**Tech Science Press**

check for updates

# Novel Optimized Feature Selection Using Metaheuristics Applied to Physical Benchmark Datasets

**Doaa Sami Khafaga[1], El-Sayed M. El-kenawy[2], Fadwa Alrowais[1,\*], Sunil Kumar[3], Abdelhameed Ibrahim[4] and Abdelaziz A. Abdelhamid[5,6]**

[1]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia
[2]Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, 35111, Egypt
[3]School of Computer Science, University of Petroleum and Energy Studies, Dehradun, 248001, India
[4]Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura, 35516, Egypt
[5]Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, 11566, Egypt
[6]Department of Computer Science, College of Computing and Information Technology, Shaqra University, 11961, Saudi Arabia
*Corresponding Author: Fadwa Alrowais. Email: fmalrowais@pnu.edu.sa
Received: 05 June 2022; Accepted: 12 July 2022

**Abstract:** In data mining and machine learning, feature selection is a critical part of the process of selecting the optimal subset of features based on the target data. There are $2^n$ potential feature subsets for every $n$ features in a dataset, making it difficult to pick the best set of features using standard approaches. Consequently, in this research, a new metaheuristics-based feature selection technique based on an adaptive squirrel search optimization algorithm (ASSOA) has been proposed. When using metaheuristics to pick features, it is common for the selection of features to vary across runs, which can lead to instability. Because of this, we used the adaptive squirrel search to balance exploration and exploitation duties more evenly in the optimization process. For the selection of the best subset of features, we recommend using the binary ASSOA search strategy we developed before. According to the suggested approach, the number of features picked is reduced while maximizing classification accuracy. A ten-feature dataset from the University of California, Irvine (UCI) repository was used to test the proposed method's performance *vs*. eleven other state-of-the-art approaches, including binary grey wolf optimization (bGWO), binary hybrid grey wolf and particle swarm optimization (bGWO-PSO), bPSO, binary stochastic fractal search (bSFS), binary whale optimization algorithm (bWOA), binary modified grey wolf optimization (bMGWO), binary multiverse optimization (bMVO), binary bowerbird optimization (bSBO), binary hybrid GWO and genetic algorithm

(bGWO-GA), binary firefly algorithm (bFA), and bGA methods. Experimental results confirm the superiority and effectiveness of the proposed algorithm for solving the problem of feature selection.

## 1 Introduction

Feature selection is the process of selecting the most significant variables in predictive modeling problems. Feature selection is considered a principal concept in the field of machine learning (ML). Some predictive modeling problems have high-dimensional datasets which may contain redundant or irrelevant variables. Employing all input variables in training the model increases the likelihood of including non-informative features. This may cause uncertainty in the predictions and reduce the model's effectiveness. Moreover, using such a large number of features in developing the model slows the training process and makes it computationally intensive. Therefore, in the process of feature selection, irrelevant and redundant attributes are discarded, and the number of features to be used in training the model is decreased. Using a reduced and significant set of features in training the model contributes in improving the prediction performance and decreases the computational complexity and time.

Feature selection techniques are categorized into supervised and unsupervised methods based on the dependence on the target variable in selecting relevant features. In supervised feature selection approaches, the relation between the input predictors and the target variable is used to distinguish the effective features. However, this is not the case in unsupervised methods. Supervised methods have been shown to provide better feature selection performance and hence are commonly used [1]. Filter, wrapper, embedded, and intrinsic methods are the most popular methods under the umbrella of supervised learning feature selection techniques [2]. In filter methods, statistical techniques are employed to evaluate the relevance between the input features and the target. Commonly, statistical correlation between the input predictors and the target variable is used as the basis for filter feature selection. A correlation score is computed for each feature, and based on these scores; the input predictors are filtered. Only features with the strongest relationship with the target are used for the model development.

In wrapper methods, multiple models are generated with different subsets of input predictors. The performance of these models is evaluated to find the best combination of features that optimizes the prediction performance. The main difference between the filter and wrapper methods is that the latter assesses the goodness of selected features based on the predictive model accuracy measure, which makes it more efficient but computationally intensive. On the other hand, the selection criteria of the filter approaches are model-independent, which makes it more popular, and has a less computational cost but is less efficient in selecting relevant features [3–5]. In order to get the advantages of both methods, the embedded approaches integrate the statistical-based and model-based feature selection into a single algorithm. Interestingly, some ML algorithms contain built-in feature selection properties such as decision trees, random forest, Lasso algorithms, and others. These algorithms automatically select the features that maximize the model performance as part of model learning. Such algorithms are referred to as intrinsic feature selection methods. It is worth mentioning that the performance of a feature selection approach is dependent on the problem in hand. In other words, the performance of a specific approach in selecting the significant features may vary from one problem to another.

The type of the input and output variables, the relevance, noise, and the number of features definitely affect feature selection performance. This makes it challenging to decide on the best feature selection approach to be used especially when there are a vast number of input attributes. In this case, the feature selection problem could be tackled as an optimization problem to select the optimal set of input features. Stochastic optimization algorithms can search and select optimum features. Among the powerful and modern optimization approaches is the metaheuristic optimization algorithms. Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Simulated Annealing, Ant Colony, and Bee algorithms are the most popular metaheuristic optimization algorithms. These optimizers have been used for feature selection in many studies to enhance the performance of classification algorithms [6].

## 2 Literature Review

Feature selection has been widely employed in a variety of applications, including spam email filtering, disease diagnosis, text classification, fraudulent detection, deoxyribonucleic acid (DNA) microarray analysis, and others. Extracting significant features from the input dataset is crucial for creating an efficient decision-making model. To increase the accuracy of prediction models, most studies apply filter and wrapper feature selection strategies [7]. For instance, authors in [8] developed a new feature selection strategy based on an improved Chi-square algorithm to improve the performance of Arabic text classification using support vector machines (SVM) classifier. According to authors in [9], an anomaly network-based system can achieve a high detection rate and minimize false positives by employing an innovative hybrid method. The AdaBoost technique was used to classify the data, and an artificial bee colony was employed to choose features. The authors of [10] proposed a feature selection method for mammography classification utilizing a Biogeography-based optimization approach employing the adaptive neuro-fuzzy inference system and artificial neural network. According to [11], robust feature selection based on the L1, and L2 norm may eliminate outliers and identify attributes across all data points with joint sparsity while reducing loss function and regularization. The scientists utilized regression as an aim and conducted multiple tests on six datasets to evaluate the proposed model. A novel unsupervised feature selection approach based on L1, and L2 norm regularization was developed by authors in [12] for human action recognition. The model extracts and chooses features simultaneously in order to create optimum features for the classification framework. Another work [13] developed a novel linear dimensionality reduction approach called robust discriminant regression (RDR) that leverages the L2; 1-norm as the core measure for feature extraction. RDR was unable to pick features in sparse projections, according to the authors.

