**EAOA: An Enhanced Archimedes Optimization Algorithm for Feature Selection in Classification**

ABEER S. DESUKY1, SADIQ HUSSAIN2, SAMINA KAUSAR3, MD. AKHTARUL ISLAM4, AND LAMIAA M. EL BAKRAWY1

1Faculty of Science, Al-Azhar University, Cairo 11754, Egypt
2Examination Branch, Dibrugarh University, Dibrugarh, Assam 786004, India
3Department of Computer Science and IT, University of Kotli Azad Jammu and Kashmir, Kotli Azad Kashmir 11100, Pakistan
4Statistics Discipline, Science Engineering & Technology School, Khulna University, Khulna 9208, Bangladesh

Corresponding author: Samina Kausar (saminamalik7@yahoo.com)

**ABSTRACT** Feature selection plays a crucial role in order to mitigate the high dimensional feature space in different classification problems. The computational cost is reduced, and the accuracy of the classification is improved by reducing the dimension of feature space. Hence, in the classification task, finding the optimal subset of features is of utmost importance. Metaheuristic techniques have proved their efficacy in solving many real-world optimization issues. One of the recently introduced physics-inspired optimization methods is Archimedes Optimization Algorithm (AOA). This paper proposes an Enhanced Archimedes Optimization Algorithm (EAOA) by adding a new parameter that depends on the step length of each individual while revising the individual location. The EAOA algorithm is proposed to improve the AOA exploration and exploitation balance and enhance the classification performance for the feature selection issue in real-world data sets. Experiments were performed on twenty-three standard benchmark functions and sixteen real-world data sets to investigate the performance of the proposed EAOA algorithm. The experimental results based on the standard benchmark functions show that the EAOA algorithm provides very competitive results compared to the basic AOA algorithm and five well-known optimization algorithms in terms of improved exploitation, exploration, local optima avoidance, and convergence rate. In addition, the results based on sixteen real-world data sets ascertain that reduced feature subset yields higher classification performance when compared with the other feature selection methods.

**INDEX TERMS** Archimedes optimization algorithm, classification, feature selection, machine learning, metaheuristic algorithms, optimization algorithms.

**I. INTRODUCTION**

Machine classification is a prominent job in data mining and machine learning, which focuses on classifying that systematically, categorized each object in the data set based on their similar features [1]. It isn’t easy to ascertain which features are beneficial without prior knowledge. As an outcome, numerous features are introduced in the data set comprising redundant, irrelevant, and relevant features. Irrelevant and redundant features mitigate the performance of the classifier because of the enormous search space termed as “the curse of dimensionality”. Feature selection can solve this problem by selecting only significant features for the task of classification. Thus, feature selection could improve the classifier’s performance by eradicating the number of features, simplifying the learned classification method, curtail the training time, data visualization, and data understanding.

There are four crucial phases in feature selection: a generation process that produces a feature subset. A search technique is essential to find out the most favorable candidates of feature subset and determine the most suitable feature subset among the subsets; an evaluation process is employed [2]. A stopping criterion is adopted afterward, and to test the validity of the subset, a validation process is exploited. Choosing an optimal subset from a large number of features is a considerable hazard. Various search strategies exist to solve
this big menace, namely heuristic search strategy, complete search strategy, and random search strategy. Search strategies suffer from certain limitations. High computational cost, multifaceted modality of the search setting, growing problem dimensionality, and convergence towards local minima mention a few. The feature selection method suffers from overfitting and substantial computational time when the number of features is large. Classification errors may occur because of an overfitted model that may blunder slight fluctuations for vital variance in the data.

Metaheuristic methods have produced outstanding outcomes in different real-life optimization issues [3]. These methods are easy to implement, simple, and can evade local minima. In the case of such methods, trade the balance between exploitation and exploration is the key. The successful techniques are those that uphold such stability on a huge range of optimization issues. New optimization methods derived from physics, mathematics, chemistry, and their powerful phenomena are introduced besides animals and insects. The essential techniques inspired from the discipline of Physics are Henry law based Henry gas solubility optimization [4], based on Coulomb and Newtonian’s laws, an optimization method called as charged system search [5], ray optimization based on ray theory [6] and gravitational theory is observed by gravitational search algorithm [7]. The laws of physics have already yielded state-of-the-art performances when formulated into optimization methods. Hashim et al. [8] devised Archimedes Optimization Algorithm (AOA), whose basis is Archimedes’ principle from the physics domain. AOA is a population-based metaheuristic method. Population-based methods perform a search with several preliminary points in an analogous style like swarm-based metaheuristics. Some of the metaheuristic algorithms have been effectively solved several optimization problems, while other methods have been utilized to enhance the classifier’s performance.

The law of buoyancy is obeyed by Archimedes Principle [8]. It explains the association between the buoyant force and an item submerged in fluid. The upward force exerted is an item’s buoyancy that is equivalent to the weight of the displaced fluid. If the displaced fluid is less than the item’s weight, the item will sink. If the weight of the item and displaced fluid are equal, then the item will float. The population entities are the objects submerged in the fluid in the case of AOA. The items have acceleration, density, and volume, which are the vital players in the buoyance of the item. To touch a mark where items are impartially buoyant and displaced fluid are equal, then the item will float. If the weight of the item’s weight, the item will sink. If the weight of the item and displaced fluid are equal, then the item will float. The optimization problems are handled by AOA suitably with numerous local best solutions. The paramount global solution can be found in a broader area by investigating a population of solutions.

We had reviewed some of the original or modified optimization methods to solve the feature selection issue. Liu et al. [9] introduced an improved feature selection (IFS) technique by utilizing support vector machines with the F-score method and a modified Multi-Swarm Particle Swarm Optimization (MSPSO). Higher generalization capability was achieved by performing feature selection and kernel parameters simultaneously. To tackle the feature selection issue, Al-Tashi et al. [10] hybridized grey wolf optimization (GWO) and particle swarm optimization (PSO). The wrapper-based technique K-nearest neighbors classification method with Euclidean separation metric was applied to find the best feature subset. The method was assessed using the 18 standard benchmark data sets. Neggaz et al. [11] introduced a new technique based on Henry gas solubility optimization (HGSO) to choose key features and improve classification accuracy. They employed various data sets with different feature sizes, from small to massive, and evaluated against well-known recent metaheuristic algorithms. Wilcoxon’s rank-sum non-parametric statistical test was performed to make a statistical comparison of the proposed method with other methods. Based on the graph clustering approach and ant colony optimization, a novel feature selection technique was devised by Moradi et al. [12]. Their method worked in three stages. In the first stage, the complete feature set was embodied as a graph. In the next stage, a community detection method was employed to discriminate the features into various clusters. In the last stage, the ACO technique was applied to select the final subset of features using a new search strategy. To tackle the same feature selection problem, a hybrid optimization method termed SSAPSO was proposed by Ibrahim et al. [13] by integrating PSO with the slap swarm algorithm (SSA). The efficacy of the exploitation and the exploration steps was enhanced. The empirical results provided evidence that the accuracy and performance were improved without affecting the computational effort. The same issue was addressed by Too et al. [14] by introducing a binary form of Harris hawk optimization. To convert the continuous variable into a binary one, the proposed method was equipped with a V-shaped or S-shaped transfer function. Quadratic binary Harris hawk optimization, another variant, was applied to enhance the performance of the earlier version.

In the literature, different metaheuristic algorithms were proposed to enhance the performance of feature selection issue. However, as far as the authors are aware, this is the first time AOA algorithm or a modified version of it is proposed to solve the feature selection issue. In this paper, EAOA will be utilized to improve the AOA exploration/exploitation balance and enhance the classification performance in real-world data sets by optimized feature selection.

The main contributions of this work are summarized as follows:
1. A new Enhanced Archimedes Optimization Algorithm (EAOA) is proposed to enhance the optimization efficiency and accuracy of AOA by adding a new parameter that depends on the step length of each individual while revising the individual location.
2. The proposed EAOA method is tested on twenty-three benchmark functions.
3. The significant Wilcoxon Rank-sum test is applied to present a fair comparison. The results provide evidence that the suggested EAOA escapes the local optima, provides a better convergence rate with an enhanced balance between exploration/exploitation processes and performs statistically significant as compared to the basic AOA and five well-known optimization algorithms.

4. As a potential application, EAOA is used as a feature selection algorithm and sixteen real-world data sets are employed to test the performance of the proposed EAOA algorithm on selecting the optimal subset of features of the training data and the results of the same features from the test set are compared with some well-known optimization algorithms.

5. EAOA is used to mitigate the features, and learning is performed using the Support Vector Machines (SVM) classifier in all performed experiments. Wilcoxon signed-rank test is also applied, and the results showcase that the proposed feature selection method statistically outperforms the AOA algorithm and other comparable optimization techniques based on the recorded results in terms of accuracy, G-Mean, F-Measure metrics. Also, EAOA yields superior performance than the basic AOA and the comparable feature selection methods in terms of Runtime, AUC, specificity, precision, and sensitivity for almost all used real-world data sets.

All experimental results prove that the proposed algorithm has a better improved ability than other state-of-the-art algorithms to choose the optimal subset of features that reduce the computational complexity (Runtime and space) by reducing the dimension of feature space while improving the classification performance.

