Chapter 7

A Hybrid Feature Selection Approach based on Information Theory and Dynamic Butterfly Optimization Algorithm for Data Classification

7.1. Introduction

Considering the incredible performance of BOA in the channel selection problem, we employ a novel dynamic variant of BOA (DBOA) [138] to solve the Curse of Dimensionality (CoD) problem. This issue describes the explosive nature of increasing data dimensions or features and its resulting exponential increase in computational efforts required for its processing and analysis. To resolve these issues, in this work, we develop a new hybrid feature selection method, namely, the Iterative Feature Selection using Dynamic Butterfly Optimization Algorithm-based Interaction Maximization (IFS-DBOIM) that combines Dynamic Butterfly Optimization Algorithm (DBOA) with a mutual information-based Feature Interaction Maximization (FIM) [22] scheme for selecting the optimal feature subset. There is evidence that DBOA performs better in exploration, exploitation, and avoidance of local optima entrapment, and FIM comparatively scores the maximum relevancy with minimum redundancy of the new features with previously selected ones. The performance of the proposed method is compared using twenty publicly available datasets with ten baseline feature selection approaches. The results revealed that IFS-DBOIM outperforms other approaches on most datasets, maximizing the percent classification accuracy with the least number of features. The nonparametric Wilcoxon rank test [95] confirms the statistical significance of these outcomes. Moreover, this method realizes the best trade-off between accuracy and stability. The detailed code of this chapter can be accessed from Link 1 given in the footer.

---

[1] https://drive.google.com/drive/folders/1IroiDhW-DlFxn74eEjMoCJaEIfvlXFO_?usp=sharing

7.2. Limitations of the Existing Literature

Most of the conventional feature selection methods implemented are based on filter and wrapper approaches that suffer from poor classification accuracy, high computational cost, and selection of irrelevant and redundant features. This is due to the limitations of the employed objective functions leading to overestimation of the feature significance. On the contrary, hybrid feature selection methods formulated from information theory and nature-inspired metaheuristic algorithms are preferred because of their high computational efficiency, scalability in avoiding redundant and less informative features, and independence from the classifier. However, these methods have the following drawbacks:

I. A poor trade-off between the exploration and exploitation phase
II. Getting stuck into an optimal local solution
III. Avoiding irrelevancy and redundancy of selected features.

The first and the second drawback is related to metaheuristic algorithm implementation, while the third is concerned with applied information-theoretic paradigms. This chapter focuses on all three above-mentioned issues to improve the robustness and preciseness of hybrid feature selection methods.

7.3. Our Contribution

The contributions of our method are recapitulated as follows:

I. Maximizing relevance between a set of features and class labels to improve the classification accuracy of the proposed model

II. Using a tri-objective function to determine the quality of each solution generated by the dynamic butterfly optimization, the objectives are: (1) classification accuracy maximization, (2) minimizing the size of optimal feature subset, and (3) mean FIM score

III. Understanding the compatibility of the proposed feature selection method with the set of classifiers mentioned above

7.4. Preliminaries

This section explains the basic concepts used within our study, namely, Dynamic Butterfly Optimization Algorithm (DBOA) and Feature Interaction Maximization (FIM). In the first subsection, we cover the basic search strategy used by butterflies and the solution quality improvement scheme introduced in the DBOA. Similarly, a three-way feature interaction approach for determining the relationship between previously and newly selected features is presented in the second subsection.

7.4.1. Dynamic Butterfly Optimization Algorithm

Dynamic Butterfly Optimization Algorithm (DBOA) is an improved variant of the conventional Butterfly Optimization Algorithm (BOA). Previously, DBOA has shown its ability to solve different high-dimensionality optimization problems such as stagnation into local minima and lack of solution diversity during the optimization process. The details of conventional BOA are given in Chapter 6. DBOA improves the performance of BOA by employing a Local Search Algorithm based on the Mutation (LSAM) operator on either the current best solution (g^*) or other solutions computed by the fitness function. LSAM receives the current best solution at the end of each iteration and updates it using the mutation operator. If the new mutated solution is better than the existing solution, then the new solution will replace the existing solution (g^*). Otherwise, LSAM selects a random solution from other solutions and performs the same mutation operation. If the fitness value of the mutated solution is better than the fitness value of the selected solution, then the process is repeated by *Num_iterations (N)* times. Finally, the computed solution is assigned as g^* for further processing. The pseudocode of the DBOA is discussed in Algorithm 7.1.

Algorithm 7.1: General Pseudocode of DBOA

```
• Initialize n butterflies population positions $x_i$ ($i = 1,2,\dots,n$)
• Set the initial value of parameters (switching probability $\rho$, sensory modality $c$, power exponent $a$,
  and the number of iterations $N$)
while not reach $N$ do
    for each butterfly bf in the population do
        Compute the fragrance value $f$ for each bf using Eq. 1
    end for
    Find the best butterfly bf
    Assign the best butterfly to $g^*$
    for each butterfly bf in the population do
        Generate a random value $r$ over the interval [0,1]
        if ($r < \rho$)
            Update bf position by using Equation 6.2      --- (Exploration phase)
        else
            Update bf position by using Equation 6. 3
        end if
        Evaluate the new butterfly
        If the new butterfly is better, update it in the population
    end for
    Update the value of the power exponent, and variable $c$
    Apply LSAM on the current best solution using the mutation operator (to improve the
    quality of solutions using algorithm 2)
    Update the best global solution if find the better solution
end while
Return the best solution found by the DBOA                ---- (Exploitation phase)
```

7.4.2. Feature Interaction Maximization

In information theory, the term interaction is mostly used to indicate the amount of information shared by a set of variables. It is a useful measure as it does not consider any prior assumptions about variables and can effectively deal with the nonlinear dependency between them. In Shannon's information theory, Entropy is a quantitative measure to show the uncertainty of a random variable. suppose $X = \{x_1, x_2, \dots, x_n\}$ is a discrete random variable and $Y = \{y_1, y_2, \dots, y_m\}$ is a class label. If the probability density function is denoted by $p(\mathrm{x})$ and $p(\mathrm{x})$ = Probability(X=x), then the entropy of X will be given by Eq. 7.1.

$$H(X) = - \sum_{i=1}^{N} p(x_i) * \log(p(x_i)) \tag{7.1}$$

where $0 \leq H(X) \leq 1$. The joint entropy $H(X,Y)$ of two discrete random variables X and Y is defined in Eq. 7.2.

$$H(X,Y) = -\sum_{i=1}^{N} \sum_{j=1}^{M} p(x_i, y_i) \log(p(x_i, y_i)) \tag{7.2}$$

If one of the two variables is known and the other is not, the remaining uncertainty is termed conditional entropy, and it is defined in Eq. 7.3.

$$H(X,Y) = -\sum_{i=1}^{N} \sum_{j=1}^{M} p(x_i, y_i) \log(p(x_i/y_i)) \tag{7.3}$$

The amount of information shared by both variables is termed mutual information, which can be computed by Eq. 7.4.

$$I(X,Y) = \sum_{i=1}^{N} \sum_{j=1}^{M} p(x_i, y_i) \log \frac{(p(x_i,\ y_i))}{p(x_i)*p(y_i)} \tag{7.4}$$

The value of MI given by the $I(X,Y)$ is always positive. A high MI value refers to significant associations between both variables; MI is zero if both variables are independent. In the feature selection, the MI can find the relationship between a variable X and a target class Y. This relation is also termed information gain, and a feature with higher mutual information is considered the most relevant to the target feature. The measures of information theory that can be derived from MI and used in our study are conditional mutual information $I(X_j;\ Y/X_i)$ and three-way interaction method $(X_j;\ X_i;\ Y)$. In both cases, the relation between a feature and target class is studied in the context of other features and can be computed from Equation 7.5 and Equation 7.6, respectively.

$$I\left(X_j; Y/X_i\right) = H\left(X_j;\ Y\right) - H\left(X_j/\,Y, X_i\right) \tag{7.5}$$

$$I\left(X_j;\ X_i;\ Y\right) = I\left(X_j, X_i;\ Y\right) - I\left(X_j; Y\right) - I(X_i;\ Y) \tag{7.6}$$

Unlike MI, three-way interaction measures can be either positive, negative, or zero. A positive score refers to the combined information associated with two features that cannot be provided by each of them individually. It is negative when any of the two features can compute the combined information. A zero score shows that both features are independent and don't share any information. Suppose X_i is a candidate feature and X_s is a feature belonging to the subset S (i.e., it has already been selected), and C is a target class (attribute), Feature Interaction Maximization (FIM) can be defined as

$$FIM = \arg\max(I\ (X_i; C) + \min_{X_s \in S}(I(X_i; X_s; C)) \qquad (7.7)$$

where $$I(x_i; X_x; C) = I(x_i; X_s; C) - I(X_i; C) - I(X_s; C) \qquad (7.8)$$

In Equation 7.8, the mutual information between the feature X_i and C computes the relationship between the candidate feature and the class attribute. The interaction information among X_i, X_s, and C is the redundancy term. The feature which is selected is the one that maximizes the objective function defined in Equation 7.8. It has the maximum relevance to the class attribute and the minimum interaction with the selected features. The advantage of this criterion is its ability to select the features that have the highest discriminative power. The pseudocode of the three-way interaction maximization approach is given in Algorithm 7.2. Our main objective is to estimate a subset of those features that have maximum interaction with the candidate feature and some resemblance to those that are already present in the subset. This process is iteratively performed for all possible solutions until a subset with a maximum interaction score is determined for further processing.

Algorithm 7.2: Pseudocode of feature interaction maximization scheme

```
• Initialization phase:
    1. Set $X \leftarrow$ "initial set of $n$ features," $S \leftarrow$ "empty set."
    2. (Computation of the MI with the output class) For each feature $x_i \in X$ $I(C; x_i)$.
    3. Find the first feature $x$ that maximizes $I(C; x_i)$; set $X \leftarrow X \setminus \{x_i\}$; set $S \leftarrow \{x_i\}$.
• Greedy selection phase
    1. Repeat until $|S| \leftarrow k$;
    2. (Computation of the interaction information between variables) For all pairs of variables $I(x_i; X_x; C)$
       with $x_i \in X$, $x_s \in S$, compute $I(x_i; X_x; C)$ if it is not already available.
    3. (Selection of next feature) Choose feature $x_i$ as the one that maximizes Equation 7.8.
• Output the set $S$ with the selected features.
```

7.5. Proposed Methodology

The proposed IFS-DBOIM model is implemented in two phases: (i) the training phase using the DBOA, and (ii) the feature selection phase. In the first phase, DBOA is employed iteratively to determine the diversity associated with training samples. Then, feature interaction maximization is adopted in the second stage to select the best feature subset. Most feature selection methods opt for the fitness function with two contradictory objectives: (1) maximum classification accuracy rate, and (2) relative number of selected relevant features. Here, the relative number of selected features is derived by dividing the number of selected features by the total number of available features. Moreover, researchers have shown that the relevance of the selected features with class labels in each solution can be an effective objective that helps to obtain optimal solutions with higher classification accuracy. Therefore, in this work, we use a three-objective fitness function to evaluate the quality of each solution. The higher the classification accuracy rate, the lesser the relative number of selected features, and the higher the average feature interaction score in the solution leads to a better solution. Therefore, the proposed work defines a multiobjective fitness function for the most optimal feature set with the minimum number of features.

