+03| 2.73E+03| 2.77E+03|
|          | MWOA      | 2.66E+03| 2.70E+03| 2.74E+03|
| F22      | WOA       | 8.78E+03| 9.70E+03| 1.03E+04|
|          | MWOA      | 8.70E+03| 9.58E+03| 1.00E+04|
| F23      | WOA       | 3.25E+03| 3.30E+03| 3.39E+03|
|          | MWOA      | 3.20E+03| 3.24E+03| 3.33E+03|
| F24      | WOA       | 3.36E+03| 3.42E+03| 3.55E+03|
|          | MWOA      | 3.33E+03| 3.36E+03| 3.44E+03|
| F25      | WOA       | 3.92E+03| 4.39E+03| 5.07E+03|
|          | MWOA      | 3.60E+03| 3.89E+03| 4.37E+03|
| F26      | WOA       | 9.27E+03| 1.01E+04| 1.08E+04|
|          | MWOA      | 8.98E+03| 9.31E+03| 1.01E+04|
| F27      | WOA       | 3.62E+03| 3.77E+03| 3.88E+03|
|          | MWOA      | 3.55E+03| 3.66E+03| 3.80E+03|
| F28      | WOA       | 5.34E+03| 6.11E+03| 7.02E+03|
|          | MWOA      | 4.70E+03| 5.29E+03| 6.27E+03|
| F29      | WOA       | 6.00E+03| 6.18E+03| 7.05E+03|
|          | MWOA      | 5.62E+03| 5.90E+03| 6.32E+03|
| F30      | WOA       | 1.88E+08| 4.04E+08| 7.14E+08|
|          | MWOA      | 1.30E+08| 2.16E+08| 3.63E+08|

Success: 29/29
MOFREH MOHAMED SALEM received the Ph.D. degree from Strathclyde University, U.K., in 1985. He was the Director of Software Engineering Unit, Faculty of Engineering, from 2001 to 2006. He was the Head of the Computers Engineering and Control Department, Faculty of Engineering, Mansoura University, Egypt, from 2004 to 2008, where he is currently a member of the Computer Center Council. He was the Dean of the High Institute for Computers in Mansoura, from 2008 to 2011. He has published 92 scientific articles in international journals or periodicals and conferences of computers engineering. His current research interests include software engineering, computer systems design, parallel processing, computer networks, cloud computing, and big data. He was a member of High Scientific Consul for Colleges of Technology, Kingdom of Saudi Arabia, from 1998 to 2001. He received the Encouragement Award in computer engineering, Egypt, in 1992. He has supervised 42 M.Sc. and Ph.D. degree students in the fields of computer engineering. He has approved and discussed 157 of M.Sc. of Ph.D. thesis. He is also a Consultant with the Studies and Researches and Consultant Engineering Center, Egypt. He is Arbitration and Examine Research Faculty Member of promotion to the Rank of a Professor and an Associate Professor with the Universities of Egypt, Saudi Arabia, and Jordan. He is an Arbitration Scientific Researcher of publication in international journals, periodicals, and conferences.

MAHMOUD BADAWY received the M.Sc. and Ph.D. degrees in computer engineering and control systems from Mansoura University, Egypt. He is currently an Associate Professor with the Computer and Systems Engineering Department, Faculty of Engineering, Mansoura University. His research interests include computer networking, distributed control systems, database management systems, mobile robots, the Internet of Things, and cloud computing.