II. MATERIALS AND METHODS

A. ARCHIMEDES OPTIMIZATION ALGORITHM (AOA) [8]

Archimedes Optimization Algorithm (AOA) is an algorithm that is based on a population where the population individuals are immersed objectives. Similar to algorithms that are based on population, AOA also instigates examination procedure with a primary population of individuals with accelerations, densities, and volumes of randomness. AOA first starts the evaluation of the fitness of the preliminary population then begins with the iteration until the conclusion condition happens. Every individual’s density and volume were updated in each iteration. Considering the condition of its collision with the other neighboring individuals, the acceleration of individuals is revised. The reorganized volume, density, and acceleration fix the new location of an individual. The elaborated mathematical appearance of AOA steps.

Step 1—Initialization: Prepare the situations of all individuals utilizing (1):

\[ x_i = \text{ll}_i + \text{rand} \times (\text{ul}_i \times \text{ll}_i) \quad i = 1, 2, \ldots, N \]  

(1)
TABLE 1. Unimodal functions (F₁ – F₇).

| Function                                      | V_no | Range     | fₘᵟₚᵣᵦ|
|-----------------------------------------------|------|-----------|---------|
| F₁(x) = Σᵢ₌₁ⁿ xᵢ²                               | 30   | [-100, 100] | 0       |
| F₂(x) = Σᵢ₌₁ⁿ |xᵢᵢ| + Πᵢ₌₁ⁿ |xᵢᵢ|                      | 30   | [-10, 10]     | 0       |
| F₃(x) = Σᵢ₌₁ⁿ (Fᵢ₋₁₋₁ xᵢ)²                   | 30   | [-100, 100] | 0       |
| F₄(x) = maxᵢ(|xᵢᵢ|, 1 ≤ i ≤ n)              | 30   | [-100, 100] | 0       |
| F₅(x) = Σᵢ₌₁ⁿ [100(xᵢ₊₁ᵢ₋₁ – xᵢᵢ)² + (xᵢᵢ – 1)²] | 30   | [-30, 30]   | 0       |
| F₆(x) = Σᵢ₌₁ⁿ (|xᵢᵢ| + 0.5)²                  | 30   | [-100, 100] | 0       |
| F₇(x) = Σᵢ₌₁ⁿ i xᵢᵢ + rand(0, 1)             | 30   | [-1.28, 1.28] | 0       |

where in a population of N individuals, xᵢᵢ is the iᵗʰ individual. lᵢᵢ is the lower limit, uᵢᵢ is the upper limit of the search space and rand is a vector with D dimension arbitrarily produces numeral within the range [0, 1].

Reset density (d) and volume (v) for each iᵗʰ individual utilizing (2):

\[ dᵢᵢ = \text{rand}; \quad vᵢᵢ = \text{rand} \]  

(2)

And lastly, start acceleration (ac) of iᵗʰ object utilizing (3):

\[ acᵢᵢ = lᵢᵢ + \text{rand} \times (uᵢᵢ \times lᵢᵢ) \]  

(3)

In this stage, assess starting population and choose the individual which has the finest fittest value. Allocate xᵢᵢ, dᵢᵢ, vᵢᵢ and acᵢᵢ, where xᵢᵢ is the finest individual, dᵢᵢ, vᵢᵢ and acᵢᵢ are the density, volume and acceleration related with the finest xᵢᵢ found.

**Step 2—Renew Volumes and Densities:** The volume and density of individual i for the repetition t + 1 is reorganized utilizing (4)

\[ dᵢᵢ₊₁ = dᵢᵢ + \text{rand}(dᵢᵢ – dᵢᵢ) \]  

\[ vᵢᵢ₊₁ = vᵢᵢ + \text{rand}(vᵢᵢ – vᵢᵢ) \]  

(4)

where dᵢᵢ and vᵢᵢ are the density and volume related with the finest individual found to this point, and a uniformly distributed random number is rand.

**Step 3—Density Factor and Transfer Operator:** In the commencement, individual’s collision occurs and, after a while, the individuals attempt to grasp at symmetry condition.

In AOA, this condition is applied with the assistance of transfer operator TO, which converts examination from exploration to exploitation, definite utilizing (5):

\[ TO = \exp\left(\frac{t – tₘᵟₚᵣᵦ}{tₘᵟₚᵣᵦ}\right) \]  

(5)

where transfer operator TO upsurges progressively with time till attainment of one. Where t is the number of iteration and tₘᵟₚᵣᵦ is the maximum iterations.

Likewise, density declining factor (ddf) also supports AOA on a global search to local. Which declines as time passes utilizing (6):

\[ ddfᵢ₊₁ = \exp\left(\frac{tₘᵟₚᵣᵦ – t}{tₘᵟₚᵣᵦ} \right) - \left(\frac{t}{tₘᵟₚᵣᵦ}\right) \]  

(6)

where ddfᵢ₊₁ declines as time passes that provides the capability to converge in previously recognized hopeful region. Worthwhile to indicate that appropriate handling of this variable will settle equilibrium amongst exploitation and exploitation in the algorithm of AOA.

**Step 4.1—Exploration Phase (Collision Between Individuals Happens):** If TO ≤ 0.5, a collision between individuals happens, choose a material which is random and revise the object’s acceleration for repetition t + 1 using (7):

\[ acᵢᵢ₊₁ = \frac{dᵢᵢ + vᵢᵢ + acᵢᵢ}{dᵢᵢ₊₁ + vᵢᵢ₊₁} \]  

(7)

where dᵢᵢ is density, vᵢᵢ is volume, and acᵢᵢ is the acceleration of iᵗʰ individual. Likewise, vᵢᵢ, acᵢᵢ, and dᵢᵢ are the volume, acceleration, and density of the material which is random. It is significant to indicate that TO ≤ 0.5 guarantees exploration in the time of 1/3 of iterations. Utilizing a value that is not 0.5 leads to alteration exploration-exploitation performance.
Step 4.2—Exploitation Phase (no Collision Between Individuals): In the condition where \( \text{TO} > 0.5 \), there exist no collision amongst individuals, revise individual’s acceleration to iteration \( t + 1 \) utilizing (8):

\[
ac_i^{t+1} = \frac{d_{\text{best}} + v_{\text{best}} + ac_{\text{best}}}{d_i^{t+1} + v_i^{t+1}}
\]

Step 4.3—Normalize Acceleration: Utilizing equation (9), acceleration need to normalize to compute the proportion of variation:

\[
a_i^t_{\text{norm}} = u \times \frac{a_i^{t+1} - \min(ac)}{\max(ac) + \min(ac)} + l
\]

where \( u = 0.9 \) and \( l = 0.1 \) are the range of normalization. The \( a_i^t_{\text{norm}} \) illustrate the proportion of steps that every agent will vary. In the condition of individual \( i \) which is afar from global optimal, acceleration value will be high, meaning that the individual will remain in the exploitation stage or exploration phase which explains the way of exploration transforms from the phase of exploration to the stage of exploitation. In the usual condition, the acceleration element starts by a huge value that declines as time passes. This assists the search agents to the passage concerning the global finest results and consequently they back off from the solutions of local. Nevertheless, significant things indicate that there can persist a small number of search agents which require additional time to visit in the stage of exploration than usual circumstances. Therefore, AOA accomplishes stability concerning exploitation and exploration.

Step 5—Update Position: If \( \text{TO} \leq 0.5 \) (exploitation phase), the \( i^{th} \) individual’s location for following iteration \( t + 1 \) using (10):

\[
x_i^{t+1} = x_i^t + K + Con_2 \times \text{rand} \times a_i^{t+1}_{\text{norm}} \times ddf \times (x_{\text{rand}} - x_i^t)
\]

where \( Con_1 \) is constant which is 2. Otherwise, if \( \text{TO} > 0.5 \) (exploitation phase), the individuals revise their locations utilizing (11).

\[
x_i^{t+1} = x_{\text{best}} + Con_2 \times \text{rand} \times ddf \times (x_{\text{best}} - x_i^t)
\]

where \( Con_2 \) is considered as constant which is 6. \( T \) upsurges more and more which is proportionate to the transfer operator and it is definite utilizing \( T = Con_3 \times \text{TO} \), where \( Con_3 \) is considered as constant which is 2. \( T \) upsurges gradually considering the time in limit \( [Con_3 \times 0.3, 1] \) and takes a convinced proportion from the finest location, primarily. It begins with small proportions as this produces step-size of the random walk will be large and large difference concerning the finest position and current position.

While the search is occurring, this proportion upsurges progressively to decline the difference between the finest location and the current situation and consequently attain a suitable equilibrium between exploration and exploitation. \( K \) is the flag to variate the direction of motion utilizing (12).

\[
k = \{+1 \text{ if } q \leq 0.5 \text{ and } -1 \text{ if } q > 0.5\}
\]

Algorithm 1 Pseudocode of EAOA

**Procedure** EAOA \((N, t_{\text{max}}, Con_1, Con_2, Con_3, Con_4, l_l, \text{and } u_l)\)

Start individual’s population with random locations, volumes, and densities utilizing the equations (1), (2), and (3), respectively.