Firstly, all three objective functions are individually computed and then assigned to the three-objective fitness function defined in Equation 7.9. In the proposed IFS-DBOIM algorithm, each solution is evaluated according to the proposed fitness function, and the solution with the highest fitness score is assigned as the best solution.

$$Fitness\_function = f(f_1, f_2, f_3) = w_1 f_1 + w_2 f_2 + w_3 f_3 \tag{7.9}$$

where f_1 is the classification accuracy rate, f_2 is the relative number of selected relevant features, and f_3 is the average feature interaction score computed in Equation 7.7. The sum of all the coefficients equals 1. Second, the interaction maximization approach selects the subset of best interactive features from each solution obtained in the first phase. Since all the feature subset solutions can not be implemented simultaneously in the classification process, they are arranged in decreasing order of their overall interaction score. Finally, the feature subset with the maximum interaction score is used in the classification process. The general structure of the proposed feature selection scheme is displayed in Figure 7.1.

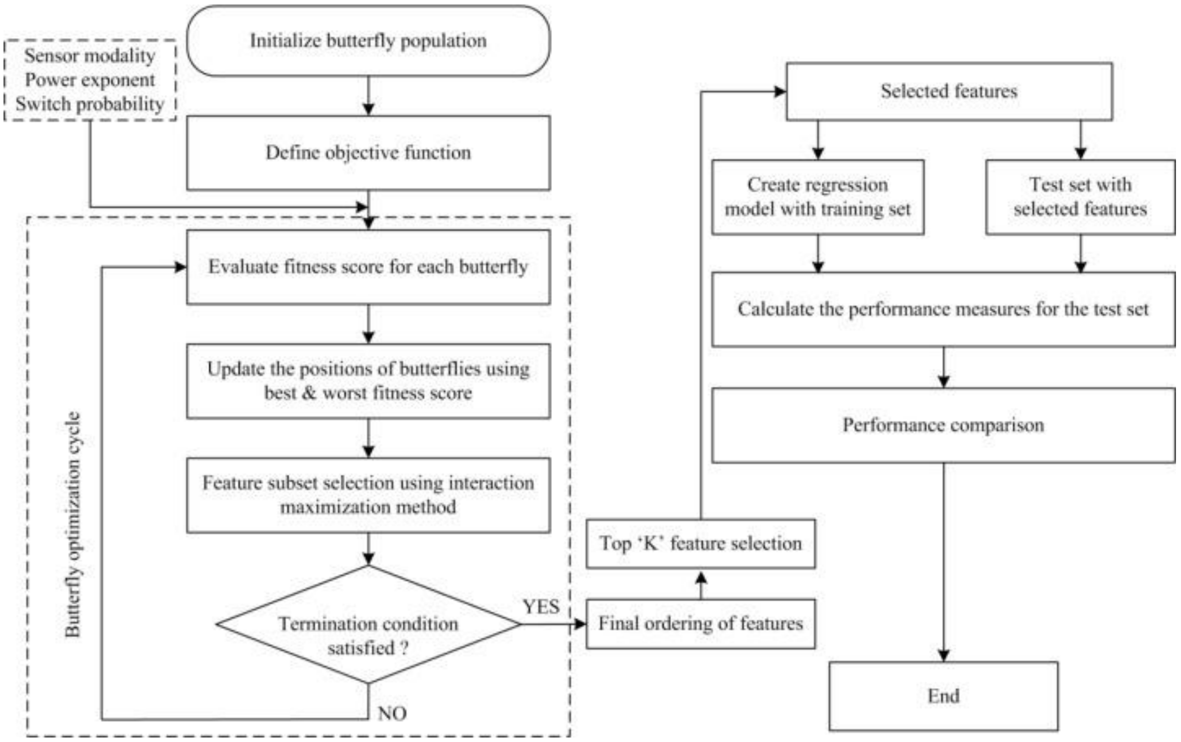


Figure 7.1. Block diagram of proposed feature selection method for data classification

7.5.1. Dynamic Butterfly Optimization-based Interaction Maximization Algorithm

The DBOA is a practical approach to covering the dynamics of the different butterflies' positions. Here, DBOA updates the EEG instances using Eq. 6.1 and Eq. 6.2 and computes the best solution in terms of minimum fitness value in the first iteration. During regular iterations, our approach intended to compute a better solution compared to the previous one. To do this step, it first updates the instances and then explores a fitness function with three objectives given in Eq (7.9). Here, DBOA updates the EEG instances according to the generated random number (r). The algorithm executes the global update scheme if $r < 0.5$ otherwise a local update policy is employed and fitness value is computed for each updated instance. If the estimated solution is better than the previous solution then we replace the previous one with the updated one otherwise a local search scheme is applied using a mutation operator ($\alpha = 0.2$). This modification helps to realize a better solution in the same iteration. Ultimately, only the best among all the solutions is stored as the best solution. This process continues until the criteria are matched. The pseudocode of the proposed IFS-DBOIM algorithm is given in Algorithm 7.4.

Algorithm 7.4: General Pseudocode of the IFS-DBOIM algorithm

```
Objective function $f(x)$, $x = (x_1, x_2, \ldots, x_d)$;
Initialize $n$ butterflies population positions $x_i$ $(i = 1,2, \ldots, n)$
Set the initial value of parameters (switching probability: $\rho$, sensory modality: $c$, power exponent: $a$)
while the stopping criteria are not met, do
    for each butterfly in the population do
        Calculate the fragrance using Equation 6.1;
    end for
    Find the best butterfly;
    for each butterfly in the population do
        Generate a random number $rand$ from [0,1];
        if $rand < \rho$
            Move towards the best butterfly using Equation 6.2;
        else
            Move randomly
        end if
    end for
    Update the value of the power exponent, and variable c
    Apply LSAM on the current best solution using the mutation operator
    Evaluate the new butterfly
    Update the best global solution and find the better solution
end while
Return the best solution found by the DBOA algorithm
```

7.5.2. Complexity Analysis of IFS-DBOIM

The computational complexity of the IFS-DBOIM approach is estimated as a performance indicator that mainly depends on three steps: (1) position update using the DBOA, (2) selecting the best features using the FIM method, and (3) classifier training time. Therefore, the complexity can be mathematically represented as $O(C_{DBOA} + C_{FIM} + C_{cl})$ where O denotes worst-case time complexity, C_{DBOA}, C_{FIM}, and C_{cl} indicate the complexity of DBOA implementation while modifying the location of each instance, the FIM method, and the execution time of the classifier in the training phase, respectively. Here, the overall complexity is represented in terms of the number of iterations (K), population size (N), dimension size (d), and training time (T_t) and can be calculated as $O\ (K\ \{(N + K) + d\} + T_t)$ where $K(N + K)$ denotes C_{DBOA}, Kd is C_{FIM}, and T_t is classifier training time (C_{cl}). It is deduced from the above expression that the complexity of the

IFS-DBOIM method mainly depends on the number of iterations and the population size. In the case of $K \gg N$, the total complexity will be $\mathrm{O}\ (\{K^2 + dK\} + T_t)$ but if $N \gg K$, it will become $\mathrm{O}\ (N + T_t)$.

7.6 Experimental Results & Discussion

In this section, the performance of the proposed model is validated on twenty benchmark datasets taken from the University of California Irvine (UCI) repository and compared with ten state-of-the-art optimization algorithms. The details of all ten datasets are already discussed in Chapter 2.

7.6.1. Experimental Setup

In our work, a five-fold cross-validation scheme is applied to each dataset to test the effectiveness of the machine learning model and avoid overfitting issues. In other words, the datasets are divided into training and testing data samples in the following manner. In the first iteration, 80% of feature vectors are used for training, and the remaining 20% are employed for testing purposes. In the next, another 20% of feature vectors are used for testing, and the rest of the 80% are employed for the training set. This process is repeated until all the feature vectors are used for testing the proposed algorithm. All the data instances are normalized in the interval of 0 and 1. To quantify results statistically, each fold is repeated 30 times, and every experiment is continually performed 100 times, giving a total of 15000 runs for each dataset. Next, the predictive classification model is developed on the training data and validated on the testing data, and the results are computed. Finally, the results are averaged over all the folds and compared with state-of-the-art methods. All parameter settings for each of the baseline algorithms and the proposed algorithms are given in Table 7.1.

Table 7.1. The setting of algorithm-specific parameters

Algorithm	Parameter Setting
GA	Crossover_ratio = 0.9, Mutation_ratio = 0.1, M (number of runs) = 30, N (number of iterations) = 100
GOA	c_Max = 1, c_Min = 0.0004, M (number of runs) = 30, N (number of iterations) = 100
PSO	Acceleration_constants ($C1$ = 2, $C2$ = 2), M (number of runs) = 30, N (number of iterations) = 100
ALO	I = 1 set as in the original article [96]
SCA	a- Power exponent = 2 as in the original article [97]
BOA	a- Power exponent =0.1 as in the original article [8]
CBOA	Control parameter (P) =0.5, Chaotic numbers ∈ (0,1), Constant (b) = 0.2 These parameters are listed in the original article [9]
DBOA	a- Power exponent =0.1, nm = 20, mu= 0.1 [143]
OEbBOA	M (number of runs) = 20, P (number of search agents) = 7, N (number of iterations) = 100, Search domains = [0,1], a- Power exponent = 0.1, c- Sensor modality = [0.01-0.25], τ_{max} (upper bound of shape tune parameter) = 4, τ_{min} (lower bound of shape tune parameter) = 0.01, F (scaling parameter) = [0,1], Crossover_ratio = 0.7, P_r (random variation paramter) = 0.7. These parameters are listed in the original article [169]
S-bBOA	K for cross-validation = 5, M (number of runs) = 20, P (number of search agents) = 7, N (number of iterations) = 100, Search domains = [0,1], Crossover_ratio = 0.9, Mutation_ratio = 0.1, a- Power exponent = 0.1, c- Sensor modality [min, max] = [0.01-0.25]. These parameters are listed in the original article [8].
IFS-DBOIM	K for cross-validation = 5, M (number of runs) = 30, N (number of iterations) = 100, a- Power exponent = 0.5, c- Sensor modality [min, max] = [0.01-0.50], SVM parameters (C= 0.01, γ = 100)

7.6.2. Results Analysis

In this subsection, two groups of experiments are conducted to evaluate the performance of the proposed IFS-DBOIM method. The main reason for categorizing the experiments is to understand the comparative performance of two different sets of optimization algorithms. The first set contains six optimization algorithms that are conventionally used, while the second set consists of four popular variants of the original BOA. For example, in the first group of experiments, the results obtained by the IFS-DBOIM method are compared with the following metaheuristic methods, namely, Ant Lion Optimization (**ALO**) [140], original Butterfly Optimization Algorithm (**BOA**) [128], Genetic Algorithm (**GA**) [108], Grasshopper Optimization

Algorithm (**GOA**) [141], Particle Swarm Optimization (**PSO**) [109], and Sine- Cosine Algorithm (**SCA**) [142]. In contrast, in the other group, the results of our method are compared with (1) Chaotic BOA (**CBOA**) [143], (2) Dynamic Butterfly Optimization Algorithm (**DBOA**) [138], (3) Optimization and Extension of binary Butterfly Optimization Algorithm (**OEbBOA**) [144], and (4) S-shaped binary Butterfly Optimization Algorithm (**S-bBOA**) [145]. Note that all results related to the aforementioned algorithms are obtained from Sadeghian et al. (2021) [47]. In addition, the maximum number of fitness evaluations is used to fix the number of iterations in the current work to have a fair comparison between the state-of-the-art methods and the proposed IFS-DBOIM algorithm. The main reason for selecting the number of fitness evaluations to fix the number of iterations is that each time the fitness is evaluated; it conveys some information about the problem instance. Thus, limiting the number of fitness evaluations reflects the total "amount of information" that one algorithm can obtain from the problem. Here, the obtained information is represented in terms of two performance measures: (1) classification accuracy and (2) fitness score to compare the performance of all the listed algorithms.