Finding the best combination of features is a challenging task. In response to this demand, optimization algorithms have been recently incorporated into modern feature selection approaches to select the best subset of attributes from datasets. Metaheuristic optimization algorithms have been utilized by many recent studies to find the optimal set of input features. A metaheuristic is a problem-independent algorithmic framework that gives a set of recommendations for developing heuristic optimization algorithms at a high level. Optimization algorithms often use a combination of specific algorithmic parameters and regulating parameters in the evaluation process. These parameters highly impact the feature selection process and significantly affect the performance of the employed machine learning models. Many studies used various optimization algorithms, such as the Genetic Algorithm and Particle Swarm Optimization, and others for attribute selection to improve the effectiveness of machine learning models. For instance, authors in [14] established a method for feature selection using GA by ranking genes using T-Statistics, F-Test, and signal-to-noise ratio (SNR) measurements. The selected features were fed afterward to K-nearest neighbor (KNN) and SVM classifiers. The study in [15] used the wrapper technique to develop a novel fusion method that combines GA and

SVM to perform feature selection. When compared to other well-known methods, the hybrid GA-SVM produced greater classification precision. Another study [16] suggested a GA-based feature selection method that uses KNN classification error as a fitness to provide a set of features to improve the classifier performance. Moradi [17] combined a local search technique with a PSO-based feature selection method. They tested the system's performance using various filter and wrapper-based methods and found that classifiers provided superior accuracy using the optimized features. Authors in [18] introduced a new feature selection approach based on chicken swarm optimization. The results of that study showed that the Chicken swarm optimization outperforms the original GA and PSO optimization algorithms on standard datasets. A feature selection algorithm that is based on elephant search optimization and a deep learning network was proposed in [19] to investigate microarray data. Rodrigues et al. [20] proposed a wrapper feature selection strategy based on Bat Algorithm (BA) and Optimum-Path Forest. Authors in [21] employed the Binary Dragonfly optimization algorithm to select a subset of features from UCI datasets and achieved better results than GA and PSO algorithms. Sayed et al. [22] developed a new metaheuristic method for feature selection based on the crow search algorithm. Their results showed superior results with benchmark datasets. To solve local optima and poor convergence difficulties, authors in [23] proposed a hybrid feature selection method combining a swarm algorithm with chaos theory and applied it to a variety of benchmark tasks. To reduce the number of attributes in breast cancer datasets, authors in [24] employed the Modified Correlation Rough Set feature selection method. For identifying CT images of cervical cancer, a new feature selection approach based on Artificial Bee Colony and KNN algorithms was introduced in [25]. To increase classification performance, authors in [26] proposed a new feature selection method for Arabic text classification utilizing the firefly algorithm. The authors conducted experiments on the OSAC dataset and achieved a precision of 0.994 with the firefly algorithm. Authors in [6] developed a metaheuristics-based feature selection model to select optimal features for breast cancer classification. Another study [27] introduced a new prediction framework called IGWO-KELM by combining an enhanced grey wolf optimization (IGWO) and kernel extreme learning machine (KELM). In the suggested method, a genetic algorithm (GA) was utilized to generate a diverse set of initial positions, and then grey wolf optimization (GWO) was used to update the population's current positions in the discrete searching space, resulting in the best feature subset for KELM classification. The authors claimed that the IGWO-KELM outperformed the native GA and GWO for breast cancer classification. In another study [28], a wrapper feature selection algorithm is developed using the chimp optimization algorithm (ChOA) for the classification of high-dimensional biomedical data. Authors in [29] introduced an enhanced version of the Binary Dragonfly Algorithm (BDA) to select optimal feature classification problems. They proposed a new wrapper-based Hyper Learning Binary Dragonfly Algorithm (HLBDA) in which a hyper learning strategy is utilized to improve searching and help escape local optima. Authors in [30] developed a modified Binary Grey Wolf optimizer (MbGWO) to find optimal features using the right balance between exploration and exploitation. Their proposed optimizer is based on Stochastic Fractal Search (SFS) to find the best solution for the MbGWO. KNN classifier was used to assess the performance of the selected features over nineteen datasets. Their results reveal the superiority of the introduced MbGWO.

## 3 Methodology

The proposed binary adaptive squirrel search optimization algorithm is based on the flying squirrels' search mechanism [31]. To utilize the optimization algorithm in the task of feature selection, the optimization algorithm is applied, and the fitness function is evaluated for each feature; then, the resulting continuous solution is converted to binary. Therefore, this section starts with presenting the

main idea of the squirrel search algorithm, followed by presenting the fitness calculation and binary conversion.

### 3.1 The Squirrel Search Algorithm

The basic idea of the squirrel search algorithm takes into account that squirrels travel between three different types of trees: hickory, oak, and regular trees. Nuts come from oak and hickory trees, whereas other trees provide no nourishment. For simplicity, the squirrel search (SS) algorithm treats three oak trees and one hickory tree as $N_{fs}$ available food supplies for the $n$ flying squirrels (FS). The squirrel locations ($SL$) and velocities ($SV$) are represented by the following matrices:

$$SL = \begin{bmatrix} SL_{1,1} & SL_{1,2} & SL_{1,3} & \ldots & SL_{1,d} \\ SL_{2,1} & SL_{2,2} & SL_{2,3} & \ldots & SL_{2,d} \\ SL_{3,1} & SL_{3,2} & SL_{3,3} & \ldots & SL_{3,d} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ SL_{m,1} & SL_{m,2} & SL_{m,3} & \ldots & SL_{m,d} \end{bmatrix} \tag{1}$$

$$SV = \begin{bmatrix} SV_{1,1} & SV_{1,2} & SV_{1,3} & \ldots & SV_{1,d} \\ SV_{2,1} & SV_{2,2} & SV_{2,3} & \ldots & SV_{2,d} \\ SV_{3,1} & SV_{3,2} & SV_{3,3} & \ldots & SV_{3,d} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ SV_{m,1} & SV_{m,2} & SV_{m,3} & \ldots & SV_{m,d} \end{bmatrix} \tag{2}$$