Assess the primary population and choose one individual with the best value.

Set iteration counter \( t = 1 \)

While \( (t \leq t_{\text{max}}) \)

For every individual \( i \)

Revise volume and density of every individual utilizing equation (4).

Revise transfer and density declining factors \( TO \) and \( ddf \) utilizing equations (5) and (6) respectively.

If \( (TO \leq 0.5) \)

Exploration phase

Revise acceleration utilizing (7) and utilize equation (9) to normalize acceleration.

Revise location utilizing equation (10)

else

Exploitation phase

Revise acceleration utilizing (8) and utilize equation (9) to normalize acceleration.

Revise direction flag utilizing (12)

\( T = Con_3 \times TO \)

\( Mu = pi \times 10 \times \text{(TO)} \)

\( yr = 2 \times \text{rand} \times \text{(size}(x_i^t)) - 1 \)

\( dx = pi \times ((1 + mu_{\text{abs}}(yi) - 1)/mu \times \text{sign}(yr) \times (u_l - l_l)) \)

Revise location utilizing equation (11)

End If

End While

Evaluate every individual and select the one with the finest value

Set \( t = t + 1 \)

End Procedure

where \( q = 2 \times \text{rand} - Con_4 \) and \( Con_4 \) is a constant equals 0.5.

Step 6—Evaluation: Assess every object utilizing the objective function and remember the finest result originate till now. Allocate \( x_{\text{best}}, d_{\text{best}}, v_{\text{best}} \text{ and } ac_{\text{best}} \)

B. ENHANCED ARCHIMEDES OPTIMIZATION ALGORITHM (EAOA) AND ITS ADAPTATION FOR FEATURE SELECTION

In the AOA algorithm, the individuals revise their locations based on the revised densities, volumes, and accelerations. EAOA algorithm is proposed to enhance the basic AOA performance by adding a new parameter \( Mu \) depends on the step length of each individual. In the exploitation phase, the parameter \( Mu \), upper and lower limits of search space are employed to enhance the balance between exploration and exploitation of the search space while revising the individual
TABLE 2. Multimodal functions ($F_8 - F_{13}$).

| Function                                                                 | $V_{no}$ | Range          | $f_{min}$   |
|--------------------------------------------------------------------------|----------|----------------|-------------|
| $F_8(x) = \sum_{i=1}^{n} -x_i \sin \left(\sqrt{|x_i|}\right)$           | 30       | [-500, 500]    | -418.9829x5 |
| $F_9(x) = \sum_{i=1}^{n} [x_i^2 - \cos \cos (2\pi x_i) + 10]$          | 30       | [-5.12, 5.12]  | 0           |
| $F_{10}(x) = -20 \exp \left(-0.2 \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i^2 \right) - \exp \left(-\frac{1}{\pi} \sum_{i=1}^{n} \cos \cos (2\pi x_i)\right) + 20 + e$ | 30       | [-32, 32]      | 0           |
| $F_{11}(x) = \frac{1}{4000} \sum_{i=1}^{n} x_i^2 - \pi \sum_{i=1}^{n} \cos \cos \left(\frac{x_i}{\sqrt{i}}\right) + 1$ | 30       | [-600, 600]    | 0           |
| $F_{12}(x) = \frac{\pi}{n} \left[10 \sin \left(\pi y_i\right) + \sum_{i=1}^{n} (y_i - 1)^2 \left[1 + 10 \left(y_i, 1\right)\right] + \left(y_i - 1\right)^2 + \sum_{i=1}^{n} u(x_i, 10, 100, 4)\right]$ | 30       | [-50, 50]      | 0           |
| $y_i = 1 + \frac{x_i + 1}{4} \left[u(x_i, a, k, m) = \{k(x_i - a)^m x_i > a; 0 - a < x_i < a k(-x_i - a)^m x_i < -a\}\right]$ |                       |                |             |
| $F_{13}(x) = 0.1 \left[3(\pi x_i) + \sum_{i=1}^{n} (x_i - 1)^2 \left[1 + (3\pi x_i + 1)\right] + (x_i - 1)^2 \left[1 + (2\pi x_i)\right]\right] + \sum_{i=1}^{n} u(x_i, 5, 100, 4)$ | 30       | [-50, 50]      | 0           |

The performance of EAOA is validated using 23 benchmark functions and utilized for solving feature selection problems on real-world data sets. The proposed EAOA has the ability to escape the local optima when tested on benchmark functions and presented an improved convergence rate with enhanced exploration/exploitation balance comparing to the basic AOA algorithm and Ant Lion Optimizer (ALO) [15], Grey Wolf Optimizer (GWO) [16], Whale Optimization Algorithm (WOA) [17], Particle Swarm Optimization (PSO) [18], and Giza Pyramids Construction (GPC) [19].

Moreover, the EAOA method is applied for feature selection on different real-world data sets. The application of EAOA for feature selection issues is performed by selecting the optimal subset of features from the training set using the enhanced optimization capability of the proposed EAOA, which starts its search with randomly generated search individuals. The proposed algorithm uses the binary encoding type as its representation scheme for selecting the optimal features. In this type of encoding, a binary array represents each search individually, and the features are treated as either existed (1) or removed (0). The 1s represent the stayed features, while 0s represent the deleted features from the training set. The G-mean of search individuals in EAOA is evaluated, which is defined in Equation (17), as fitness value.

Figure 1 depicts the flowchart of EAOA method and its implication in the feature selection problem for the real-world data sets. As shown in the figure, the original data set is divided into training, validation, and testing subsets. The proposed EAOA focuses to find an optimal feature subset from the training set and is tested using the validation set. When the termination criteria are satisfied, the best search individual is assumed as converged and this search individual that has the highest fitness value is decoded for the solution that consists of a reduced feature set.

To assess the performance of the EAOA feature selection method, experiments are conducted, and the results are compared with state-of-the-art feature selection methods.

Support Vector Machine (SVM) is used in our experiments to measure the classification performance of the reduced data resulting from the EAOA feature selection.
TABLE 3. Fixed-dimension multimodal functions ($F_{14} – F_{23}$).

| Function | $V_{no}$ | Range    | $f_{min}$ |
|----------|----------|----------|-----------|
| $F_{14}(x) = \left( \frac{1}{500} + \frac{\Sigma_{i=1}^{25} 1}{j + \Sigma_{i=1}^{3}(x_i - a_{ij})^6} \right)^{-1}$ | 2        | [-65, 65] | 1         |
| $F_{15}(x) = \Sigma_{i=1}^{11} a_i x_i (b_i x_1 + b_j x_2)^2$ | 4        | [-5, 5]  | 0.00030   |
| $F_{16}(x) = 4x_1^2 - 2.1x_1^8 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$ | 2        | [-5, 5]  | -1.0316   |
| $F_{17}(x) = \left( x_2 - \frac{5.1}{4\pi^2} x_1^2 + \frac{2}{\pi} x_1 - 6 \right)^2 + 10(1 - \frac{1}{8\pi}) \cos \cos x_1 + 10$ | 2        | [-5, 5]  | 0.398     |
| $F_{18}(x) = (1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)) \times (30 + (2x_1 - 3x_2)^2 \times (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2))$ | 2        | [-2, 2]  | 3         |
| $F_{19}(x) = -\Sigma_{i=1}^{4} c_i \exp (-\Sigma_{j=1}^{4} a_{ij}(x_j - p_{ij})^2)$ | 3        | [1, 3]   | -3.86     |
| $F_{20}(x) = -\Sigma_{i=1}^{4} c_i \exp (-\Sigma_{j=1}^{4} a_{ij}(x_j - p_{ij})^2)$ | 6        | [0, 1]   | -3.32     |
| $F_{21}(x) = -\Sigma_{i=1}^{4} [(x - a_i)(x - a_i)^T + c_i]^{-1}$ | 4        | [0, 10]  | -10.1532  |
| $F_{22}(x) = -\Sigma_{i=1}^{4} [(x - a_i)(x - a_i)^T + c_i]^{-1}$ | 4        | [0, 10]  | -10.4028  |
| $F_{23}(x) = -\Sigma_{i=1}^{4} [(x - a_i)(x - a_i)^T + c_i]^{-1}$ | 4        | [0, 10]  | -10.5363  |

Accuracy = $\frac{TP + TN}{TP + FN + FP + TN}$  
Recall = Sensitivity = $TPR = \frac{TP}{TP + FN}$  
Specificity = $TNR = \frac{TN}{TN + FP}$  
Precision = $\frac{TP}{TP + FP}$

**Accuracy**

For evaluation metrics, let $TN$ represent true negatives, and $TP$ represents true positives, $FN$ implies false negatives, and $FP$ denotes false positives [20]–[22].

Accuracy = $\frac{TP + TN}{TP + FN + FP + TN}$  
Recall = Sensitivity = $TPR = \frac{TP}{TP + FN}$  
Specificity = $TNR = \frac{TN}{TN + FP}$  
Precision = $\frac{TP}{TP + FP}$

**Recall**

**Specificity**

**Precision**

G-mean = $\sqrt{TPR \times TNR}$  
F-Measure = $\frac{2 \times Precision \times Recall}{Precision + Recall}$

Area under the ROC (Receiver Operating Characteristics) curve (AUC) is a performance measurement for the classification problems at numerous threshold situations. ROC is a likelihood curve and AUC is used to measure or degree of separability.