7.6.2.1. Performance Comparison in the First Group Experiment

Table 7.2 summarizes the results that are obtained from the first group of experiments. The performance of the IFS-DBOIM method and other aforementioned metaheuristic feature selection methods are compared in achieving classification accuracy (*CA*) and the number of selected features (*P*) by each method over 30 iterations. The maximum classification accuracy percent and the optimal number of selected features are shown in bold. In addition, the mean, standard deviation (S.D), and rank correspond to the average classification accuracy, standard deviation, and effectiveness of the respective algorithm. The approach with a lower rank is considered more effective than the one with a higher rank.

According to Table 7.2, the IFS-DBOIM method has shown a superior classification accuracy (93.36%) than six other baseline feature selection methods on all the datasets, except for four (Australian, Exactly, M-of-N, and Vowel) on which PSO and SCA outperform the other approaches. It should be noted that these have the lowest number of features among all the datasets. This observation shows that the PSO and SCA can effectively detect the change in the average fitness function in the early phase of optimization when employed on low-dimensional datasets. A few other factors such as the initialization method, the search mechanism, and the effective parameter tuning might also favor the better performance of PSO and SCA. However, the IFS-DBOIM achieves the best mean classification accuracy of 90.43%, followed by PSO which realizes a mean classification accuracy of 86.15%. Otherwise, our approach performs exceptionally well on high-dimensional datasets (Tox-171, Yale, and Penglungew) compared to the PSO method. On these three datasets, our method obtains a mean classification accuracy of 94.71%, which is 10.71% higher than the mean classification accuracy of 84% achieved by the PSO method.

Table 7.2 ranks the methods' performances in terms of their average classification accuracy rates. According to the ranking, the IFS-DBOIM approach realizes the best overall classification performance, followed by PSO, GA, SCA, BOA, GOA, and ALO methods. This phenomenon is due to the improved local search ability and increased solution diversity of the introduced IFS-DBOIM approach, which finds the best outcome from the pool of various solutions. The outstanding performance of the proposed method on high-dimensional datasets can be explained in terms of the working of DBOA and FIM schemes. In the early phase of each iteration, LSAM employs mutation operation on the current best solution and computes the mutated solution. This intermediate solution is compared with other candidates' solutions, forming an exploration strategy. Therefore, this step enhances the algorithm's ability to discover improved solutions based on the current ones populating the search space. However, the IFS-DBOIM algorithm outperforms others on the high-dimensional datasets. Though the applied LSAM strategy effectively produces mutated solutions, they are the best in the event of increased features. It helps in avoiding the local optima, thereby replacing the worst solution with a better one.

In contrast to DBOA, the other competitive metaheuristic algorithms do not scale well with complexity. That is, where the number of elements exposed to mutation is large, there is often an exponential increase in search space size. This makes it extremely difficult to use these algorithms on high-dimensional optimization problems like medical image classification and microarray gene expression datasets. In addition, the role of FIM is also important to explain the outstanding performance of the proposed feature selection method. It is known that FIM filters the relevant set of features based on their relevance and redundancy with the previously selected features. Therefore, the probability of determining the set of more informative features increases with the dimensionality of datasets. This step not only helps to realize better classification accuracy but also reduces irrelevant and redundant features during search space reduction. However, other baseline algorithms lack filtering optimal feature sets because they don't use any selection criterion to measure the significance of the features. Based on the combined performance of DBOA and FIM, it can be concluded that the proposed IFS-DBOIM method effectively maintains the trade-off between the classification accuracy and the number of selected features.

In Table 7.2, the average (avg.) number of selected features using the IFS-DBOIM and six similar methods are reported. The experimental results demonstrate that the proposed method selects the least number of average features (154.28) from all the datasets compared to the overall average number of features (383.70). Henceforth IFS-DBOIM achieves a 59.79% feature reduction rate followed by GA, PSO, ALO, GOA, SCA, and BOA. It shows that the proposed method effectively reduces the large feature search space without compromising the classification accuracy. In a nutshell, the proposed IFS-DBOIM approach identifies the optimal feature combination with the highest probability. To summarize, our method shows great compatibility with the SVM classifier as it obtains the best classification accuracy on sixteen out of twenty datasets, thereby proving its potential to avoid the overfitting problem faced by other algorithms on high-dimension datasets. Table 7.3 recapitulates the fitness values of seven feature selection algorithms for all twenty datasets. It concludes the effectiveness of the IFS-DBOIM method over the remaining approaches except for four datasets (Australian, Exactly, M-of-N, and Vowel) on which PSO outperforms the others. Since the minimum fitness value indicates superior solution quality, the classification accuracy achieved by the PSO is better than the IFS-DBOIM method on the datasets mentioned above. Globally, the proposed feature selection method realized the lowest fitness values on the remaining sixteen datasets. Due to this, the maximum classification accuracy rate is achieved by our method on the corresponding datasets, followed by PSO, GA, SCA, BOA, GOA, and ALO.

Figure 7.2 shows the classification accuracies of seven feature selection methods on twenty experimental datasets. We selected Penglungew, TOX-171, and Yale datasets as the benchmark to interpret the average classification accuracy achieved by the algorithms because they are high-dimensional datasets with an imbalanced ratio between the number of instances and original features. On the Penglungew dataset, the ALO method realizes maximum classification accuracy in the initial iterations. Moreover, after forty iterations, the proposed method outperforms the others, showing that the effective number of features selected from the IFS-DBOIM approach bears more relevance than those computed by other methods. The average classification accuracy of our method is better than those of the other six methods after sixty iterations except over the Australian, Exactly, M-of-N, and Vowel datasets. Interestingly, PSO realized maximum classification accuracy on all these four datasets, which shows that its global search ability has the upper hand over our interaction maximization strategy in selecting the relevant features.

However, the drawback is that PSO selected more features in all four cases than our method. One plausible reason for this behavior is that the proposed method's FIM algorithm eliminates less informative features in the early iterations without considering their roles in classification accuracy. Overall, it can be concluded that the proposed method maintains the most robust balance between the lowest number of selected features and the classification accuracy.

The classification accuracy of the IFS-DBOIM approach for the remaining two high-dimensional datasets, TOX-171, and Yale is always higher than the other methods. The TOX-171 is a high-dimensional microarray dataset containing a huge number of irrelevant and redundant features. Our method achieves the best classification accuracy of 97.02% using only 2404 features compared to other feature reduction methods. Similarly, Yale is another popular high-dimensional face recognition-based imaging dataset over which our method achieves superior classification accuracy. The proposed method realized a maximum classification accuracy of 87.92% with the minimum number of selected features. The performance of the IFS-DBOIM approach indicates that it can effectively deal with the Curse of Dimensionality (CoD) of high-dimensional datasets.

7.6.2.2. Performance Comparison in the Second Group Experiment

In this subsection, another experiment is conducted to evaluate the performance of the IFS-DBOIM method. Table 7.4 summarizes the experimental results regarding classification accuracy and the number of selected features of the IFS-DBOIM and the four variants of the original BOA (CBOA, DBOA, OEbBOA, S-bBOA). The maximum classification accuracy and the optimal number of selected features are shown in bold. According to Table 7.4, the IFS-DBOIM method achieves maximum classification accuracy except for four datasets (Australian, Exactly, M-of-N, and Vowel) on which DBOA and S-bBOA outperform the other approaches with a slight performance difference from the proposed method. The IFS-DBOIM approach realizes much higher classification accuracy than other similar methods on sixteen datasets. The mean classification accuracy of the IFS-DBOIM method is 90.43% which is a 1.28% enhancement in the conventional DBOA, which forms a part of our method. Since the standard deviation of our method is lower than that of the DBOA, most of the individual classification results are ideal and clustered around the mean classification accuracy.

Table 7.2. Performance of the IFS-DBOIM method in terms of classification accuracy (*CA*) and the number of selected features (*P*) for UCI datasets in the first group of experiments.

No.	ALO		BOA		GA		GOA		PSO		SCA		IFS-DBOIM			
	CA	*P*	*CA*	*P*	*CA*	*P*	*CA*	*P*	*CA*	*P*	*CA*	*P*	SVM	NB	DT	*P*
1	79.17	06.61	78.64	07.06	84.10	05.63	79.54	06.42	**85.04**	06.39	78.91	06.43	82.10	85.32	80.12	04.18
2	71.03	09.44	72.30	11.63	75.05	09.03	72.51	09.93	76.23	09.97	72.68	12.36	**83.68**	67.80	69.33	07.30
3	93.40	11.11	93.05	13.06	94.52	09.33	93.05	10.97	95.69	10.18	93.49	12.66	**98.46**	86.52	79.66	04.20
4	70.40	07.08	75.11	11.13	81.70	08.03	72.30	06.69	**87.46**	06.73	87.46	10.90	83.20	77.54	69.12	05.78
5	68.34	09.71	67.98	11.46	70.63	08.36	68.54	09.46	71.70	09.01	68.50	11.60	**80.02**	89.66	91.28	06.02
6	55.23	49.19	55.70	60.56	60.02	46.33	56.03	49.38	61.48	48.57	55.73	57.06	**74.66**	83.69	85.20	39.60
7	88.04	15.79	88.37	20.13	91.41	13.80	88.85	16.64	93.07	15.80	88.09	17.76	**98.33**	81.32	80.33	08.38
8	75.27	45.08	75.50	52.06	80.13	40.16	75.64	44.37	80.97	42.33	75.73	48.23	**96.18**	83.18	89.70	29.02
9	84.20	07.18	90.20	11.66	91.15	08.50	85.40	07.22	**95.96**	06.97	92.48	10.30	77.80	61.11	65.50	06.42
10	93.78	09.36	93.59	08.60	94.45	05.36	93.56	09.87	95.00	08.74	93.89	07.83	**99.34**	64.40	60.33	02.73
11	90.81	160.86	91.18	162.00	93.77	135.96	91.24	161.40	93.94	151.74	91.86	178.46	**99.20**	80.03	80.02	98.62
12	84.58	20.24	85.22	24.83	87.86	19.20	85.13	20.24	88.81	19.81	85.11	26.23	**92.18**	**77.30**	80.30	14.20
13	83.47	28.97	84.20	40.96	89.57	26.56	84.77	29.25	91.80	28.59	85.32	35.83	**98.50**	81.20	89.12	19.40
14	89.58	28.16	90.60	42.33	91.71	28.80	90.19	28.88	93.02	29.00	90.62	41.03	**96.12**	72.40	76.50	21.68
15	75.09	10.82	76.59	14.96	80.67	11.03	75.65	10.86	81.73	10.50	77.90	14.06	**88.20**	68.02	73.10	07.25
16	74.51	2875.7	75.59	3621.3	82.14	2825.9	78.53	2878.4	85.36	2846.2	76.26	3268.73	**97.02**	87.58	81.18	2403.7
17	94.66	07.85	95.05	09.93	95.94	07.16	94.77	07.93	96.38	07.58	95.05	08.73	**99.33**	91.04	78.66	03.82
18	90.11	07.31	91.80	10.50	93.24	08.86	90.57	07.25	**93.50**	07.05	92.02	10.63	87.20	84.33	77.54	07.92
19	79.96	11.32	82.68	19.40	82.54	13.93	80.64	11.79	83.17	11.47	83.09	19.00	**89.22**	88.10	68.12	12.33
20	62.57	509.28	63.05	591.20	68.15	492.96	64.08	509.80	72.70	501.21	64.17	563.23	**87.92**	71.10	67.33	383.16
Mean	80.21	191.55	81.32	237.23	84.43	186.24	81.04	191.84	86.15	188.89	82.41	218.05	**90.43**	80.08	78.62	**154.28**
S.D.	10.75		10.74		09.67		10.40		09.41		10.69		**07.09**	**07.55**	08.21	
Rank	7	IV	5	VII	3	II	6	V	2	III	4	VI	**1**			I