For the $j^{th}$ dimension and $i^{th}$ squirrel, the flying squirrel is denoted by $FA_{i,j}$ for $i \in 1, 2, 3, \ldots, n$ and $j \in 1, 2, 3, \ldots, d$. The uniform distribution within upper and lower bounds is used to determine the initial locations of $FA_{i,j}$. For each flying squirrel, the fitness values $f = f_1, f_2, f_3, \ldots, f_n$ are measured in terms of the following matrix.

$$f = \begin{bmatrix} f_1 \left( SL_{1,1}, SL_{1,2}, SL_{1,3}, \ldots, SL_{1,d} \right) \\ f_2 \left( SL_{2,1}, SL_{2,2}, SL_{2,3}, \ldots, SL_{2,d} \right) \\ f_3 \left( SL_{3,1}, SL_{3,2}, SL_{3,3}, \ldots, SL_{3,d} \right) \\ \ldots \\ f_m \left( SL_{m,1}, SL_{m,2}, SL_{m,3}, \ldots, SL_{m,d} \right) \end{bmatrix} \tag{3}$$

As a result, flying squirrels look for food sources with a higher fitness value than those with lower fitness values. It is a hickory tree that provides the best value. The values are then rearranged in order of increasing importance. In the case of hickory nut trees, the first best solution has been determined to be $FA_{ht}$, followed by three further best solutions that have been determined to be $FA_{at}$. The rest of the answers should be $FA_{nt}$ on standard trees. In mathematics, there are three possible cases to generate new locations for a flock of flying squirrels. The global and local optima of the ASSOA algorithm are shown in the plot of Fig. 1. As shown in the figure, the movement of the squirrel based on the three cases can lead to the optimum global location.
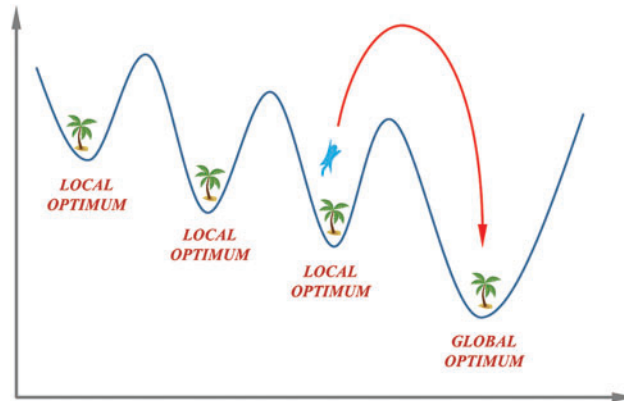
**Figure 1:** Global and local optima during the search by the ASSOA algorithm

**Case 1**: Moving to the hickory nut tree and location of $FL_{at}$:

$$FL_{at}(t+1) = f(x) = \begin{cases} FL_{at}(t) + d_g \times G_c(FL_{ht}(t) - FL_{at}(t)), & if \ R_1 < P_{dp} \\ Random location, otherwise \end{cases} \tag{4}$$

**Case 2**: Moving to the acorn nut trees and location of $FL_{nt}$:

$$FL_{nt}(t+1) = f(x) = \begin{cases} FL_{nt}(t) + d_g \times G_c(FL_{ht}(t) - FL_{nt}(t)), & if \ R_2 < P_{dp} \\ Random location, otherwise \end{cases} \tag{5}$$

**Case 3**: moving to the hickory nut tree and location of $FL_{nt}$:

$$FL_{nt}(t+1) = f(x) = \begin{cases} FL_{nt}(t) + d_g \times G_c(FL_{ht}(t) - FL_{nt}(t)), & if \ R_3 < P_{dp} \\ Random location, otherwise \end{cases} \tag{6}$$

where the variables $R_1$, $R_2$, and $R_3$ are selected randomly from the range [0,1], $t$ denoted the iteration number, distance for gliding is referred to as $d_g$, which is selected randomly. To achieve balance the exploitation and exploration, the value of $G_c$ is set as 1.9. For the three cases, the value of $P_{dp}$ is set as 0.1. More details about the ASSOA are presented in [32].

### 3.2 The Proposed Feature Selection Method

The binary ASSOA search strategy for selecting wrapper features is introduced in this section. Using the wrapper-based strategy, several classifiers are required to choose the optimal subset of features from a high-dimensional dataset. The wrapper-based technique is used to handle the challenge of high-dimensional feature selection.

1. Search algorithm
2. Feature selection basis
3. Classification method

Classification algorithms such as SVM, KNN, and artificial neural networks (ANN) are all prevalent. As a result, this research employs the KNN classifier, which can be used simply and easily. In addition, for the feature selection problem, KNN is one of the most commonly used classifiers. SVM and ANN can also be used to evaluate the quality of features that have been picked [33–38].

The proposed bASSOA employs a KNN classifier to make sure that only decency-preserving features are used. Due to its primary goal of maximizing classification accuracy while minimizing the number of selected features and the error rate, feature selection methods aim to decrease both of these factors. As a result, Eq. (1) is employed in the proposed feature selection technique. To maximize the feature assessment norm, the proposed bASSOA approach proposes an adaptive search space exploration presented in Eq. (1).

$$Fitness = \alpha \gamma_R(D) + \beta \frac{|C - R|}{|C|} \tag{7}$$

In which the condition attribute set $R$ in relation to the decision $D$ has a classification quality denoted by $\gamma_R(D)$. The total number of features is denoted by C, and the number of selected features is referred to as R. Additional parameters that affect the classification accuracy are $\alpha \in [0,1]$ and $\beta = 1 - \alpha$. Based on the error rate and selected feature ratio, Eq. (1) can be converted into the minimization problem. Therefore, the minimization of the error rate can be performed using the following fitness function, where $E_R(D)$ is the error rate.

$$Fitness = \alpha E_R(D) + \beta \frac{|R|}{|C|} \tag{8}$$

To select the best set of features, the resulting best solution is converted into binary 0 or 1. To perform this conversion, the sigmoid function is usually employed as represented by the following equation where $S_{best}$ refers to the best position at iteration $t$.

$$S_d(t+1) = \begin{cases} 0, \text{if } Sigmoid(S_{best}) < 0.5 \\ 1, \text{otherwise} \end{cases} \tag{9}$$

The steps of the proposed feature selection method are listed in Algorithm (1).