**III. RESULTS AND DISCUSSION**

In this section, Experiments were performed on twenty-three benchmark functions and sixteen real-world data sets...
to investigate the performance of the proposed EAOA algorithm.

A. RESULTS OF EAOA ON BENCHMARK FUNCTIONS

Benchmark functions can be divided into three groups: (1) unimodal functions ($F_1$ – $F_7$), which have one global optimum and can evaluate the exploitation ability and convergence of an algorithm; (2) multimodal functions ($F_8$ – $F_{13}$), which have many local optima and can test the exploration ability and local optima avoidance of an algorithm; (3) fixed-dimensional multimodal functions ($F_{14}$ – $F_{23}$), which can test the consistency of an algorithm in finding a global optimum solution.

The details of these functions are presented in Tables 1, 2, and 3. In these tables, $V_{no}$ is the number of variables, $Range$ represents the range of variation of variables, and $F_{min}$ is the optimum solution reported in the
TABLE 5. Statistical results of algorithms on benchmark functions.

| Fn. | MAOA | AOA | ALO | GWO | WOA | PSO | GPC |
|-----|------|-----|-----|-----|-----|-----|-----|
| Ave | 0    | 0   | 1.2E-22 | 0.42E-09 | 1.66E-09 | 8.15E-08 | 1.81E-07 | 4.1E-14 | 9.2E-14 | 3.38E-33 | 7.5267E-33 | 1.6E-34 | 7.8E-34 |
| Std | 0    | 0   | 0     | 5.26E-5 | 8.87E-56 | 7.2E-102 | 1.4E-101 | 2.32E-19 | 4.17E-14 | 4.5264E-17 | 8.16E-17 | 8.8462E-18 | 1.58E-17 |
| Ave | 0    | 0   | 3E-196 | 0.001009 | 8.72E-05 | 3.65E-04 | 3.16E-04 | 25.0556 | 47.3877 | 4.72E-10 | 3.3329E-10 | 3.4555E-33 | 7.72E-33 |
| Std | 0    | 0   | 5.9E-106 | 0.0264 | 0.0536 | 1.44E-31 | 1.44E-31 | 0.6693 | 1.3219 | 1.19E-8 | 1.3602E-08 | 8.8462E-18 | 1.58E-17 |
| Ave | 0    | 0   | 2.7E-106 | 8.76E-51 | 2.3E-22 | 6.4294 | 0.8065 | 6.5411 | 0.1807 | 4.6276 | 0.5216 | 7.7264 | 0.1341 |
| Std | 0    | 0   | 0     | 4.54E-09 | 5.83E-10 | 1.16E-06 | 4.63E-07 | 7.04E-04 | 0.0011 | 4.93E-33 | 6.75E-33 | 1.0433 | 0.0781 |
| Ave | 0    | 0   | 3.2E-04 | 0.0232 | 0.0097 | 3.62E-04 | 2.82E-04 | 0.0012 | 6.19E-04 | 0.0075 | 0.0036 | 4.77E-05 | 3.29E-05 |
| Std | 0    | 0   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| Ave | 0    | 0   | 1.2E-04 | 0.001009 | 8.72E-05 | 3.65E-04 | 3.16E-04 | 25.0556 | 47.3877 | 4.72E-10 | 3.3329E-10 | 3.4555E-33 | 7.72E-33 |
| Std | 0    | 0   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| Ave | 0    | 0   | 1.2E-04 | 0.001009 | 8.72E-05 | 3.65E-04 | 3.16E-04 | 25.0556 | 47.3877 | 4.72E-10 | 3.3329E-10 | 3.4555E-33 | 7.72E-33 |
| Std | 0    | 0   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |

Generally, all metaheuristic algorithms face some challenges due to their stochastic nature and design. Failing to solve any of these challenges produces algorithms limitations. Some of these limitations are: getting trapped in confined areas of the search space, the exploration/exploitation imbalance, and the high computational cost. The main limitation that the AOA algorithm faced was the computational cost. Since the proposed EAOA is a modification of AOA then, it has the same limitation. Several techniques are applied for parameter sensitivity analysis in metaheuristic algorithms to solve the computational cost problem. AOA authors have applied a sensitivity analysis for general parameter configuration guidance for the AOA control variables $Con_1$ to $Con_4$ to solve this limitation. In our modification we have used the same values of AOA parameters to focus on the limitation relates to improving the exploration/exploitation balance that we try to solve in this work.
TABLE 6. P-Values of wilcoxon rank-sum test for EAOA vs. the other used optimizers.

| Fn. | AOA | ALO | GWO | WOA | PSO | GPC |
|-----|-----|-----|-----|-----|-----|-----|
| F1  | 1.21E-12 | 1.21E-12 | 1.21E-12 | 1.21E-12 | 1.21E-12 | 1.21E-12 |
| F2  | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 |
| F3  | 1.21E-12 | 1.21E-12 | 1.21E-12 | 1.21E-12 | 1.21E-12 | 1.21E-12 |
| F4  | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 |
| F5  | 2.35E-04 | 0.3183 | 3.02E-11 | 4.08E-11 | 8.48E-09 | 4.62E-10 |
| F6  | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 |
| F7  | 1.53E-05 | 8.35E-08 | 8.35E-08 | 8.35E-05 | 3.34E-11 | 8.56E-04 |
| F8  | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 | 3.02E-11 |
| F9  | 6.61E-05 | 1.21E-12 | 0.1608 | 0.3337 | 1.20E-12 | NaN |
| F10 | 0.0814 | 1.21E-12 | 6.12E-14 | 2.17E-07 | 4.17E-13 | 0.1607 |
| F11 | 0.1608 | 1.21E-12 | 8.67E-07 | 6.62E-04 | 6.62E-04 | NaN |
| F12 | 3.02E-11 | 3.99E-04 | 1.53E-05 | 7.20E-05 | 2.65E-11 | 8.99E-11 |
| F13 | 3.02E-11 | 3.02E-11 | 1.39E-06 | 8.15E-08 | 2.21E-11 | 3.02E-11 |
| F14 | 0.0144 | 0.1523 | 0.008 | 0.2643 | 0.0818 | 8.89E-10 |
| F15 | 1.64E-05 | 6.74E-06 | 0.0073 | 1.55E-09 | 4.69E-08 | 0.2519 |
| F16 | 2.01E-04 | 6.01E-08 | 0.0339 | 0.1154 | 7.57E-12 | 3.02E-11 |
| F17 | 1.43E-05 | 3.01E-11 | 1.33E-10 | 4.08E-11 | 1.21E-12 | 0.0292 |
| F18 | 1.10E-08 | 9.26E-09 | 1.21E-10 | 1.46E-10 | 1.09E-11 | 3.02E-11 |
| F19 | 3.02E-11 | 3.69E-11 | 3.02E-11 | 3.02E-11 | 4.08E-12 | 3.02E-11 |
| F20 | 2.37E-10 | 0.0703 | 0.5947 | 1.34E-05 | 0.0727 | 3.02E-11 |
| F21 | 0.2116 | 0.0406 | 1.73E-07 | 0.1297 | 0.0493 | 9.92E-11 |
| F22 | 0.001 | 0.1023 | 9.03E-04 | 0.1907 | 4.86E-04 | 1.09E-10 |
| F23 | 0.0016 | 0.3329 | 0.001 | 0.0351 | 3.73E-07 | 1.33E-10 |

The statistical results (average (Ave) and standard deviation (Std) obtained by each algorithm are shown in Table 5. This table shows that the proposed EAOA algorithm is very competitive with the basic AOA algorithm, ALO, GWO, WOA, PSO, and GPC on most functions. For unimodal functions, the EAOA algorithm achieves exact optimum results for functions $F_1$ and $F_3$ and the best optimal solutions for functions $F_2$ and $F_4$ in terms of average and standard deviation. For function $F_5$, PSO algorithm is able to obtain the best average optimal value, but the standard deviation of an algorithm is the best. In the case of function $F_7$, EAOA algorithm obtains the second-best average optimal value. These results prove that the EAOA algorithm has better exploitation capability. For multimodal and fixed-dimensional multimodal functions, the EAOA algorithm obtains the exact optimum value for functions $F_9$ and $F_{11}$. Moreover, the EAOA algorithm achieves the best optimal solutions for function $F_{10}$ and $F_{20}$ in terms of average and standard deviation and provides the best average optimal value for functions $F_8$, $F_{16}$, $F_{18}$, and $F_{20}$. EAOA algorithm also achieves the second-best average optimal value for functions $F_{22}$ and $F_{23}$. It is observed from the results that the EAOA algorithm has better exploration capability as compared to other algorithms. Figure 3 shows the convergence curves of the algorithms for six benchmark functions. As illustrated, the convergence speed of the proposed EAOA algorithm towards the optimum is faster than other algorithms only in final iterations for functions $F_1$, $F_3$, and $F_{10}$. Figure 3 also shows that
the proposed EAOA algorithm provides a better convergence rate from the initial steps of iterations for functions $F_8$, $F_{19}$ and $F_{20}$, and can avoid being trapped into local optima.