Table 7.3. The average fitness values of all competing algorithms over 30 runs

No.	ALO	BOA	GA	GOA	PSO	SCA	IFS-DBOIM
1	0.21	0.21	0.16	0.20	**0.15**	0.21	0.15
2	0.29	0.28	0.25	0.27	0.23	0.27	**0.17**
3	0.07	0.07	0.05	0.07	0.04	0.07	**0.03**
4	0.29	0.25	0.18	0.27	**0.12**	0.23	0.14
5	0.31	0.32	0.29	0.31	0.28	0.31	**0.22**
6	0.44	0.44	0.40	0.44	0.38	0.44	**0.35**
7	0.12	0.12	0.08	0.11	0.07	0.12	**0.05**
8	0.24	0.24	0.20	0.24	0.19	0.24	**0.16**
9	0.16	0.10	0.09	0.15	**0.04**	0.08	0.06
10	0.06	0.06	0.05	0.06	0.05	0.06	**0.04**
11	0.09	0.09	0.06	0.09	0.06	0.08	**0.04**
12	0.15	0.15	0.12	0.15	0.11	0.15	**0.09**
13	0.16	0.16	0.10	0.15	0.08	0.15	**0.07**
14	0.10	0.10	0.08	0.10	0.07	0.10	**0.06**
15	0.25	0.23	0.19	0.24	0.18	0.22	**0.15**
16	0.25	0.24	0.18	0.21	0.14	0.24	**0.10**
17	0.05	0.05	0.04	0.05	0.03	0.05	**0.01**
18	0.10	0.08	0.07	0.09	**0.07**	0.08	0.07
19	0.20	0.18	0.17	0.19	0.17	0.17	**0.13**
20	0.37	0.37	0.32	0.36	0.27	0.36	**0.18**
Mean	0.20	0.19	0.15	0.19	0.14	0.18	**0.11**
S.D.	0.10	0.10	0.09	0.10	0.09	0.10	0.07
Rank	4	4	2	3	2	3	**1**

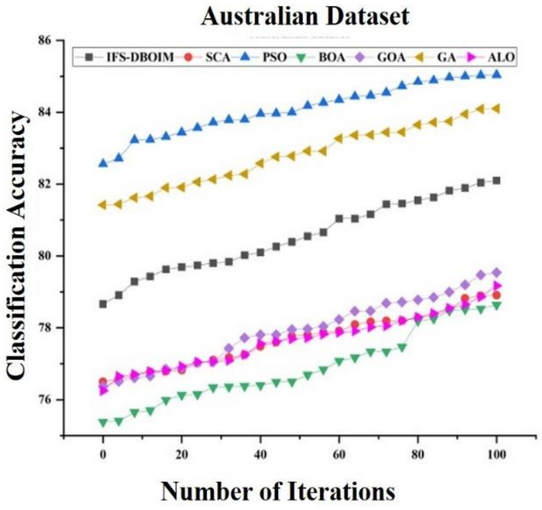
Table 7.4. Performance of the IFS-DBOIM method in terms of classification accuracy rates (*CA*) and the number of selected features (*P*) for UCI datasets in the second group of experiments.

No.	CBOA		DBOA		OEbBOA		S-bBOA		IFS-DBOIM			
	CA	*P*	*CA*	*P*	*CA*	*P*	*CA*	*P*	SVM	NB	DT	*P*
1	79.77	06.12	**88.04**	04.53	83.22	07.18	79.12	09.12	82.10	85.32	80.12	04.18
2	74.35	10.08	77.71	08.70	78.11	13.55	74.42	12.52	**83.68**	67.80	69.33	07.30
3	94.21	07.12	**98.46**	04.53	97.15	06.20	96.58	05.44	**98.46**	86.52	79.66	04.20
4	82.31	08.30	**99.85**	06.50	89.25	08.92	97.24	10.16	83.20	77.54	69.12	05.78
5	68.77	07.22	73.15	06.56	71.33	06.46	68.11	07.34	**80.02**	89.66	91.28	06.02
6	56.11	38.72	63.98	43.20	61.44	52.40	57.88	44.12	**74.66**	83.69	85.20	39.60
7	97.70	11.33	95.48	09.16	96.65	10.44	90.70	08.40	**98.33**	81.32	80.33	08.38
8	79.44	38.42	83.75	33.93	81.00	41.22	76.61	36.14	**96.18**	83.18	89.70	29.02
9	92.66	07.30	99.76	06.76	96.99	06.98	97.20	08.26	77.80	61.11	65.50	06.42
10	93.77	04.36	97.50	02.73	95.11	03.33	94.76	05.25	**99.34**	64.40	60.33	02.73
11	91.20	164.48	96.74	104.63	92.91	142.23	87.75	192.41	**99.20**	80.03	80.02	98.62
12	87.55	23.50	90.28	18.20	91.03	24.75	86.22	27.62	**92.18**	77.30	80.30	14.20
13	94.20	18.24	96.13	21.40	95.14	25.28	93.62	23.68	**98.50**	81.20	89.12	19.40
14	90.91	31.20	94.13	33.00	91.40	29.68	91.11	27.58	**96.12**	72.40	76.50	21.68
15	82.90	11.78	86.54	09.70	85.16	13.21	84.63	12.44	**88.20**	68.02	73.10	07.25
16	77.12	3320.11	89.07	2743.86	83.07	2907.46	76.11	3120.1	**97.02**	87.58	81.18	2403.74
17	96.44	04.98	98.16	04.70	98.44	05.76	96.53	07.98	**99.33**	91.04	78.66	03.82
18	91.77	14.86	94.73	08.46	92.12	11.16	95.13	10.28	87.20	84.33	77.54	07.92
19	80.30	19.30	84.42	15.26	83.10	18.40	74.29	21.14	**89.22**	88.10	68.12	12.33
20	67.22	612.33	75.18	466.96	70.15	524.42	64.54	532.16	**87.92**	71.10	67.33	383.16
Mean	83.93	217.98	89.15	177.63	86.63	192.95	84.12	206.10	**90.43**	80.08	78.62	**154.28**
S.D.	10.91	723.90	09.85	597.17	10.01	632.92	11.72	678.69	**07.09**	07.55	08.21	**522.61**
Rank	15		2		3		4		**1**			

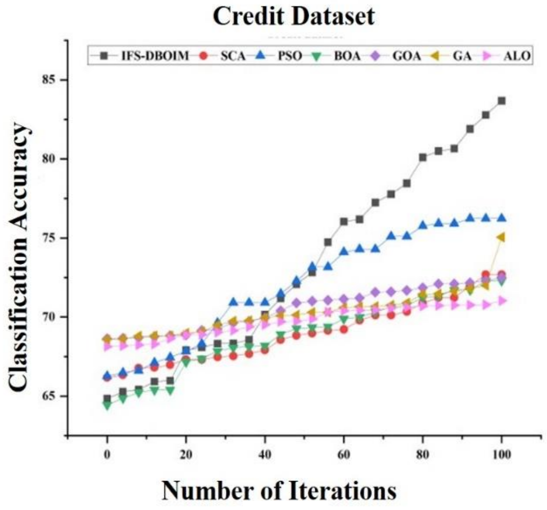
Although the number of selected features is less than the original set, it improved the classification accuracy showing that all features are not required for achieving the best results. The results demonstrate that the IFS-DBOIM realized a maximum feature reduction rate on all the datasets by selecting an average of only 154.28 features. It clearly shows that our approach can effectively determine an optimal set of features to realize the maximum feature reduction rate on all datasets.

7.6.2.3. Sensitivity and Specificity Comparison

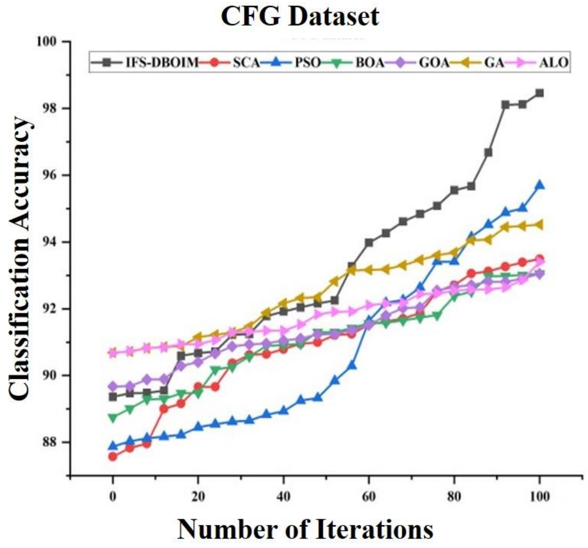
In addition to classification accuracy and the number of selected features, sensitivity, and specificity were also computed to compare the performance of IFS-DBOIM with those from other baseline feature selection methods. From Eq. 15 and 16, it can be deduced that both sensitivity and specificity are only applied to binary classification problems as they are two instance-based ratios. Therefore we have used 13 binary datasets out of 20 to evaluate both performance measures. Figures 7.3 and 7.4 compare all the state-of-the-art methods' sensitivity and specificity scores with the proposed IFS-DBOIM approach. It can be observed that the IFS-DBOIM method realizes a higher score compared to the baseline algorithms on eleven datasets with Diabetic and Sonar as exceptions. Similarly, as shown in Fig. 4, the IFS-DBOIM method again achieves the best specificity score on eleven datasets except for the credit and sonar. Henceforth, the IFS-DBOIM approach can be an effective feature reduction method for binary decision-based medical datasets where a machine learning model must correctly identify a truly infected person. As discussed earlier, due to its global search ability, the PSO algorithm is second-best on four datasets (Australian, Credit, Exactly, Spect), thereby achieving a higher sensitivity score than our method. Two algorithms (ALO and GOA) performed almost equally well and are the second-best approaches that gain a good specificity score after the IFS-DBOIM method on different datasets.