---

**Algorithm 1:** The proposed bASSOA algorithm for feature selection

---

1 **Initialize** ASSOA population locations $FL_i$, and velocities $FV_i$, for i = 1, 2, … , n, and max iterations *iter_max*
2 **Evaluate** fitness function *Fn* for each $FL_i$
3 **Find** best squirrel $FL_{best}$
4 **Train** k-NN based model and then calculate error
5 **While** t < *iter_max* **do**
6　**Apply** ASSOA algorithm
7　**Evaluate** fitness function for each agent
8　**Change** solution to binary solution (0 or 1) based on Eq. (9)
9　**Update** best solution
10　**Update** parameters
11 **end while**
12 **return** best solution

---

## 4 Experimental Results

Ten feature selection benchmark datasets from the UCI repository have been used to evaluate the proposed binary feature selection method's performance (UCI Feature selection dataset 2017). The number of features and occurrences in each of the databases is different. Tab. 1 provides a

comprehensive overview of all datasets. In order to choose the best features, the suggested bASSOA was used. The suggested technique selects the optimal features based on the accuracy given by the KNN classifier. Using empirical data, the KNN classifier sets K to 5. In cross-validation, each dataset's instances are randomly divided into three sets: training, validation, and testing. The suggested bASSOA algorithm's performance was verified using K-fold cross-validation that makes use of $K - 1$ folds for training, validation, and testing.

**Table 1:** Physical datasets included in the conducted experiments

| Dataset | #Features | #Samples | #Classes | Notation |
|---|---|---|---|---|
| Climate | 20 | 540 | 2 | D1 |
| Glass | 9 | 214 | 7 | D2 |
| IonosphereEW | 34 | 351 | 2 | D3 |
| Robot-failures-lp1 | 90 | 88 | 4 | D4 |
| Robot-failures-lp2 | 90 | 47 | 5 | D5 |
| Robot-failures-lp3 | 90 | 47 | 4 | D6 |
| Robot-failures-lp4 | 90 | 114 | 3 | D7 |
| Robot-failures-lp5 | 90 | 164 | 5 | D8 |
| WineEW | 13 | 178 | 3 | D9 |
| SonarEW | 60 | 208 | 2 | D10 |

### 4.1 Evaluation Criteria

Performance evaluation measures used to evaluate the suggested strategy are shown in Tab. 2. There are six criteria that make up these metrics: standard deviation, average fitness size, mean fitness, average error, worst fitness, and best fitness. In this part, the evaluation of the proposed approach using these metrics is presented and discussed.

**Table 2:** Performance metrics used in measuring the performance of the achieved results

| Metric | Value |
|---|---|
| Best fitness | $bmin_{i=1}^{M} g_{*}^{i}$ |
| Worst fitness | $Max_{i=1}^{M} g_{*}^{i}$ |
| Average error | $\frac{1}{M} \sum_{j=1}^{M} \frac{1}{N} \sum_{i=1}^{N} mse\,(C_i, L_i)$ |
| Mean fitness | $\frac{1}{M} \sum_{M}^{1} g_{*}^{i}$ |
| Average fitness size | $\frac{1}{M} \sum_{i=1}^{M} \frac{size\,(g_{*}^{i})}{D}$ |
| Standard deviation | $\sqrt{\frac{1}{M-1} \sum \left(g_{*}^{i} - Mean\right)^{2}}$ |

### 4.2 Algorithm Configuration

The configuration parameters of the proposed approach are presented in Tab. 3. In this table, the number of iterations is 50, and the number of search agents is 10. In addition, the ranges of the variables of the optimization equations are presented in the table.

**Table 3:** The configuration parameters of the proposed bASSOA algorithm

| Parameter | Value |
|---|---|
| *Num of iterations* | 50 |
| *Num of agents* | 10 |
| $R_1$ | [0,1] |
| $R_2$ | [0,1] |
| $R_3$ | [0,1] |
| $G_c$ | 1.9 |
| $P_{dp}$ | 0.1 |
| $c_1, c_2, c_3$ | [0,1] |
| $P_a$ | [0,1] |
| $P_d$ | [0,1] |
| *a, r, b, p* | [0,1] |

### 4.3 Evaluation Results

There are eleven current state-of-the-art ways to compare the proposed bASSOA feature selection approach with; bGWO, PSO, bPSO, SFS bWOA, bMGWO and SBO, bSBO, bPSO, bGWO GA and the suggested method, bASSOA. Mean values for each approach have been obtained for comparison after each method was run 30 times. In order to evaluate the suggested bASSOA approach, the six performance criteria listed above were taken into consideration.

Data transformation and feature selection approaches were tested on the original datasets without deleting any attributes in order to show the impact of the suggested methods on the original datasets. Tab. 4 shows the average inaccuracy of the proposed bASSOA and the other approaches investigated. According to the table, the bASSOA technique has the best overall average error of 0.63, which is lower than the other methods' average errors. Using data transformation, we can examine the effect that the suggested technique of feature selection has on the average select size of chosen features, as shown in Tab. 5. Based on this chart, it's clear that using bASSOA returns the largest select size.