In order to validate the results statistically, nonparametric Wilcoxon Rank-sum test [23] was used to create meaningful comparison between the proposed EAOA algorithm and the other used optimizers. The calculated $p$-values of the Wilcoxon Rank-sum test are shown in Table 6. The $p$-value was compared against a significance level 0.05.

The NaN in Table 6 means “Not a Number” returned by the significant Wilcoxon Rank-sum test. The symbols $\wedge$, $\lor$ and $\approx$ respectively indicate that the proposed EAOA algorithm is significantly better, significantly inferior, and statistically similar to that of other compared state-of-the-art optimizers. According to Table 6, EAOA algorithm is significantly better than AOA and GPC for seventeen functions.

| DATA SET     | Instances | Features | positive | Negative |
|--------------|-----------|----------|----------|----------|
| yeast-2 vs 4 | 514       | 8        | 51       | 463      |
| wisconsin    | 683       | 9        | 239      | 444      |
| Vowel(2-others) | 528     | 10       | 48       | 480      |
| new-thyroid1 | 215       | 5        | 35       | 180      |
| Libra(123-others) | 360   | 90       | 72       | 288      |
| Libra(456-others) | 360   | 90       | 72       | 288      |
| breastEW     | 569       | 30       | 212      | 357      |
| Tic-tac-toe  | 958       | 9        | 332      | 626      |
| Wdbc         | 569       | 31       | 212      | 357      |
| Ionsphere    | 351       | 34       | 126      | 225      |
| Bupa         | 345       | 6        | 145      | 200      |
| glass0       | 214       | 9        | 70       | 144      |
| Heart        | 270       | 13       | 120      | 150      |
| Parkinsons   | 195       | 22       | 28       | 147      |
| Sonar        | 208       | 60       | 97       | 111      |
| Breast       | 277       | 9        | 81       | 196      |
### TABLE 8. Performances of EAOA, AOA, ALO, GWO, AND WOA on real-world data sets.

| DATA SET       | Measure | Or   | EAOA  | AOA   | ALO   | GWO   | WOA   |
|----------------|---------|------|-------|-------|-------|-------|-------|
| yeast-2_vs_4   | Accuracy| 92.97| 92.97 | 89.84 | 89.84 | 90.63 | 89.84 |
|                | G-Mean  | 55.47| 55.47 | 0     | 0     | 27.74 | 0     |
|                | F-Measure | 47.06| 47.06 | NaN   | NaN   | 14.29 | NaN   |
|                | Runtime | 0.4543| 0.0073| 0.0075| 0.0083| 0.0088| 0.0084|
|                | # Features | 8 | 6 | 2 | 2 | 2 | 2 |
| wisconsin      | Accuracy | 94.12| 96.47 | 95.88 | 95.88 | 94.71 | 93.53 |
|                | G-Mean  | 93.49| 96.5  | 95.65 | 95.65 | 94.35 | 92.17 |
|                | F-Measure | 91.53| 95   | 94.12 | 94.12 | 92.44 | 90.43 |
|                | Runtime | 0.45 | 0.0082| 0.0102| 0.0109| 0.0134| 0.0114|
|                | # Features | 9 | 8 | 6 | 6 | 6 | 7 |
| Vowel(2-others)| Accuracy | 88.64| 90.15 | 90.91 | 88.64 | 90.91 | 90.91 |
|                | G-Mean  | 62.36| 49.16 | 0     | 0     | 62.36 | 0     |
|                | F-Measure | 40 | 31.58 | NaN   | NaN   | 40   | NaN   |
|                | Runtime | 0.4740| 0.0083| 0.012 | 0.0081| 0.0083| 0.0093|
|                | # Features | 10 | 8 | 4 | 10 | 2 | 2 |
| new-thyroid1   | Accuracy | 98.11| 100  | 96.23 | 100  | 98.11 | 98.11 |
|                | G-Mean  | 93.54| 100  | 86.6  | 100  | 93.54 | 93.54 |
|                | F-Measure | 93.33| 100  | 85.71 | 100  | 93.33 | 93.33 |
|                | Runtime | 0.4653| 0.0069| 0.009 | 0.0078| 0.0296| 0.007 |
|                | # Features | 5 | 3 | 2 | 3 | 3 | 3 |
| Libra(123-others)| Accuracy | 88.89| 90 | 88.89 | 90 | 88.89 | 90 |
|                | G-Mean  | 66.67| 70.71 | 66.67 | 70.71 | 66.67 | 70.71 |
|                | F-Measure | 61.54| 66.67 | 61.54 | 66.67 | 61.54 | 66.67 |
|                | Runtime | 0.4525| 0.0088| 0.011 | 0.0106| 0.0208| 0.0126|
|                | # Features | 99 | 63 | 68 | 45 | 81 | 59 |
| Libra(456-others)| Accuracy | 94.44| 95.56 | 94.44 | 94.44 | 94.44 | 93.33 |
|                | G-Mean  | 90.01| 90.65 | 87.58 | 90.01 | 90.01 | 86.96 |
|                | F-Measure | 85.71| 88.24 | 84.85 | 85.71 | 85.71 | 82.35 |
|                | Runtime | 0.4695| 0.0086| 0.0089| 0.0085| 0.0142| 0.0088|
|                | # Features | 90 | 48 | 44 | 54 | 51 | 47 |
| breastEW       | Accuracy | 94.37| 95.07 | 93.66 | 92.96 | 92.96 | 93.66 |
|                | G-Mean  | 95.08| 95.3  | 94.5  | 93.91 | 93.91 | 94.5  |
|                | F-Measure | 92.86| 93.58 | 92.04 | 91.23 | 91.23 | 92.04 |
|                | Runtime | 0.4659| 0.0105| 0.022 | 0.0116| 0.0116| 0.0126|
|                | # Features | 39 | 16 | 23 | 29 | 29 | 29 |
| Tic-tac-toe    | Accuracy | 98.33| 73.64 | 70.29 | 65.27 | 65.27 | 65.27 |
|                | G-Mean  | 97.56| 62.39 | 57.98 | 0     | 0     | 0     |
|                | F-Measure | 67.53| 53.33 | 47.41 | NaN   | NaN   | NaN   |
|                | Runtime | 0.4579| 0.0157| 0.0107| 0.011 | 0.0133| 0.0083|
|                | # Features | 9 | 8 | 7 | 2 | 2 | 2 |
| Webc           | Accuracy | 94.37| 95.07 | 94.37 | 94.37 | 95.07 | 93.66 |
|                | G-Mean  | 95.08| 95.3  | 94.73 | 95.08 | 95.66 | 94.5  |
|                | F-Measure | 92.86| 93.58 | 92.73 | 92.86 | 93.69 | 92.04 |
|                | Runtime | 0.8307| 0.0089| 0.0095| 0.0123| 0.0152| 0.0105|
|                | # Features | 39 | 14 | 15 | 29 | 28 | 27 |
| Irisphere      | Accuracy | 90.8 | 90.8 | 89.66 | 87.76 | 82.76 | 88.51 |
|                | G-Mean  | 86.14| 86.14 | 84.24 | 80.32 | 74.83 | 83.49 |
|                | F-Measure | 85.19| 85.19 | 83.02 | 78.43 | 70.59 | 81.48 |
|                | Runtime | 0.8301| 0.0087| 0.0097| 0.0104| 0.0158| 0.011 |
|                | # Features | 34 | 18 | 17 | 21 | 16 | 22 |
significantly inferior for three and two functions, respectively, and AOA is statistically similar for three and four functions, respectively. EAOA algorithm is also significantly better than GWO and WOA for twelve functions, significantly inferior for nine and six functions, respectively, and statistically similar for two and five functions, respectively. Moreover, EAOA algorithm is significantly better than ALO for ten functions, significantly inferior for nine functions and statistically similar for four functions. While EAOA algorithm is significantly better than PSO for nine functions, significantly inferior for twelve functions and statistically similar for two functions. The preceding results prove that the EAOA algorithm performs statistically significant and comparatively better performance over the other state-of-the-art optimizers.

**B. RESULTS OF EAOA ON REAL-WORLD DATA SET**

In this subsection, experiments were conducted on 16 real-world data sets with different numbers of instances and attributes. These data were chosen from the UCI machine learning repository [24]. Their classification results were obtained using some state-of-the-art methods compared with the proposed method results to assess its classification performance. Table 7 shows the characteristics of the real-world data sets.