Australian Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
86
84
82
80
78
76
0
20
40
60
80
100
Number of Iterations

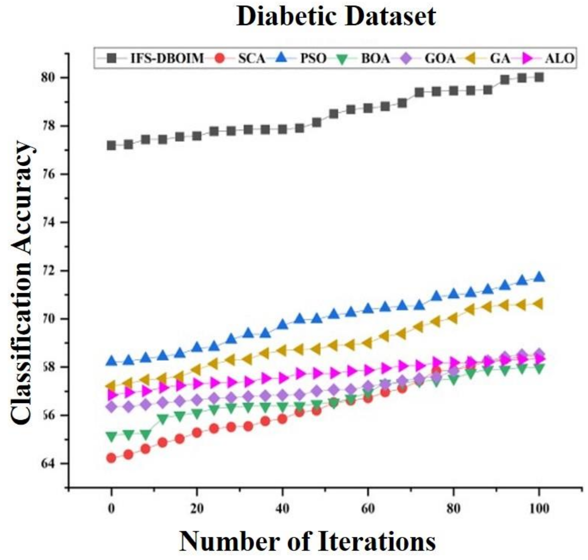
(A)


Credit Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
85
80
75
70
65
0
20
40
60
80
100
Number of Iterations

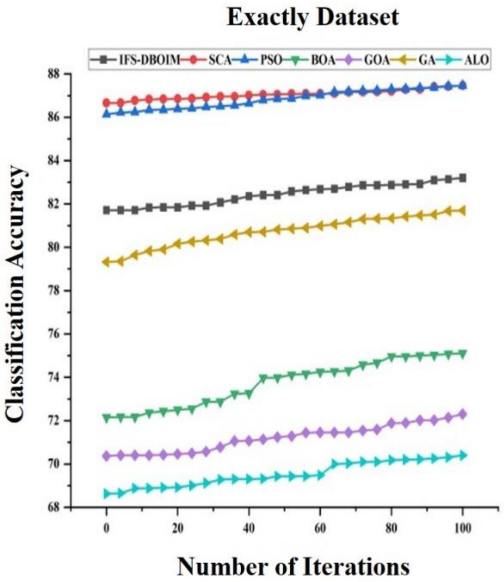
(B)


CFG Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
100
98
96
94
92
90
88
0
20
40
60
80
100
Number of Iterations

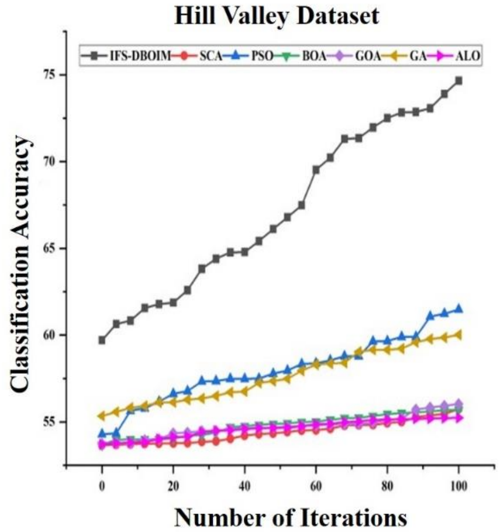
(C)


Diabetic Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
80
78
76
74
72
70
58
56
64
0
20
40
60
80
100
Number of Iterations

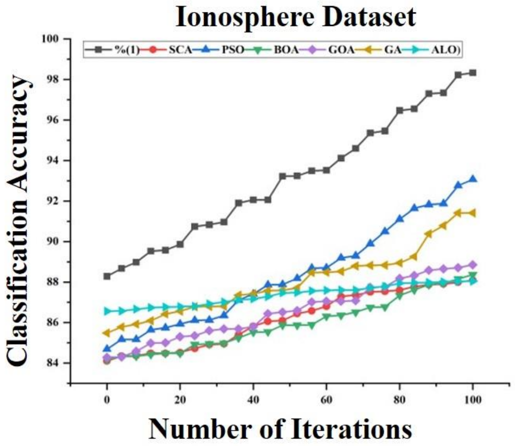
(D)


Exactly Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
Number of Iterations
88
86
84
82
80
78
76
74
72
70
68
0
20
40
60
80
100

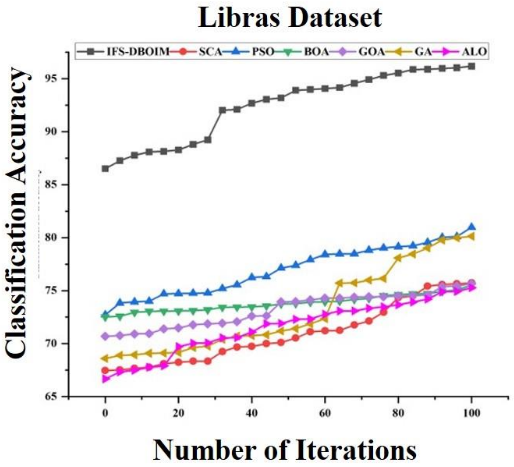
(E)


Hill Valley Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
Number of Iterations
75
70
65
60
55
0
20
40
60
80
100

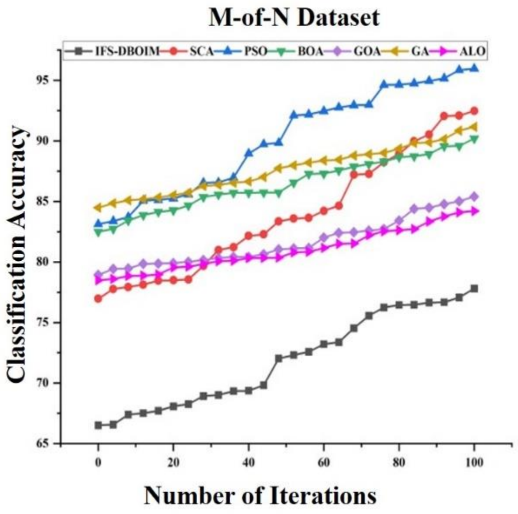
(F)


Ionosphere Dataset
%(1)
SCA
PSO
BOA
GOA
GA
ALO)
Classification Accuracy
Number of Iterations
100
98
96
94
92
90
88
86
84
0
20
40
60
80
100

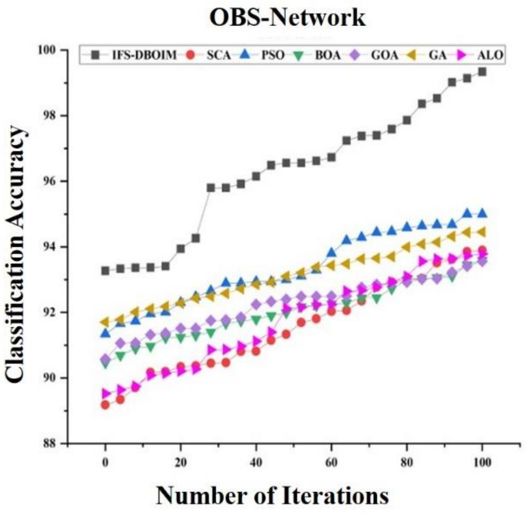
(G)


Libras Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
Number of Iterations
95
90
85
80
75
70
65
0
20
40
60
80
100

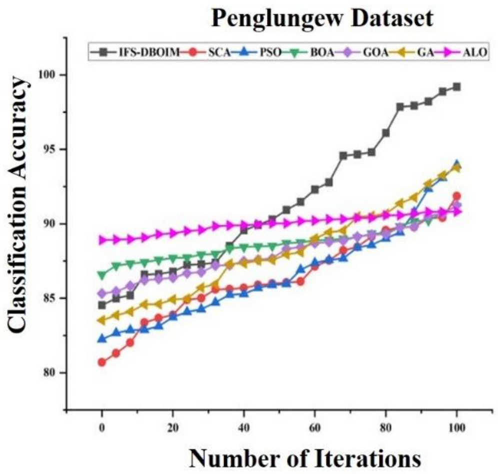
(H)


M-of-N Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
Number of Iterations
65
70
75
80
85
90
95
0
20
40
60
80
100

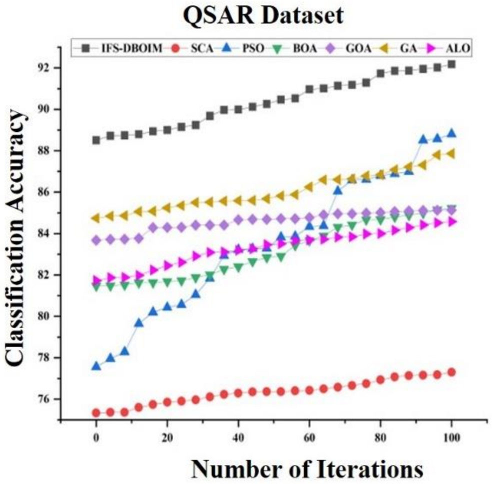
(I)


OBS-Network
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
Number of Iterations
88
90
92
94
96
98
100
0
20
40
60
80
100

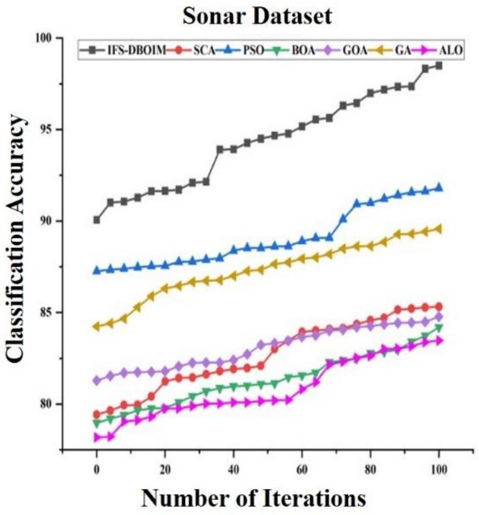
(J)


Penglungew Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
Number of Iterations
80
85
90
95
100
0
20
40
60
80
100

(K)

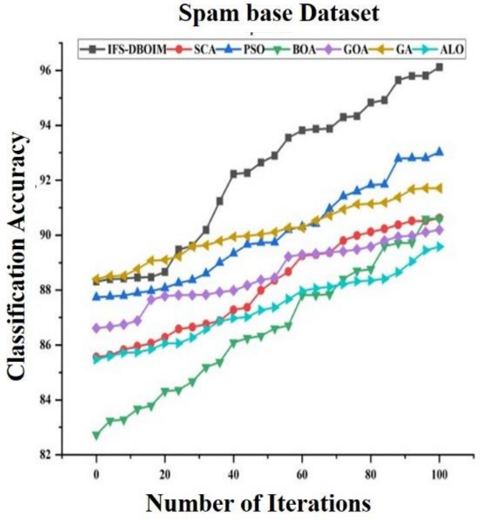

QSAR Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
Number of Iterations
76
78
80
82
84
86
88
90
92
0
20
40
60
80
100

(L)



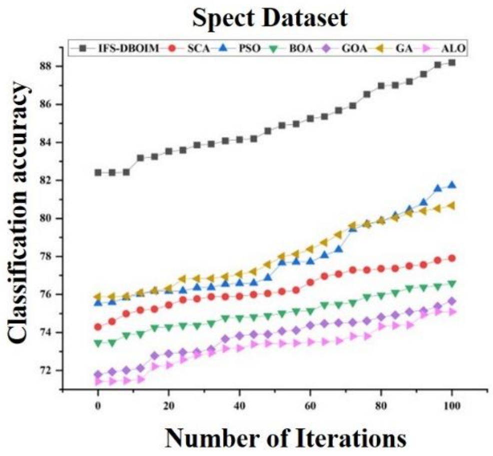
Sonar Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
Number of Iterations
100
95
90
85
80
0
20
40
60
80
100

(M)



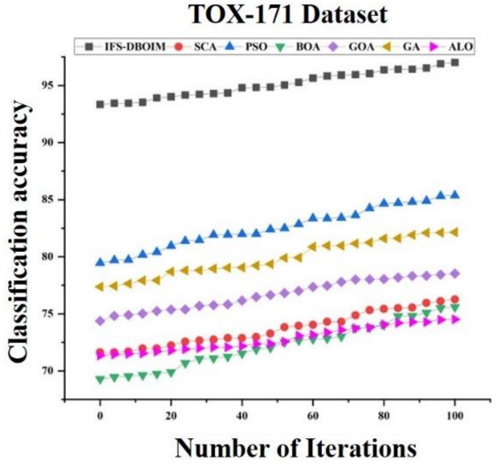
Spam base Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification Accuracy
Number of Iterations
96
94
92
90
88
86
84
82
0
20
40
60
80
100