**Table 4:** Average error

| Dataset | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **bASSOA** | **0.33** | **0.27** | **0.34** | **0.42** | **0.27** | **0.29** | **0.38** | **0.18** | **0.41** | **3.40** | **0.63** |
| bGWO | 0.35 | 0.28 | 0.36 | 0.43 | 0.29 | 0.30 | 0.39 | 0.18 | 0.40 | 3.56 | 0.65 |
| bFA | 0.34 | 0.30 | 0.37 | 0.43 | 0.28 | 0.30 | 0.38 | 0.18 | 0.41 | 3.63 | 0.66 |
| bPSO | 0.34 | 0.31 | 0.36 | 0.42 | 0.29 | 0.29 | 0.39 | 0.18 | 0.41 | 3.63 | 0.66 |

(Continued)

**Table 4:** Continued

| Dataset | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|
| bSFS | 0.35 | 0.29 | 0.36 | 0.39 | 0.28 | 0.29 | 0.40 | 0.18 | 0.50 | 3.48 | 0.65 |
| bGA | 0.33 | 0.29 | 0.36 | 0.42 | 0.28 | 0.29 | 0.39 | 0.18 | 0.42 | 3.33 | 0.63 |
| bWOA | 0.33 | 0.30 | 0.35 | 0.43 | 0.28 | 0.30 | 0.39 | 0.18 | 0.41 | 3.73 | 0.67 |
| bMVO | 0.33 | 0.31 | 0.37 | 0.42 | 0.29 | 0.30 | 0.38 | 0.18 | 0.41 | 3.90 | 0.69 |
| bSBO | 0.35 | 0.29 | 0.37 | 0.39 | 0.27 | 0.30 | 0.38 | 0.18 | 0.39 | 3.69 | 0.66 |
| bGWO_GA | 0.36 | 0.33 | 0.36 | 0.39 | 0.30 | 0.29 | 0.37 | 0.18 | 0.39 | 3.55 | 0.65 |
| bGWO_PSO | 0.34 | 0.28 | 0.35 | 0.39 | 0.29 | 0.29 | 0.39 | 0.19 | 0.39 | 3.45 | 0.64 |
| bMGWO | 0.34 | 0.27 | 0.34 | 0.39 | 0.29 | 0.32 | 0.39 | 0.19 | 0.41 | 3.39 | 0.63 |

**Table 5:** Average select size

| Dataset | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **bASSOA** | **0.477** | **0.272** | **0.542** | **0.664** | **0.392** | **0.408** | **0.778** | **0.612** | **0.479** | **0.656** |
| bGWO | 0.541 | 0.402 | 0.636 | 0.643 | 0.547 | 0.564 | 0.836 | 0.661 | 0.648 | 0.511 |
| bFA | 0.676 | 0.624 | 0.644 | 0.722 | 0.606 | 0.661 | 0.898 | 0.773 | 0.711 | 0.594 |
| bPSO | 0.671 | 0.627 | 0.644 | 0.836 | 0.604 | 0.668 | 0.786 | 0.730 | 0.742 | 0.619 |
| bSFS | 0.496 | 0.450 | 0.639 | 0.650 | 0.428 | 0.622 | 0.747 | 0.661 | 0.611 | 0.669 |
| bGA | 0.636 | 0.527 | 0.644 | 0.686 | 0.584 | 0.643 | 0.911 | 0.717 | 0.698 | 0.566 |
| bWOA | 0.766 | 0.468 | 0.644 | 0.822 | 0.620 | 0.814 | 0.911 | 0.773 | 0.773 | 0.644 |
| bMVO | 0.641 | 0.607 | 0.652 | 0.664 | 0.672 | 0.650 | 0.873 | 0.730 | 0.723 | 0.614 |
| bSBO | 0.656 | 0.566 | 0.569 | 0.736 | 0.626 | 0.707 | 0.786 | 0.786 | 0.661 | 0.669 |
| bGWO_GA | 0.616 | 0.560 | 0.836 | 0.650 | 0.636 | 0.664 | 0.786 | 0.786 | 0.661 | 0.669 |
| bGWO_PSO | 0.536 | 0.427 | 0.636 | 0.678 | 0.527 | 0.593 | 0.736 | 0.636 | 0.661 | 0.414 |
| bMGWO | 0.587 | 0.390 | 0.625 | 0.697 | 0.445 | 0.507 | 0.679 | 0.657 | 0.561 | 0.657 |

On the other hand, the average fitness of the quality of the selected features is measured, and the results are recorded in Tab. 6. As presented in this table, the overall average fitness of the proposed approach is 0.775, which is the best value among the other approaches. In addition, the best, worst, and standard deviation fitness are measured based on the selected features, and the results are presented in Tabs. 7–9, respectively. The recorded results in these tables confirm the superiority and effectiveness of the proposed feature selection algorithm.

**Table 6:** Average fitness

| Dataset | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **bASSOA** | **0.361** | **0.282** | **0.478** | **0.500** | **0.306** | **0.449** | **0.988** | **0.459** | **0.712** | **3.209** | **0.775** |
| bGWO | 0.396 | 0.317 | 0.526 | 0.530 | 0.299 | 0.464 | 1.008 | 0.475 | 0.721 | 3.558 | 0.829 |
| bPSO | 0.383 | 0.345 | 0.532 | 0.514 | 0.315 | 0.455 | 1.015 | 0.471 | 0.729 | 3.626 | 0.838 |
| bSFS | 0.397 | 0.290 | 0.380 | 0.587 | 0.299 | 0.276 | 0.403 | 0.482 | 0.724 | 4.813 | 0.865 |

(Continued)

**Table 6:** Continued

| Dataset | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|
| bWOA | 0.384 | 0.330 | 0.523 | 0.525 | 0.306 | 0.458 | 1.008 | 0.470 | 0.728 | 3.726 | 0.846 |
| bFA | 0.391 | 0.335 | 0.538 | 0.523 | 0.305 | 0.466 | 1.000 | 0.473 | 0.727 | 3.619 | 0.838 |
| bSBO | 0.391 | 0.291 | 0.474 | 0.564 | 0.325 | 0.300 | 0.385 | 0.683 | 0.729 | 4.875 | 0.902 |
| bMVO | 0.384 | 0.339 | 0.539 | 0.512 | 0.318 | 0.459 | 1.007 | 0.471 | 0.724 | 3.888 | 0.864 |
| bGA | 0.384 | 0.325 | 0.527 | 0.513 | 0.311 | 0.454 | 1.013 | 0.473 | 0.736 | 3.322 | 0.806 |
| bMGWO | 0.362 | 0.314 | 0.359 | 0.524 | 0.307 | 0.318 | 0.385 | 0.483 | 0.722 | 3.626 | 0.740 |
| bGWO_GA | 0.355 | 0.330 | 0.461 | 0.584 | 0.299 | 0.293 | 0.374 | 0.485 | 0.722 | 5.385 | 0.929 |
| bGWO_PSO | 0.379 | 0.284 | 0.353 | 0.590 | 0.289 | 0.289 | 0.390 | 0.486 | 0.722 | 4.524 | 0.830 |