The instances of each data set were divided randomly in all experiments into three sets: train, test, and validation sets. The proposed EAOA algorithm was applied to perform a feature selection of each set where the SVM classifier was used to evaluate the selected features. 30 runs of the algorithm were independently performed to measure the performance based on accuracy, G-Mean, F-Measure and Runtime metrics for each data set then, the average of each metric was recorded. Also, the number of selected features (# features) for each method was recorded. Figure 4 shows the number of features obtained by compared methods EAOA, AOA, ALO, GWO, and WOA for each data set. Table 8 presents the performances of the compared feature selection methods on real-world data sets. The original data set (Or) results with full features are provided in the first column for references.

| Data Set | Accuracy | G-Mean | F-Measure | Runtime | # Features |
|----------|----------|--------|-----------|---------|------------|
| Bupa     | 63.95    | 58.88  | 50.79     | 1.82    | 6          |
|          | 63.95    | 57.74  | 49.18     | 1.793   | 5          |
|          | 58.14    | 0      | NaN       | 2005    | 2          |
|          | 58.14    | 0      | NaN       | 3879    | 2          |
|          | 58.14    | 0      | NaN       | 1816    | 1          |
|          | 58.14    | 0      | NaN       | 3829    | 1          |
| glass0   | 67.92    | 69.81  | 67.92     | 0.0614  | 9          |
|          | 69.81    | 67.92  | 67.92     | 0.0064  | 7          |
|          | 67.92    | 69.81  | 67.92     | 0.0065  | 8          |
|          | 67.92    | 69.81  | 67.92     | 0.0101  | 8          |
|          | 67.92    | 69.81  | 67.92     | 0.0082  | 8          |
| Heart    | 80.6     | 83.58  | 82.09     | 79.01   | 13         |
|          | 80.6     | 83.58  | 82.09     | 79.01   | 10         |
|          | 80.6     | 83.58  | 82.09     | 79.01   | 8          |
|          | 80.6     | 83.58  | 82.09     | 79.01   | 12         |
|          | 80.6     | 83.58  | 82.09     | 79.01   | 9          |
| Parkinsons | 85.42   | 69.72  | 76.19     | 1.1248  | 22         |
|          | 85.42    | 69.72  | 76.19     | 0.6129  | 11         |
|          | 85.42    | 69.72  | 76.19     | 0.1014  | 17         |
|          | 85.42    | 69.72  | 76.19     | 0.3728  | 14         |
|          | 85.42    | 69.72  | 76.19     | 0.4813  | 10         |
|          | 85.42    | 69.72  | 76.19     | 0.2765  | 10         |
| Sonar    | 78.85    | 80.77  | 78.85     | 0.8515  | 60         |
|          | 78.85    | 80.77  | 78.85     | 0.0073  | 33         |
|          | 78.85    | 80.77  | 78.85     | 0.0093  | 50         |
|          | 78.85    | 80.77  | 78.85     | 0.0081  | 58         |
|          | 78.85    | 80.77  | 78.85     | 0.0117  | 33         |
|          | 78.85    | 80.77  | 78.85     | 0.0108  | 39         |
| Breast   | 73.91    | 76.81  | 73.91     | 0.8296  | 9          |
|          | 73.91    | 76.81  | 73.91     | 0.0068  | 5          |
|          | 73.91    | 76.81  | 73.91     | 0.0094  | 5          |
|          | 73.91    | 76.81  | 73.91     | 0.0076  | 4          |
|          | 73.91    | 76.81  | 73.91     | 0.0127  | 3          |
|          | 73.91    | 76.81  | 73.91     | 0.0075  | 5          |

**TABLE 8. (Continued.) Performances of EAOA, AOA, ALO, GWO, AND WOA on real-world data sets.**
Results in Table 8 show obviously that the EAOA outperforms the compared methods in terms of the classification accuracy on all data sets except “Vowel(2-others)” data, where the AOA method gained the best accuracy for this data only. On the G-Mean and F-Measure metrics, our proposed feature selection method also reaped the best values for fourteen out of the sixteen used data sets, proving the stability of our proposal against other feature selection methods. Also, results in Table 8 clearly indicate that all the feature selection methods have reduced the time consumed for classification.
and it can be noticed that the runtime of the proposed EAOA is less than the AOA algorithm for almost all the data sets and has superiority for saving time over the other compared algorithms. It can also be noticed from Table 8 that although decreasing the number of features as in ‘yeast-2_vs_4’ data, EAOA can maintain the same performance obtained by the full features while for most other data sets, EAOA gained better performance than the performance obtained by the full features. From the preceding results, we can conclude that the EAOA feature selection method can gain better performance regarding the computational complexity (Runtime and space) than other methods since it had the less Runtime and the fewer features that guarantee the best performance.

To perform the comparison, the Wilcoxon signed-rank test (a paired difference, two-sided signed-rank test), which is a non-parametric statistical hypothesis [25] was used to apply a statistical significance analysis and derive strong conclusions fairly. For each data set, EAOA was compared with all the other methods; the differences were calculated and ranked from smallest ‘1’ to largest ‘16’ for each two compared methods. (‘+’ or ‘−’) signs were subsequently assigned to the corresponding differences of the ranks. R+ and R− were assigned to all the +ve and −ve ranks after summing up separately. The T value was compared against a significance level $\alpha = 0.05$, with a critical value, equals 29 for 16 data sets where $T = \min \{R+, R−\}$. The null hypothesis was that all performance differences between any two compared methods might occur by chance, and the null hypothesis was rejected only if the T value is $\leq$ the critical value 29.

Tables 9-14 present the significance test results and the addition ‘+’ symbol in tables 12-14 indicates that our proposal EAOA outperforms the compared methods. The NaN values were represented in these tables by the value −1 to be at the bottom of the ranked list and thus are excluded.

The significance test results of average accuracy for EAOA VS. AOA, ALO, GWO, and WOA, using the SVM classifier, are presented in Table 9. In the case of EAOA vs. AOA, EAOA has a positive difference for 15 data sets; hence it is better than AOA, while AOA has a negative difference and is better than EAOA for only one data set. The sum of all the positive ranks $R+$ is 133, and the sum of negative ranks $R−$ is 3. Since the data used are 16 data sets from the critical table, to reject a null hypothesis at 0.05 (the significance level), the T value should be $\leq 29$. It can be concluded that EAOA statistically outperforms AOA as $T = \min \{R+, R−\} = \min \{133, 3\} = 3 < 29$. As well, in the case of EAOA vs. ALO, GWO, and WOA, EAOA

| DATA SET       | EAOA | AOA | Difference | Rank | ALO | Difference | Rank | GWO | Difference | Rank | WOA | Difference | Rank |
|----------------|------|-----|------------|------|-----|------------|------|-----|------------|------|-----|------------|------|
| yeast-2_vs_4   | 47.06| -1  | 48.06      | 15   | -1  | 48.06      | 14   | 14.29| 32.77      | 13   | -1  | 48.06      | 14   |
| wisconsin      | 95   | 94.12| 0.88       | 2    | 94.12| 0.88       | 5    | 92.44| 2.56       | 4    | 90.43| 4.57       | 7    |
| Vowel(2-others)| 31.58| -1  | 32.58      | 14   | 40  | -8.42      | -11  | -1  | 32.58      | 12   | -1  | 32.58      | 13   |
| new-thyroid1   | 100  | 85.71| 14.29      | 13   | 100 | 0          | 1    | 93.33| 6.67       | 8    | 93.33| 6.67       | 10   |
| Libra(23-others)| 66.67| 61.54| 5.13       | 10   | 66.67| 0          | 1    | 61.54| 5.13       | 7    | 66.67| 0          | 1    |
| Libra(456-others)| 88.24| 84.85| 3.39       | 7    | 85.71| 2.53       | 7    | 85.71| 2.53       | 3    | 82.35| 5.89       | 8    |
| breastEW       | 93.58| 92.04| 1.54       | 4    | 91.23| 2.35       | 6    | 91.23| 2.35       | 2    | 92.04| 1.54       | 3    |
| Tic-tac-toe    | 53.33| 47.41| 5.92       | 11   | -1  | 54.33      | 16   | -1  | 54.33      | 16   | -1  | 54.33      | 16   |
| Wdbc           | 93.58| 92.73| 0.85       | 1    | 92.86| 0.72       | 4    | 93.69| -0.11      | -1   | 92.04| 1.54       | 3    |
| ionsphere      | 85.19| 83.02| 2.17       | 6    | 78.43| 6.76       | 10   | 70.59| 14.6       | 11   | 81.48| 3.71       | 6    |
| Bupa           | 49.18| -1  | 50.18      | 16   | -1  | 50.18      | 15   | -1  | 50.18      | 14   | -1  | 50.18      | 15   |
| glasso         | 50   | 45.16| 4.84       | 9    | 50  | 0          | 1    | -1  | 51         | 15   | 50  | 0          | 1    |
| Heart          | 80   | 78.57| 1.43       | 3    | 76.36| 3.64       | 8    | 77.19| 2.81       | 5    | 74.07| 5.93       | 9    |
| Parkinsons     | 76.19| 66.67| 9.52       | 12   | 63.16| 13.03      | 13   | 63.16| 13.03      | 10   | 66.67| 9.52       | 11   |
| Sonar          | 76.19| 74.42| 1.77       | 5    | 70  | 6.19       | 9    | 71.43| 4.76       | 6    | 78.26| -2.07      | 5    |
| Breast         | 52.94| 48.48| 4.46       | 8    | 42.42| 10.52      | 12   | 42.42| 10.52      | 9    | 42.42| 10.52      | 12   |

| EAOA vs AOA | AOA | ALO | GWO | WOA |
|-------------|-----|-----|-----|-----|
| R+          | 133 | 133 | 133 | 128 |
| R−          | 3   | 0   | 2   | 4   |
| T = \min \{R+, R−\} | 8 (+) | 0 (+) | 2 (+) | 4 (+) |

$T = \min (107, 0) = 0$  $T = \min (77, 11) = 11$  $T = \min (78, 1) = 1$  $T = \min (71, 5) = 5$
TABLE 13. The T values of significance test on averaged G-Mean for EAOA vs. AOA, ALO, GWO and WOA.

|        | EAOA vs | AOA | ALO | GWO | WOA |
|--------|---------|-----|-----|-----|-----|
| R+     | 136     | 120 | 135 | 129 |
| R−     | 0       | 13  | 1   | 5   |
| T = min {R+, R−} | 0 (−) | 13 (+) | 1 (+) | 5 (+) |

TABLE 14. The T values of significance test on averaged F-Measure for EAOA vs. AOA, ALO, GWO and WOA.

|        | EAOA vs | AOA | ALO | GWO | WOA |
|--------|---------|-----|-----|-----|-----|
| R+     | 107     | 77  | 78  | 71  |
| R−     | 0       | 11  | 1   | 5   |
| T = min {R+, R−} | 0 (−) | 11 (+) | 1 (+) | 5 (+) |

FIGURE 4. Number of obtained features by compared methods for each data set.