(N)



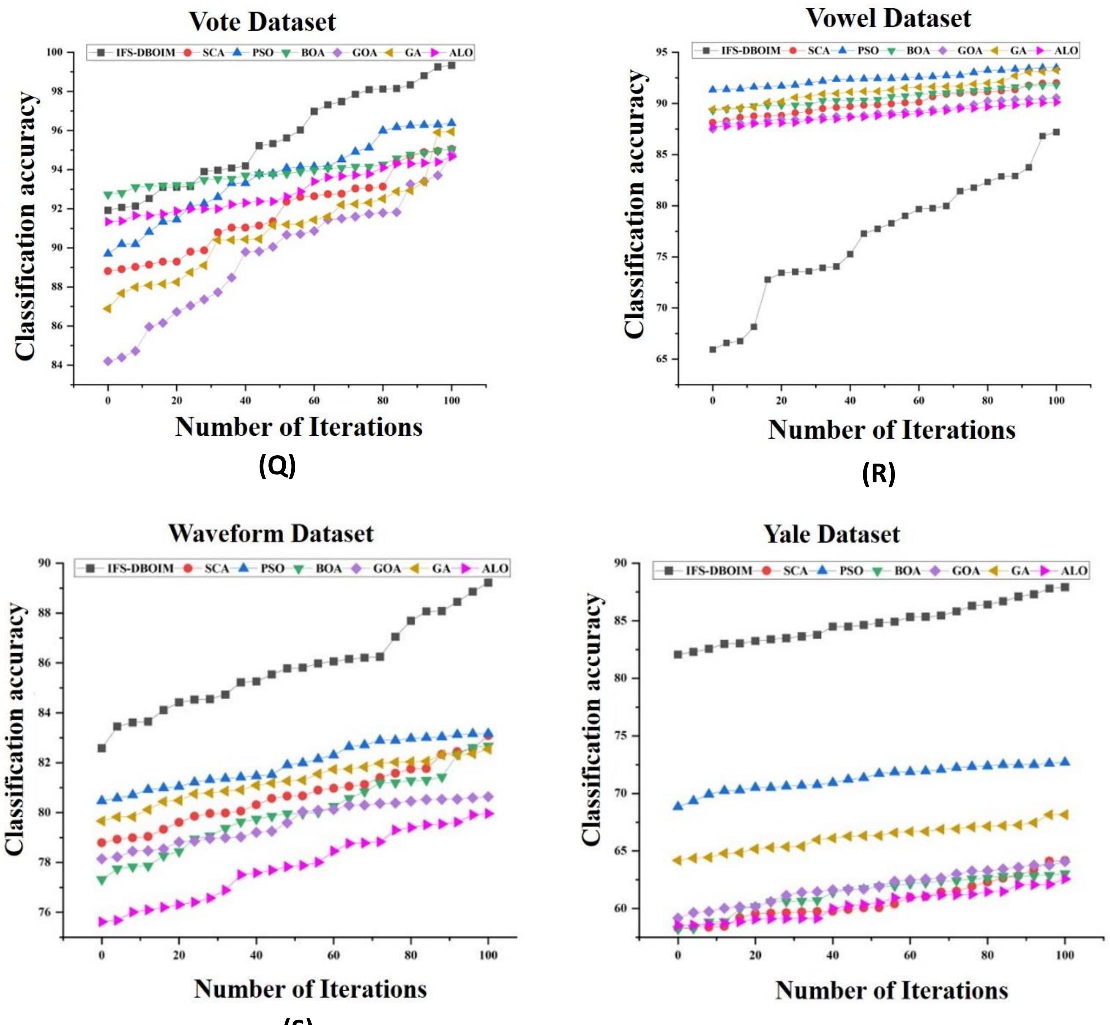
Spect Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification accuracy
Number of Iterations
88
86
84
82
80
78
76
74
72
0
20
40
60
80
100

(O)



TOX-171 Dataset
IFS-DBOIM
SCA
PSO
BOA
GOA
GA
ALO
Classification accuracy
Number of Iterations
95
90
85
80
75
70
0
20
40
60
80
100

(P)



(Q)

(R)

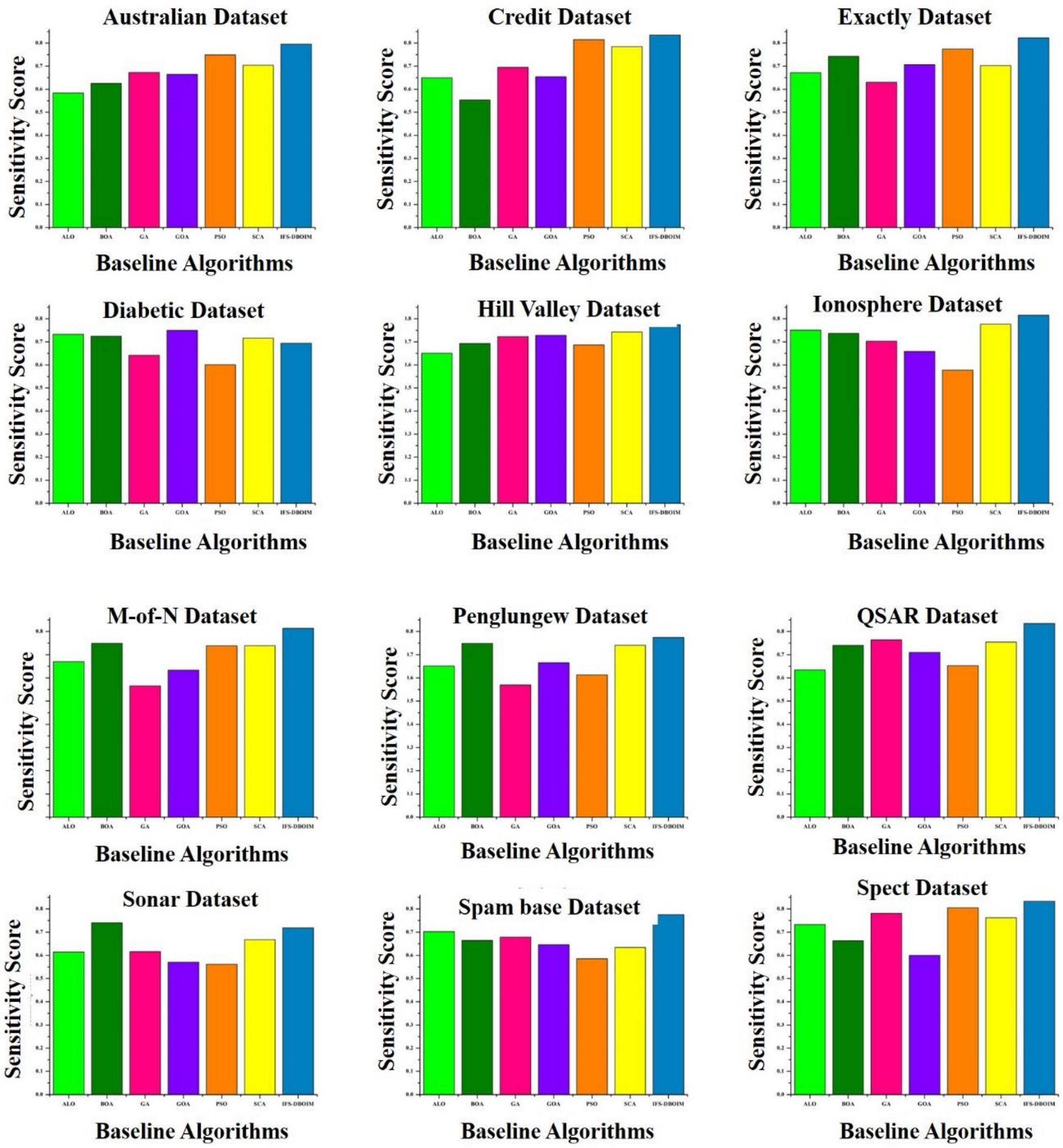
(S)

(T)

Figure 7.2. Illustration of classification accuracies of seven feature selection methods on all twenty datasets.

7.6.2.4. Time Complexity Analysis

Time complexity is another important performance measure that indicates the total time that an algorithm consumes to find the optimal global solution. Table 7.5 presents the average computation times of ten baseline algorithms and the proposed method over 30 independent runs. It can be observed that the IFS-DBOIM realized the lowest computation time on sixteen datasets out of twenty, whereas the PSO outperformed on three and the SCA on a single dataset. It ranks the approaches' performance according to their average computation times. According to the ranking, the IFS-DBOIM realizes the minimum average computation time on all 20 datasets, followed by PSO, SCA, CBOA, BOA, S-bBOA, ALO, OEbBOA, DBOA, GOA, and GA.



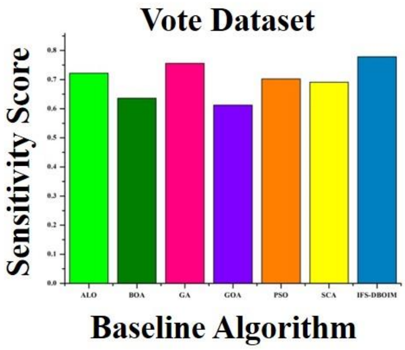
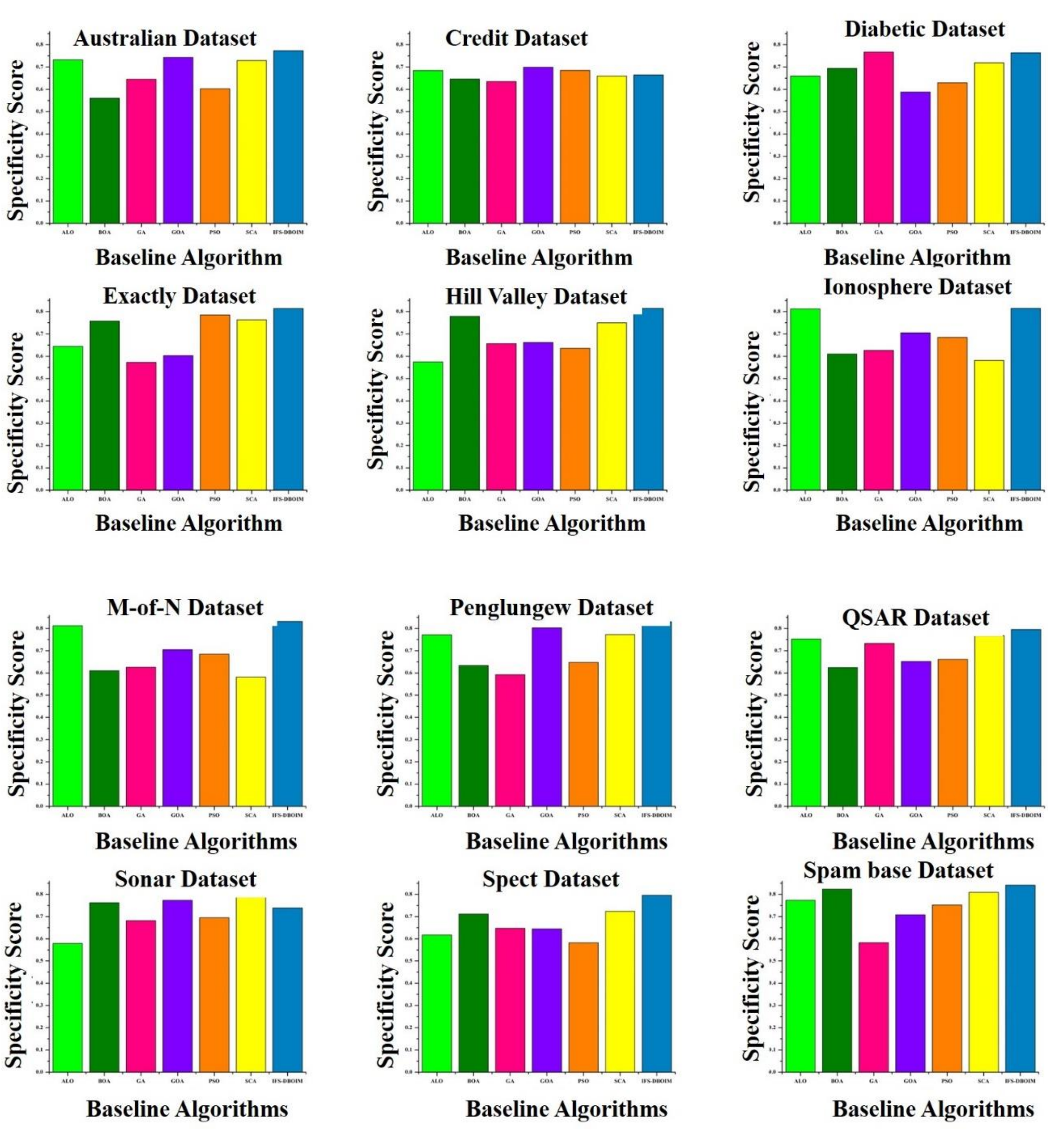