**Table 7:** Best fitness

| Dataset | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **bASSOA** | **0.265** | **0.230** | **0.444** | **0.236** | **0.196** | **0.412** | **0.968** | **0.445** | **0.665** | **2.103** | **0.596** |
| bGWO | 0.285 | 0.247 | 0.463 | 0.321 | 0.211 | 0.425 | 0.968 | 0.448 | 0.684 | 2.043 | 0.609 |
| bFA | 0.324 | 0.230 | 0.473 | 0.377 | 0.211 | 0.429 | 0.964 | 0.440 | 0.680 | 2.103 | 0.623 |
| bGA | 0.285 | 0.264 | 0.453 | 0.321 | 0.226 | 0.425 | 0.964 | 0.453 | 0.680 | 1.881 | 0.595 |
| bPSO | 0.285 | 0.281 | 0.453 | 0.292 | 0.257 | 0.421 | 0.976 | 0.453 | 0.692 | 2.043 | 0.615 |
| bSFS | 0.346 | 0.237 | 0.501 | 0.371 | 0.211 | 0.437 | 1.003 | 0.453 | 0.692 | 3.258 | 0.751 |
| bSBO | 0.343 | 0.264 | 0.511 | 0.363 | 0.226 | 0.442 | 0.974 | 0.461 | 0.684 | 3.962 | 0.823 |
| bMVO | 0.343 | 0.239 | 0.473 | 0.363 | 0.226 | 0.434 | 0.976 | 0.448 | 0.692 | 1.800 | 0.599 |
| bWOA | 0.324 | 0.256 | 0.453 | 0.292 | 0.226 | 0.416 | 0.968 | 0.444 | 0.672 | 1.618 | 0.567 |
| bGWO_GA | 0.343 | 0.315 | 0.482 | 0.406 | 0.287 | 0.442 | 0.964 | 0.465 | 0.688 | 4.023 | 0.841 |
| bGWO_PSO | 0.363 | 0.273 | 0.511 | 0.377 | 0.241 | 0.434 | 0.972 | 0.465 | 0.699 | 2.992 | 0.733 |
| bMGWO | 0.313 | 0.281 | 0.452 | 0.303 | 0.302 | 0.429 | 0.992 | 0.451 | 0.690 | 1.558 | 0.577 |

**Table 8:** Worst fitness

| Dataset | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **bASSOA** | **0.460** | **0.382** | **0.585** | **0.646** | **0.377** | **0.483** | **1.071** | **0.496** | **0.764** | **5.418** | **1.068** |
| bGWO | 0.479 | 0.391 | 0.569 | 0.787 | 0.363 | 0.533 | 1.071 | 0.512 | 0.792 | 5.376 | 1.087 |
| bFA | 0.518 | 0.434 | 0.742 | 0.688 | 0.379 | 0.632 | 1.031 | 0.499 | 0.769 | 5.235 | 1.093 |
| bPSO | 0.498 | 0.400 | 0.597 | 0.745 | 0.409 | 0.515 | 1.071 | 0.491 | 0.757 | 6.063 | 1.155 |
| bSFS | 0.421 | 0.423 | 0.559 | 0.575 | 0.394 | 0.455 | 1.099 | 0.495 | 0.734 | 6.576 | 1.173 |
| bWOA | 0.498 | 0.417 | 0.597 | 0.745 | 0.394 | 0.498 | 1.051 | 0.487 | 0.777 | 5.598 | 1.106 |
| bSBO | 0.460 | 0.366 | 0.636 | 0.575 | 0.379 | 0.472 | 1.055 | 0.491 | 0.723 | 6.791 | 1.195 |
| bMVO | 0.460 | 0.383 | 0.636 | 0.674 | 0.394 | 0.481 | 1.071 | 0.491 | 0.781 | 5.619 | 1.099 |
| bGA | 0.460 | 0.442 | 0.636 | 0.759 | 0.379 | 0.489 | 1.071 | 0.504 | 0.843 | 5.255 | 1.084 |
| bGWO_GA | 0.460 | 0.425 | 0.578 | 0.575 | 0.409 | 0.472 | 1.055 | 0.482 | 0.730 | 7.437 | 1.262 |
| bGWO_PSO | 0.440 | 0.340 | 0.530 | 0.575 | 0.379 | 0.477 | 1.055 | 0.495 | 0.715 | 5.174 | 1.018 |
| bMGWO | 0.479 | 0.400 | 0.607 | 0.555 | 0.394 | 0.571 | 1.055 | 0.504 | 0.761 | 5.763 | 1.109 |

**Table 9:** Standard deviation fitness

| Dataset | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | Average |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|---------|
| **bASSOA** | **0.169** | **0.169** | **0.162** | **0.229** | **0.168** | **0.151** | **0.156** | **0.149** | **0.152** | **0.977** | **0.248** |
| bGWO | 0.188 | 0.178 | 0.165 | 0.252 | 0.177 | 0.168 | 0.161 | 0.150 | 0.166 | 1.037 | 0.264 |
| bFA | 0.186 | 0.184 | 0.200 | 0.219 | 0.180 | 0.181 | 0.157 | 0.149 | 0.161 | 1.037 | 0.265 |
| bPSO | 0.195 | 0.169 | 0.176 | 0.240 | 0.179 | 0.158 | 0.164 | 0.145 | 0.153 | 1.272 | 0.285 |
| bSFS | 0.188 | 0.210 | 0.184 | 0.225 | 0.212 | 0.163 | 0.180 | 0.169 | 0.173 | 1.230 | 0.293 |
| bGA | 0.186 | 0.183 | 0.181 | 0.241 | 0.173 | 0.154 | 0.167 | 0.149 | 0.171 | 1.074 | 0.268 |
| bWOA | 0.182 | 0.178 | 0.173 | 0.243 | 0.188 | 0.159 | 0.161 | 0.147 | 0.158 | 0.983 | 0.257 |
| bSBO | 0.183 | 0.173 | 0.187 | 0.235 | 0.193 | 0.153 | 0.176 | 0.147 | 0.152 | 1.268 | 0.287 |
| bGWO_GA | 0.180 | 0.182 | 0.171 | 0.222 | 0.183 | 0.155 | 0.172 | 0.142 | 0.155 | 1.503 | 0.307 |
| bGWO_PSO | 0.171 | 0.174 | 0.164 | 0.217 | 0.187 | 0.155 | 0.170 | 0.148 | 0.153 | 1.016 | 0.256 |
| bMGWO | 0.178 | 0.174 | 0.173 | 0.231 | 0.171 | 0.170 | 0.158 | 0.146 | 0.167 | 1.898 | 0.347 |
| bMVO | 0.176 | 0.175 | 0.176 | 0.215 | 0.189 | 0.151 | 0.161 | 0.147 | 0.162 | 1.108 | 0.266 |