All the results of the statistical Wilcoxon signed-rank test prove that the feature selection proposed method can statistically outperform the AOA algorithm and has strong competitiveness against other comparative methods in terms of all used performance measures.

Sensitivity, specificity, precision, and AUC measures were also recorded and presented in Table 15 to show the EAOA feature selection method’s overall performance compared with the basic AOA and the three state-of-the-art feature selection techniques ALO GWO and WOA. It can be noticed from Table 15 that EAOA outperforms other feature selection methods for all the used data in terms of sensitivity, specificity, and AUC except for just four data which proves the superiority of our proposal in identifying both positive and negative examples correctly comparing to other used methods. Furthermore, it shows the best performance in terms of precision than basic AOA, ALO, GWO, and WOA for all data sets except only one data. The preced-
| DATA SET       | Measure | Or  | EAOA | AOA | ALO | GWO | WOA |
|---------------|---------|-----|------|-----|-----|-----|-----|
| yeast-2_vs_4  | Sensitivity | 30.77 | 30.77 | 0   | 0   | 7.69 | 0   |
|               | Specificity | 100  | 100  | 100  | 100  | 100  | 100  |
|               | Precision  | 100  | 100  | NaN  | NaN  | 100  | NaN  |
|               | AUC       | 4    | 4    | 0    | 0    | 1    | 0    |
| wisconsin     | Sensitivity | 91.53 | 96.61 | 94.92 | 94.92 | 93.22 | 88.14 |
|               | Specificity | 95.5  | 96.4  | 96.4  | 95.5  | 96.4  | 96.4  |
|               | Precision  | 91.53 | 93.44 | 93.33 | 93.33 | 91.67 | 92.86 |
|               | AUC       | 54   | 57   | 56   | 56   | 55   | 52   |
| Vowel(2-others) | Sensitivity | 41.67 | 25   | 0    | 41.67 | 0    | 0    |
|               | Specificity | 93.33 | 96.67 | 100  | 93.33 | 100  | 100  |
|               | Precision  | 38.46 | 42.86 | NaN  | 38.46 | NaN  | NaN  |
|               | AUC       | 5    | 3    | 0    | 5    | 0    | 0    |
| new-thyroidl  | Sensitivity | 87.5  | 100  | 75   | 100  | 87.5  | 87.5  |
|               | Specificity | 100   | 100  | 100  | 100  | 100   | 100   |
|               | Precision  | 100   | 100  | 100  | 100  | 100   | 100   |
|               | AUC       | 7    | 8    | 6    | 8    | 7    | 7    |
| Libra(123-others) | Sensitivity | 44.44 | 50   | 44.44 | 50   | 44.44 | 50   |
|               | Specificity | 100   | 100  | 100  | 100  | 100   | 100   |
|               | Precision  | 100   | 100  | 100  | 100  | 100   | 100   |
|               | AUC       | 8    | 9    | 8    | 9    | 8    | 9    |
| Libra(456-others) | Sensitivity | 83.33 | 83.33 | 77.78 | 83.33 | 83.33 | 77.78 |
|               | Specificity | 97.22 | 98.61 | 98.61 | 97.22 | 97.22 | 97.22 |
|               | Precision  | 88.24 | 93.75 | 93.33 | 88.24 | 88.24 | 87.5  |
|               | AUC       | 15   | 15   | 14   | 15   | 15   | 14   |
| breastEW      | Sensitivity | 98.11 | 96.23 | 98.11 | 98.11 | 98.11 | 98.11 |
|               | Specificity | 92.13 | 94.38 | 91.01 | 89.89 | 89.89 | 91.01 |
|               | Precision  | 88.14 | 91.07 | 86.67 | 85.25 | 85.25 | 86.67 |
|               | AUC       | 52   | 51   | 52   | 52   | 52   | 52   |
| Tic-tac-toe   | Sensitivity | 95.18 | 43.37 | 38.55 | 0    | 0    | 0    |
|               | Specificity | 100   | 89.74 | 87.18 | 100  | 100  | 100   |
|               | Precision  | 100   | 69.23 | 61.54 | NaN  | NaN  | NaN   |
|               | AUC       | 79   | 36   | 32   | 0    | 0    | 0    |
| Wdbc          | Sensitivity | 98.11 | 96.23 | 96.23 | 98.11 | 98.11 | 98.11 |
|               | Specificity | 92.13 | 94.38 | 93.26 | 92.13 | 93.26 | 91.01 |
|               | Precision  | 88.14 | 91.07 | 89.47 | 88.14 | 89.66 | 86.67 |
|               | AUC       | 52   | 51   | 51   | 52   | 52   | 52   |
| ionosphere    | Sensitivity | 74.19 | 74.19 | 70.97 | 64.52 | 58.06 | 70.97 |
|               | Specificity | 100   | 100  | 100  | 100  | 96.43 | 98.21 |
|               | Precision  | 100   | 100  | 100  | 100  | 90   | 95.65 |
|               | AUC       | 23   | 23   | 22   | 20   | 18   | 22   |
TABLE 15. (Continued.) Sensitivity, specificity, precision and AUC of EAOA, AOA, ALO, GWO, and WOA methods on real-world data sets.

| Algorithm   | Sensitivity | Specificity | Precision | AUC  |
|-------------|-------------|-------------|-----------|------|
| Bupa        | 44.44       | 41.67       | 0         | 0    | 0    | 0    |
| Specificity | 78          | 80          | 100       | 100  | 100  | 100  |
| Precision   | 59.26       | 60          | NaN       | NaN  | NaN  | NaN  |
| AUC         | 16          | 15          | 0         | 0    | 0    | 0    |
| glass0      | 47.06       | 47.06       | 41.18     | 47.06| 0    | 47.06|
| Specificity | 77.78       | 80.56       | 80.56     | 80.56| 100  | 80.56|
| Precision   | 50          | 53.33       | 50        | 53.33| NaN  | 53.33|
| AUC         | 8           | 8           | 7         | 8    | 0    | 8    |
| Heart       | 73.33       | 73.33       | 73.33     | 70   | 73.33| 66.67|
| Specificity | 86.49       | 91.89       | 89.19     | 89.19| 86.49| 89.19|
| Precision   | 81.48       | 88          | 84.62     | 84   | 81.48| 83.33|
| AUC         | 22          | 22          | 22        | 21   | 22   | 20   |
| Parkinsons  | 50          | 66.67       | 58.33     | 50   | 50   | 50   |
| Specificity | 97.22       | 97.22       | 94.44     | 97.22| 97.22| 100  |
| Precision   | 85.71       | 88.89       | 77.78     | 85.71| 85.71| 100  |
| AUC         | 6           | 8           | 7         | 6    | 6    | 6    |
| Sonar       | 60          | 64          | 64        | 56   | 60   | 72   |
| Specificity | 96.3        | 96.3        | 92.59     | 96.3 | 92.59| 88.89|
| Precision   | 93.75       | 94.12       | 88.89     | 93.33| 88.24| 85.71|
| AUC         | 15          | 16          | 16        | 14   | 15   | 18   |
| Breast      | 25          | 45          | 40        | 35   | 35   | 35   |
| Specificity | 93.88       | 89.8        | 89.8      | 87.76| 87.76| 87.76|
| Precision   | 62.5        | 64.29       | 61.54     | 53.85| 53.85| 53.85|
| AUC         | 5           | 9           | 8         | 7    | 7    | 7    |

FIGURE 5. Box and whiskers plot for sensitivity of compared methods.

FIGURE 6. Box and whiskers plot for specificity of compared methods.