Figure 7.3. Sensitivity score comparison between baseline algorithms and the IFS-DBOIM method



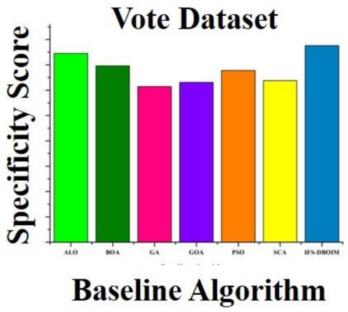


Figure 7.4. Specificity score comparison between baseline algorithms and the IFS-DBOIM method

7.6.2.5. Wilcoxon’s Test Results Analysis

Table 7.6 reports the *p*-values of the IFS-DBOIM in comparison with other baseline feature reduction methods obtained using Wilcoxon’s rank-sum test. This test is conducted to determine whether the difference between the results of the proposed method and other approaches is significant or not. Specifically, if the *p*-value is less than 0.05, the results are considered significant, whereas a greater *p*-value indicates otherwise. It can be observed that in most of the comparisons, the *p*-values obtained using the rank-sum test are less than 0.05, which proves that the effectiveness of our method is statistically significant. Compared with the original DBOA, the *p*-value is less than 0.05 for 14 out of 20 datasets (except Australian, Credit, Exactly, Hill Valley, OBS-Network, and Waveform), which shows the significance of the improvement introduced in our method.

Table 7.5. Details of average computation time of all the state-of-the-art methods and IFS-DBOIM

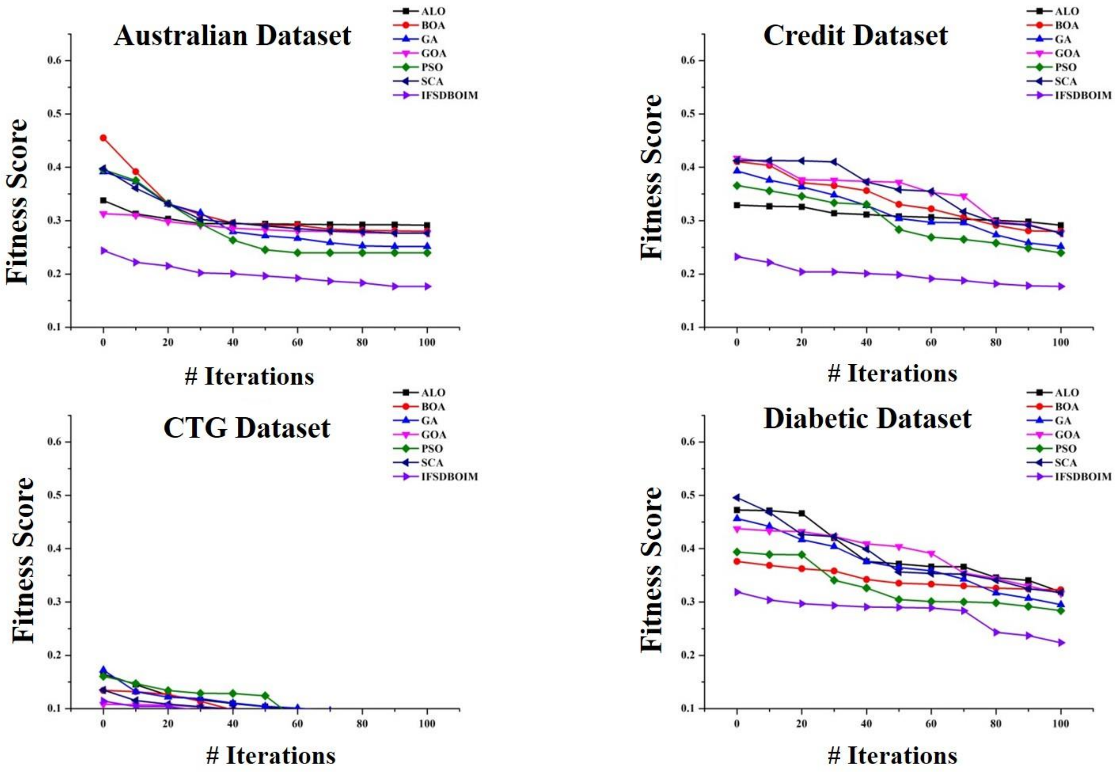
No.	ALO	BOA	GA	GOA	PSO	SCA	CBOA	DBOA	S-bBOA	OEbBOA	IFS-DBOIM
1	17.99	12.82	32.54	08.56	**14.79**	07.98	10.12	14.65	09.26	08.83	06.50
2	14.26	16.72	39.68	13.36	**12.56**	**08.05**	09.01	09.69	14.02	17.61	03.94
3	13.91	13.31	43.53	21.94	**01.61**	09.13	10.51	10.02	08.93	11.94	02.66
4	15.32	18.37	26.27	14.66	**11.59**	12.16	10.63	09.83	14.31	09.00	05.76
5	09.82	10.94	36.16	12.98	**14.35**	07.83	09.14	11.99	10.01	12.42	04.96
6	62.17	68.49	121.56	49.53	28.56	33.74	44.43	27.62	53.79	62.14	26.92
7	08.23	16.19	09.24	12.53	**10.21**	08.48	16.24	08.91	11.37	09.31	06.06
8	13.17	15.95	21.52	24.97	27.57	26.72	21.43	13.89	15.35	19.26	05.21
9	06.81	19.82	16.81	10.12	35.92	35.06	33.40	27.50	24.26	23.50	03.02
10	33.20	28.97	27.10	23.92	28.40	29.27	17.08	20.27	26.60	26.61	12.20
11	84.14	55.13	81.46	91.58	44.56	73.51	88.16	106.87	99.75	104.93	38.71
12	44.92	30.49	48.31	46.81	**14.98**	35.10	27.25	45.96	48.34	33.14	19.43
13	25.11	71.97	58.80	70.80	50.26	52.35	49.98	20.24	20.24	20.24	09.18
14	41.44	29.80	35.27	56.30	33.63	37.80	60.43	59.41	48.54	57.41	24.83
15	63.76	39.36	37.56	78.56	35.44	**17.21**	26.96	34.28	32.84	28.50	22.19
16	71.49	49.28	53.07	47.19	38.14	56.24	36.88	95.40	41.62	45.71	32.50
17	34.49	33.79	43.00	71.08	**19.68**	28.76	47.43	59.09	37.27	69.83	23.74
18	37.0	62.71	84.17	83.52	39.30	40.41	90.78	76.64	34.75	47.78	14.90
19	73.66	63.75	100.9	48.95	62.53	89.42	69.86	48.15	89.097	73.01	18.49
20	103	87.14	94.07	43.33	61.38	78.14	64.59	117.57	114.80	108.33	53.27
Mean	38.	37.25	50.56	41.53	29.27	34.37	37.22	40.90	37.76	39.47	16.72
Rank	7	5	11	10	2	3	4	9	6	8	1

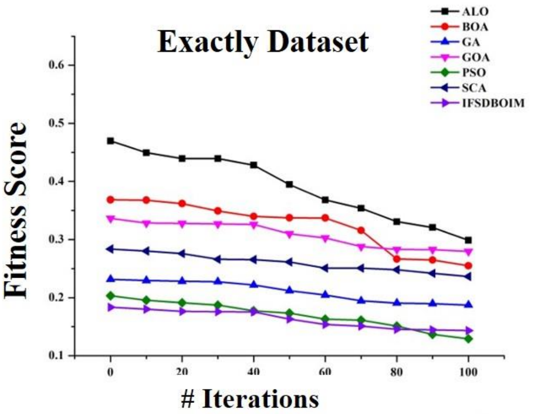
Table 7.6. The *p*-value based on the Wilcoxon test of all the algorithms in terms of average classification accuracy over 30 runs ($p \geq \mathbf{0.05}$) is marked with bold color.