In addition, the visualization of some of the achieved results is presented in Figs. 2 and 3. In these figures, the average error and average standard deviation of the achieved fitness are shown. As depicted in these figures, the best values achieved are based on the proposed bASSOA approach. These results confirm the superiority of the proposed approach.
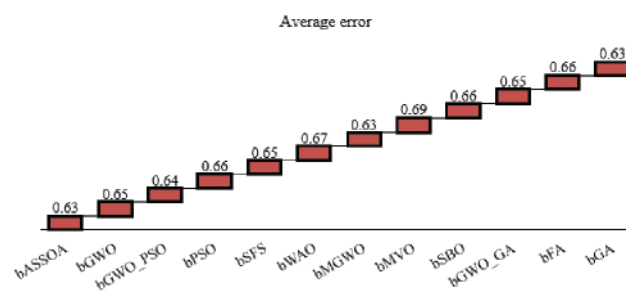


**Figure 2:** The average error of the achieved results using the proposed and other competing approaches
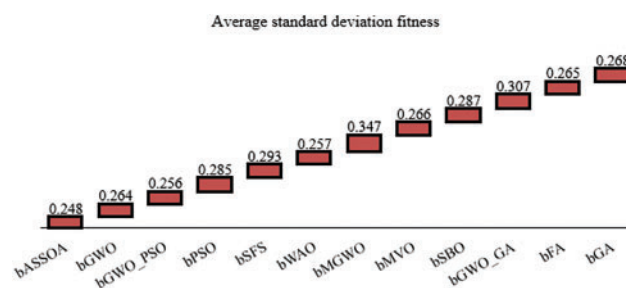


**Figure 3:** The average standard deviation of the achieved fitness

## 5  Conclusions

This paper introduces the binary adaptive squirrel search optimization algorithm (bASSOA). The suggested bASSOA approach was utilized to pick the important and non-redundant features from the high-dimensional datasets. First, the standard dataset is changed, and then an optimum selection of features is selected from the altered dataset. The performance of the proposed binary feature selection approach was evaluated using ten benchmark datasets. Six assessment criteria have been used to evaluate the proposed bASSOA technique in comparison to eleven other approaches in the literature, namely, bGWO, bGWO PSO, bPSO, bSFS, bWOA, bMGWO, bMVO, bSBO, bGWO GA, bFA, and bGA methods. Experimental results showed that the proposed bASSOA approach outperforms the other feature selection methods with an average error of 0.63. It has been shown that the bASSOA approach has the best standard deviation fitness of 0.248 when compared to the other methods. Future research will look into the possibility of increasing accuracy by adjusting various performance parameters. In addition, the proposed approach may be evaluated on additional real-world datasets.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

[1]  E.-S. M. El-Kenawy, S. Mirjalili, A. Ibrahim, M. Alrahmawy, M. El-Said *et al.,* "Advanced meta-heuristics, convolutional neural networks, and feature selectors for efficient COVID-19 X-ray chest image classification," *IEEE Access*, vol. 9, no. 1, pp. 36019–36037, 2021.

[2]  Q. Al-Tashi, H. Md Rais, S. Abdulkadir, S. Mirjalili and H. Alhussian, "A review of grey wolf optimizer-based feature selection methods for classification," in *Evolutionary Machine Learning Techniques. Algorithms for Intelligent Systems*. Singapore: Springer, 2020.

[3]  Q. Al-Tashi, H. Rais and S. Abdulkadir, "Hybrid swarm intelligence algorithms with ensemble machine learning for medical diagnosis," in *4th Int. Conf. on Computer and Information Sciences (ICCOINS)*, Kuala Lumpur, Malaysia, pp. 1–6, 2018.

[4]  M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131–156, 1997.

[5]  H. Liu, H. Motoda, R. Setiono and Z. Zhao, "Featureselection: An ever evolving frontier in data mining," in *Feature Selection in Data Mining*, Hyderabad, India, pp. 4–13, 2010.

[6]  M. Allam and M. Nandhini, "Optimal feature selection using binary teaching learning based optimization algorithm," *Journal of King Saud University–Computer and Information Sciences*, vol. 34, no. 2, pp. 329–341, 2022.

[7]  Y. Wah, N. Ibrahim, H. Hamid, S. Abdul-Rahman and S. Fong, "Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy," *Pertanika Journal of Science and Technology*, vol. 26, no. 1, pp. 329–340, 2018.

[8]  S. Bahassine, A. Madani, M. Al-Sarem and M. Kissi, "Feature selection using an improved Chi-square for Arabic text classification," *Journal of King Saud University–Computer and Information Sciences*, vol. 32, no. 2, pp. 225–231, 2020.

[9] M. Mazini, B. Shirazi and I. Mahdavi, "Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms," *Journal of King Saud University–Computer and Information Sciences*, vol. 31, no. 4, pp. 541–553, 2019.

[10] S. Thawkar and R. Ingolikar, "Classification of masses in digital mammograms using Biogeography-based optimization technique," *Journal of King Saud University–Computer and Information Sciences*, vol. 32, no. 10, pp. 1140–1148, 2020.

[11] F. Nie, H. Huang, X. Cai and C. Ding, "Efficient and robust feature selection via joint L2,1-norms minimization," in *Proc. of the 23rd Int. Conf. on Neural Information Processing Systems*, Vancouver, Canada, pp. 1813–1821, 2010.

[12] J. Wen, Z. Lai, Y. Zhan and J. Cui, "The L2,1-norm-based unsupervised optimal feature selection with applications to action recognition," *Pattern Recognition*, vol. 60, no. 1, pp. 515–530, 2016.