The results can also be noticed from Figures 5-8 which present the box and whiskers plots of sensitivity, specificity, precision, and AUC for all compared methods. These figures give an evident and qualitative specification of the performance of the compared methods, and the box plots of all other feature selection methods are visibly lower than EAOA in all mentioned measures, which more directly demonstrate the superiority of the EAOA feature selection algorithm.
This paper has enhanced the AOA algorithm and applied it in the feature selection problem. The EAOA algorithm is proposed to enhance the basic AOA performance by adding a new parameter that depends on the step length of each individual while revising the individual location over the course of iterations. The new parameter, upper and lower limits of search space, are employed to enhance the balance between the exploration and exploitation processes. The EAOA algorithm’s ability to perform the optimization compared to state-of-the-art optimizers is tested using twenty-three benchmark functions. Experimental results show that the EAOA algorithm as compared to other algorithms has better exploitation capability for unimodal functions and better exploration capability for multimodal and fixed-dimensional multimodal functions. Also, the proposed EAOA algorithm could escape the local optima and provide a better convergence rate than the basic AOA method and other state-of-the-art optimizers. After applying the significant Wilcoxon Rank-sum test, the results proved that the EAOA algorithm performs statistically significant and comparatively better performance over the other state-of-the-art optimizers.

In addition, as a potential application, EAOA is used to find the optimal set of features, and SVM is applied for classification tasks. Sixteen real-world data sets are taken into account. Each data set is divided into three sets: train, test, and validation sets. To measure the performance on each data set the average accuracy, G-Mean, F-Measure, Runtime, and the number of selected features for each method were recorded. EAOA records comparable or superior results in the case of feature selection and classification performance when compared with the basic AOA, ALO, GWO, and WOA.

Results indicate that all the feature selection methods have reduced the time consumed for classification and the proposed EAOA is less than the AOA algorithm for almost all the data sets and has superiority for saving computational time and space while maintaining the best performance over the other compared algorithms. The statistical Wilcoxon signed-rank test results also prove that the feature selection proposed method can statistically outperform the AOA algorithm and has strong competitiveness against other comparative methods in terms of accuracy, G-Mean, and F-Measure. Finally, the box and whiskers plots of sensitivity, specificity, precision, and AUC for all compared methods show the superiority of the EAOA feature selection algorithm.

**REFERENCES**

[1] B. Xue, M. Zhang, and W. N. Browne, “Particle swarm optimization for feature selection in classification: A multi-objective approach,” *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1656–1671, Dec. 2013.

[2] S. Tiwari, B. Singh, and M. Kaur, “An approach for feature selection using local searching and global optimization techniques,” *Neural Comput. Appl.*, vol. 28, no. 10, pp. 2915–2930, Oct. 2017.

[3] Z. Beheshti and S. M. H. Shamsuddin, “A review of population-based meta-heuristic algorithms,” *Int. J. Adv. Soft Comput. Appl.*, vol. 5, no. 1, pp. 1–35, 2013.

[4] F. A. Hashim, E. H. Houssein, M. S. Mabrouk, W. Al-Atabany, and S. Mirjalili, “Henry gas solubility optimization: A novel physics-based algorithm,” *Future Gener. Comput. Syst.*, vol. 101, pp. 646–667, Dec. 2019.

[5] A. Kaveh and S. Talatahari, “A novel heuristic optimization method: Charged system search,” *Acta Mechanica*, vol. 213, nos. 3–4, pp. 267–289, 2010.

[6] A. Kaveh and M. Khayatazad, “A new meta-heuristic method: Ray optimization,” *Comput. Struct.*, vol. 112, pp. 283–294, Dec. 2012.

[7] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, “GSA: A gravitational search algorithm,” *J. Inf. Sci.*, vol. 179, no. 13, pp. 2232–2248, 2009.

[8] F. A. Hashim, K. Hussain, E. H. Houssein, M. S. Mabrouk, and W. Al-Atabany, “Archimedes optimization algorithm: A new metaheuristic algorithm for solving optimization problems,” *Appl. Intell.*, vol. 51, pp. 1531–1551, Sep. 2020.

[9] Y. Liu, G. Wang, H. Chen, H. Dong, X. Zha, and S. Wang, “An improved particle swarm optimization for feature selection,” *J. Bionic Eng.*, vol. 8, no. 2, pp. 191–200, Jun. 2011.

[10] Q. Al-Tashi, S. J. A. Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, “Binary optimization using hybrid grey wolf optimization for feature selection,” *IEEE Access*, vol. 7, pp. 39496–39508, 2019.

[11] N. Neggaz, E. H. Houssein, and K. Hussain, “An efficient henry gas solubility optimization for feature selection,” *Expert Syst. Appl.*, vol. 152, Aug. 2020, Art. no. 113364.

[12] P. Moradi and M. Rostami, “Integration of graph clustering with ant colony optimization for feature selection,” *Knowl.-Based Syst.*, vol. 84, pp. 144–161, Aug. 2015.

[13] R. A. Ibrahim, A. A. Ewees, D. Oliva, M. A. Elaziz, and S. Lu, “Improved salp swarm algorithm based on particle swarm optimization for feature selection,” *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 8, pp. 3155–3169, Aug. 2019.

[14] J. Too, A. R. Abdullah, and N. M. Saad, “A new quadratic binary Harris hawk optimization for feature selection,” *Electronics*, vol. 8, no. 10, p. 1130, Oct. 2019.

[15] S. Mirjalili, “The ant lion optimizer,” *Adv. Eng. Softw.*, vol. 83, pp. 80–98, May 2015.
[16] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer,” Adv. Eng. Software, vol. 69, pp. 46–61, Mar. 2014.
[17] S. Mirjalili and A. Lewis, “The whale optimization algorithm,” Adv. Eng. Software, vol. 95, pp. 51–67, May 2016.
[18] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in Proc. Int. Conf. Neural Netw. (ICNN), vol. 4, Nov./Dec. 1995, pp. 1942–1948.
[19] S. Harifi, J. Mohammadzadeh, M. Khalilian, and S. Ebrahimeinjad, “Giza pyramids construction: An ancient-inspired metaheuristic algorithm for optimization,” Evol. Intell., vol. 13, pp. 1–19, Jul. 2020.
[20] S. Hussain and G. C. Hazarika, “Educational data mining model using rattle,” Int. J. Adv. Comput. Sci. Appl., vol. 5, no. 6, pp. 22–27, 2014.
[21] M. Gong, “A novel performance measure for machine learning classification,” Int. J. Manag. Inf. Technol., vol. 13, no. 1, pp. 1–19, 2021.
[22] A. S. Desuky and L. M. El Bakrawy, “Improved prediction of post-operative life expectancy after thoracic surgery,” Adv. Syst. Sci. Appl., vol. 16, no. 2, pp. 70–80, 2016.
[23] F. Wilcoxon, “Individual comparisons by ranking methods,” Biometrics Bull., vol. 1, no. 6, pp. 80–83, 1945.
[24] A. Asuncion and D. J. Newman, “UCI machine learning repository,” School Inf. Comput. Sci., Univ. California, Irvine, CA, USA, 2007, [Online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.html
[25] R. F. Woolson, “Wilcoxon signed-rank test,” in Wiley Encyclopedia of Clinical Trials. Hoboken, NJ, USA: Wiley, 2008, pp. 1–3.

ABEER S. DESUKY received the B.Sc. degree in science, in 2003, and the M.Sc. and Ph.D. degrees in computer science, in 2008 and 2012, respectively. She is currently an Associate Professor in computer science with the Mathematics Department, Faculty of Science, Al-Azhar University, Cairo, Egypt. She has published several research papers in the field of AI, machine learning, metaheuristic optimization, and data mining and analysis. Supervisor of some master’s and Ph.D. theses. She is a reviewer in many Scopus-indexed journals, such as IEEE Access, Egyptian Informatics Journal, and Advances in Systems Science and Applications (ASSA).

SADIQ HUSSAIN received the Ph.D. degree from Dibrugarh University, India. He is currently a System Administrator with Dibrugarh University, Assam, India. He is associated with the computerization of the examination system of Dibrugarh University. He has published various research and conference papers of international repute. He had reviewed research papers for many IEEE, SCI, and Scopus-indexed journals. His research interests include data mining, medical analytics, and machine learning. He acted as a technical member of many reputed conferences.

SAMINA KAUSAR is currently working as a Researcher and also a Ph.D. Scholar with the School of Computer Engineering and Science, Shanghai University, China. She is also working as an Assistant Professor with the University of Kotli, Azad Kashmir, Pakistan. Her research interests include big data, bioinformatics, computer networks, cloud computing, data mining, E-learning, and machine learning algorithms.

MD. AKHTARUL ISLAM received the B.Sc. and M.S. degrees in statistics biostatistics and informatics from Dhaka University, Dhaka, Bangladesh, in 2012 and 2013, respectively. He is currently working as an Assistant Professor in the statistics discipline with Khulna University, Khulna, Bangladesh. He has authored or co-authored around 12 publications in different peer-reviewed journals. His research interests include biostatistics, epidemiology, public health, infectious disease, and meta-analysis.

LAMIAA M. EL BAKRAWY received the M.S. and Ph.D. degrees in computer science from the Faculty of Science, Al-Azhar University, Egypt, in 2009 and 2018, respectively. Her master’s topic was optimization using swarm intelligence. Her Ph.D. topic was machine learning in image authentication. She is currently an Associate Professor with the Mathematics and Computer Department, Faculty of Science, Al-Azhar University. She has published many articles in well-known international journals. Her research interests include image processing, social network analysis, computational intelligence, machine learning, metaheuristic optimization, and information security. She is also a program committee member in various international conferences.