[13] Z. Lai, D. Mo, W. Keung, X. Yong, D. Miao *et al.,* "Robust discriminant regression for feature extraction," *IEEE Transactions in Cybernetics*, vol. 48, no. 8, pp. 2472–2484, 2018.

[14] C. Gunavathi and K. Premalatha, "Performance analysis of genetic algorithm with KNN and SVM for feature selection in tumor classification," *International Scholarly Science Research Innovation*, vol. 8, no. 8, pp. 1490–1497, 2014.

[15] K. Tan, E. Teoh, Q. Yu and K. Goh, "A hybrid evolutionary algorithm for attribute selection in data mining," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8616–8630, 2009.

[16] B. Oluleye, A. Leisa, J. Leng and D. Dean, "A genetic algorithm-based feature selection," *International Journal Electrical Communication and Computer Engineering*, vol. 5, no. 4, pp. 899–905, 2014.

[17] P. Moradi and M. Gholampour, "A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy," *Application Soft Computing*, vol. 43, no. December, pp. 117–130, 2016.

[18] A. Hafez, H. Zawbaa, E. Emery, H. Mahmoud and A. Hassanien, "An innovative approach for feature selection based on chicken swarm optimization," in *Int. Conf. of Soft Computing and Pattern Recognition (SoCPaR)*, Fukuoka, Japan, pp. 19–24, 2015.

[19] D. Sami Khafaga, A. Ali Alhussan, E. M. El-kenawy, A. E. Takieldeen, T. M. Hassan *et al.,* "Meta-heuristics for feature selection and classification in diagnostic breast cancer," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 749–765, 2022.

[20] D. Rodrigues, L. Pereira, R. Nakamura, K. Costa, Y. Xin-She *et al.,* "A wrapper approach for feature selection based on bat algorithm and optimum-path forest," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2250–2258, 2014.

[21] M. Mafarja, D. Eleyan, I. Jaber, A. Hammouri and S. Mirjalili, "Binary dragonfly algorithm for feature selection," in *Int. Conf. on New Trends in Computing Sciences (ICTCS)*, Amman, Jordan, pp. 12–17, 2017.

[22] G. Sayed, A. Hassanien and A. Azar, "Feature selection via a novel chaotic crow search algorithm," *Neural Computing and Applications*, vol. 31, no. 1, pp. 171–188, 2019.

[23] G. Sayed, G. Khoriba and M. Haggag, "A novel chaotic salp swarm algorithm for global optimization and feature selection," *Application Intelligence*, vol. 48, no. 1, pp. 3462–3481, 2018.

[24] T. Sridevi and A. Murugan, "A novel feature selection method for effective breast cancer diagnosis and prognosis," *International Journal of Computer Applications*, vol. 88, no. 11, pp. 0975–8887, 2014.

[25] V. Agrawal and S. Chandra, "Feature selection using artificial bee colony algorithm for medical image classification," in *Int. Conf. on Contemporary Computing (IC3)*, Noida, India, pp. 171–176, 2015.

[26] H. Sun and R. Grishman, "Lexicalized dependency paths based supervised learning for relation extraction," *Computer Systems Science and Engineering*, vol. 43, no. 3, pp. 861–870, 2022.

[27] Q. Li, H. Chen, H. Huang, X. Zhao, Z. Cai *et al.,* "An enhanced grey wolf optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis," *Computational and Mathematical Methods in Medicine*, vol. 2017, pp. 1–15, 2017.

[28] E. Pashaei and E. Pashaei, "An efficient binary chimp optimization algorithm for feature selection in biomedical data classification," *Neural Computing & Applications*, vol. 34, no. 8, pp. 6427–6451, 2022.

[29] J. Too and S. Mirjalili, "A hyper learning binary dragonfly algorithm for feature selection: A COVID-19 case study," *Knowledge-Based Systems*, vol. 212, no. 1, pp. 1–31, 2021.

[30] E. M. El-Kenawy, M. M. Eid, M. Saber and A. Ibrahim, "MbGWO-SFS: Modified binary grey wolf optimizer based on stochastic fractal search for feature selection," *IEEE Access*, vol. 8, pp. 107635–107649, 2020.

[31] N. Abdel Samee, E. M. El-Kenawy, G. Atteia, M. M. Jamjoom, A. Ibrahim *et al.,* "Metaheuristic optimization through deep learning classification of COVID-19 in chest X-ray images," *Computers, Materials & Continua*, vol. 73, no. 2, pp. 4193–4210, 2022.

[32] E.-S. M. El-Kenawy, S. Mirjalili, A. Ibrahim, M. Alrahmawy, M. El-Said *et al.,* "Advanced metaheuristics, convolutional neural networks, and feature selectors for efficient COVID-19 X-ray chest image classification," *IEEE Access*, vol. 9, no. 1, pp. 36019–36037, 2021.

[33] A. Abdelhamid and S. Alotaibi, "Optimized two-level ensemble model for predicting the parameters of metamaterial antenna," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 917–933, 2022.

[34] A. Abdelhamid and S. R. Alotaibi, "Robust prediction of the bandwidth of metamaterial antenna using deep learning," *Computers, Materials & Continua*, vol. 72, no. 2, pp. 2305–2321, 2022.

[35] D. Sami Khafaga, A. Ali Alhussan, E. M. El-kenawy, A. E. Takieldeen, T. M. Hassan *et al.,* "Metaheuristics for feature selection and classification in diagnostic breast cancer," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 749–765, 2022.

[36] D. Sami Khafaga, A. Ali Alhussan, E. M. El-kenawy, A. Ibrahim, S. H. Abd Elkhalik *et al.,* "Improved prediction of metamaterial antenna bandwidth using adaptive optimization of LSTM," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 865–881, 2022.

[37] E. -S. M. El-Kenawy, S. Mirjalili, F. Alassery, Y. Zhang, M. Eid *et al.,* "Novel meta-heuristic algorithm for feature selection, unconstrained functions and engineering problems," *IEEE Access*, vol. 10, pp. 40536–40555, 2022.

[38] A. Abdelhamid, E.-S. M. El-kenawy, B. Alotaibi, M. Abdelkader, A. Ibrahim *et al.,* "Robust speech emotion recognition using CNN+LSTM based on stochastic fractal search optimization algorithm," *IEEE Access*, vol. 10, pp. 49265–49284, 2022.