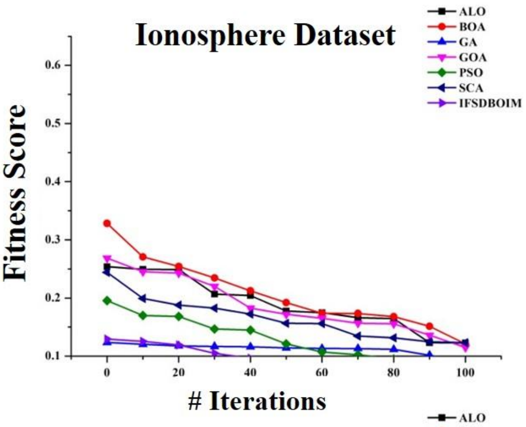
No.	ALO	BOA	GA	GOA	PSO	SCA	CBOA	DBOA	S-bBOA	OEbBOA
1	1.68E-09	2.53E-09	1.08E-05	5.41E-09	**7.25E- 04**	7.70E-09	2.0E-04	**1.0E-00**	1.8E-04	2.98E-06
2	6.61E-11	1.94E-10	3.41E-06	2.98E-10	**1.62E-02**	3.42E-10	**3.01E-03**	**2.05E-01**	3.4E-07	**2.8E-03**
3	7.30E-11	3.65E-11	3.13E-10	2.98E-11	2.09E-07	3.65E-11	**5.07E-02**	1.70E-05	3.7E-09	1.3E-04
4	1.97E-11	1.90E-11	5.41E-10	6.15E-11	1.28E-05	1.89E-11	1.37E-07	**3.10E-02**	6.2E-11	4.2E-07
5	2.95E-08	1.10E-08	1.72E-03	7.88E-07	**3.64E-02**	3.49E-08	**3.7E-04**	2.41E-07	6.03E-05	0.03E-01
6	1.07E-09	3.82E-09	1.86E-03	1.70E-08	**2.15E-02**	8.47E-09	9.1E-11	**3.8E-04**	13.1E-08	2.07E-09
7	1.61E-10	5.31E-10	1.866E-06	4.83E-10	**3.76E-03**	9.90E-11	4.7E-09	7.3E-05	**6.18E-03**	6.01E-06
8	2.02E-07	5.38E-07	2.15E-03	4.10E-07	**5.10E-02**	1.15E-06	5.21E-06	6.1E-11	9.31E-10	1.7E-11
9	1.99E-11	1.89E-11	2.97E-10	2.10E-11	**7.25E-04**	1.17E-10	0.7E-06	4.07E-07	**1.13E-03**	3.3E-05
10	2.52E-08	3.09E-09	4.15E-08	4.03E-09	1.80E-06	4.41E-09	4.0E-07	**1.8E- 01**	**2.4E-02**	2.0E-04
11	4.41E-06	3.32E-06	1.14E-03	4.42E-06	**2.19E-02**	1.43E-05	**3.0E-03**	1.1E-05	3.39E-07	3.8E-06
12	1.09E-10	2.73E-10	7.87E-05	7.34E-10	**4.52E-03**	2.60E-10	2.0E-11	1.0E--06	7-4E-09	**3.1E-03**
13	2.86E-10	1.77E-10	1.10E-06	1.68E-10	**6.19E-04**	4.60E-10	1.06E-14	4.30E-05	2.4E-13	0.3E-09
14	3.02E-11	3.34E-11	1.55E-09	4.97E-11	1.07E-04	4.50E-11	2.03E-17	2.02E-05	1.3E-08	3.4E-13
15	3.46E-10	3.63E-10	3.37E-07	1.32E-10	**3.36E-04**	1.20E-08	1.91E-10	1.16E-17	3.4E-14	2.3E-12
16	6,12E-10	6.12E-10	2.00E-06	1.85E-08	**3.58E-03**	5.97E-09	3.21E-15	6.31E-08	5.1E-11	1.6E-08
17	3.89E-07	2.16E-07	1.66E-04	1.92E-07	**4.67E-04**	2.08E-07	3.14E-15	0.19E-13	2.0E-09	1.3E-07
18	9.19E-10	1.72E-06	1.71E-03	3.06E-07	**4.82E-02**	1.86E-05	4.01E-11	3.14E-11	1.9E-16	6.6E-13
19	3.01E-11	1.87E-07	1.43E-06	3.01E-11	1.40E-04	1.49E-06	4.07E-07	**1.13E-03**	3.3E-05	3.14E-11
20	7.12E-09	1.43E-08	6.36E-05	7.09E-08	**1.71E-01**	7.70E-08	1.18E-10	4.59E-09	3.37E-13	1.49E-09

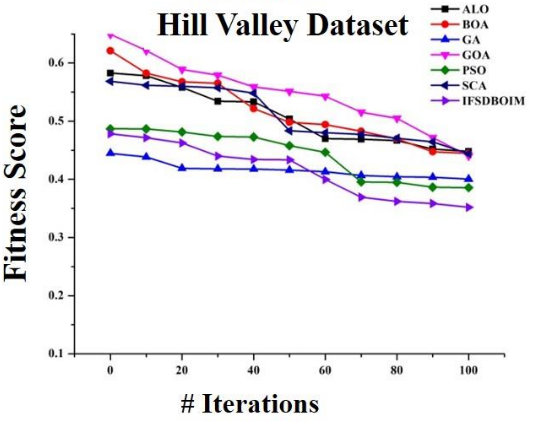
7.6.2.6. Convergence Analysis

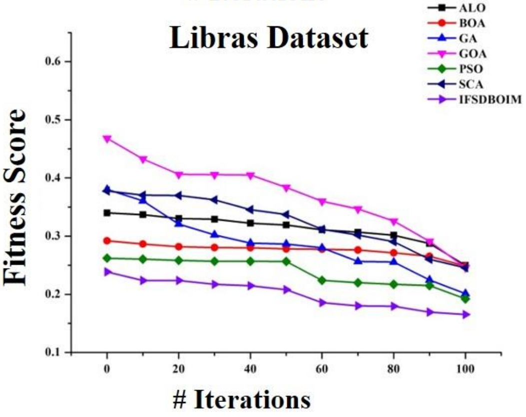
The convergence rate is an important performance measure that represents how quickly an algorithm approaches its solution. In Figure 7.5, convergence curves of all the competing algorithms were plotted between fitness scores and iteration counts to show their behavior on all the datasets. After contemplating Figure 7.5, two main comparison perspectives arise. First, in the early iterations, IFS-DBOIM achieved lower fitness scores on all the datasets except for the three (Exactly, M-of-N, and Vowel) even though it did not converge like state-of-the-art methods. This might be attributed to the executed initialization method, population size, and the search strategy of the different algorithms. Generally speaking, the convergence rate of the original DBOA is slower than other baseline algorithms, which explains a similar low execution rate of the derived IFS-DBOIM. In addition, it is assumed that the involvement of the mutual information-based FIM further slows down the convergence rate to better the fitness score in our method. This highlights the significance of the employed local search strategy (LSAM) to find the possible global optima or a closer solution.

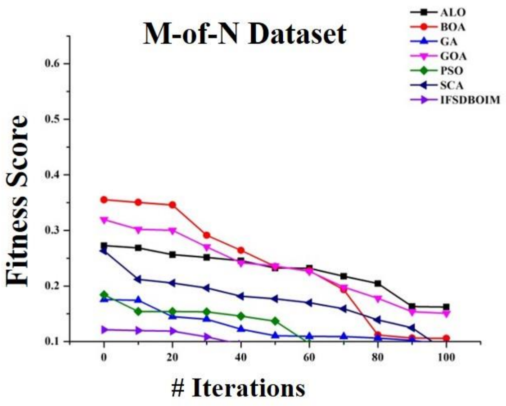

Australian Dataset
Credit Dataset
CTG Dataset
Diabetic Dataset
Fitness Score
Iterations
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100

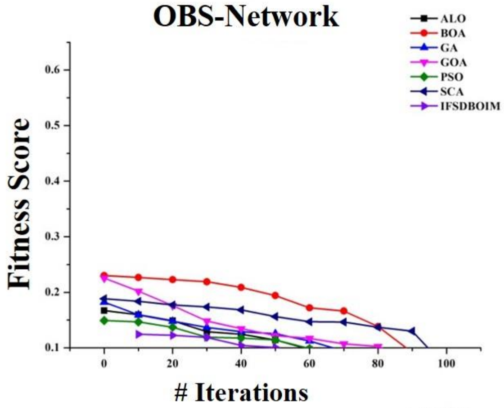

Exactly Dataset
Fitness Score
Iterations
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM

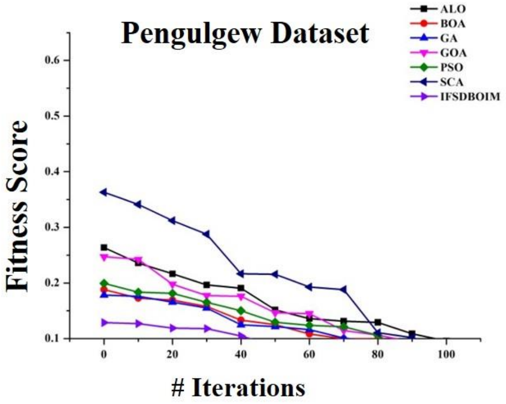

Ionosphere Dataset
Fitness Score
Iterations
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM

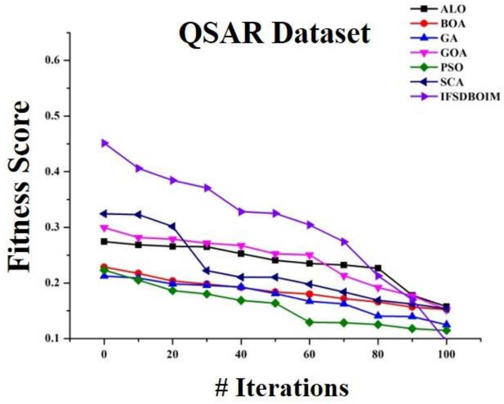

Hill Valley Dataset
Fitness Score
Iterations
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM

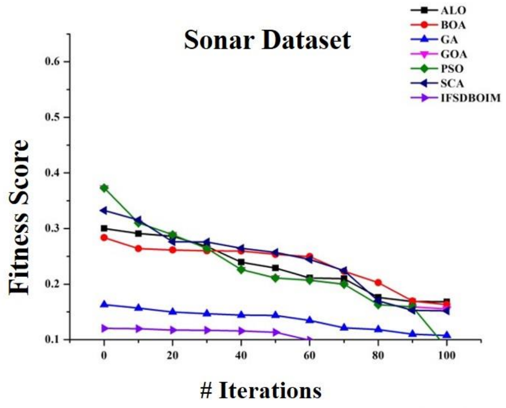

Libras Dataset
Fitness Score
Iterations
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM

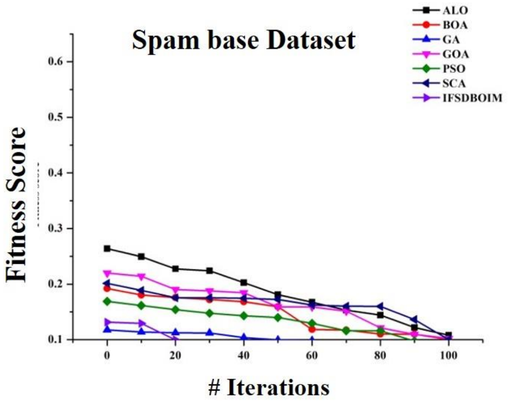

M-of-N Dataset
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM
Fitness Score
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
Iterations

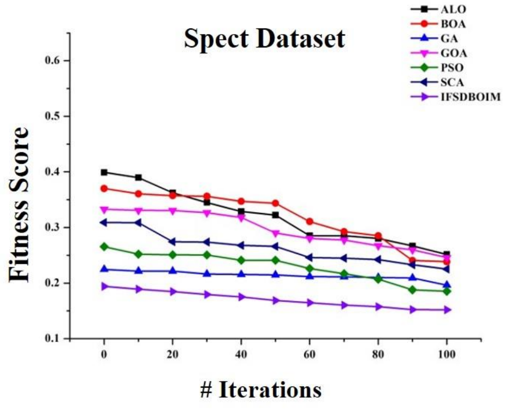

OBS-Network
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM
Fitness Score
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
Iterations

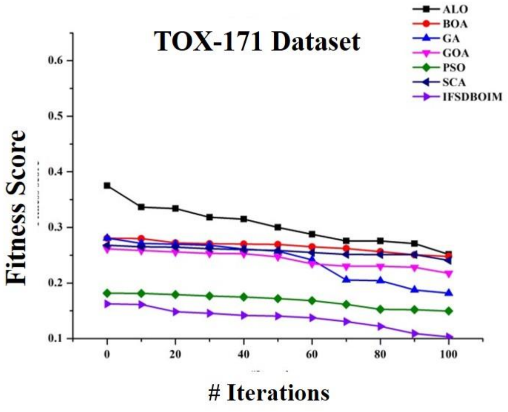

Pengulgew Dataset
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM
Fitness Score
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
Iterations


QSAR Dataset
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM
Fitness Score
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
Iterations


Sonar Dataset
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM
Fitness Score
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
Iterations


Spam base Dataset
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM
Fitness Score
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
Iterations


Spect Dataset
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM
Fitness Score
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
Iterations


TOX-171 Dataset
ALO
BOA
GA
GOA
PSO
SCA
IFSDBOIM
Fitness Score
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
Iterations

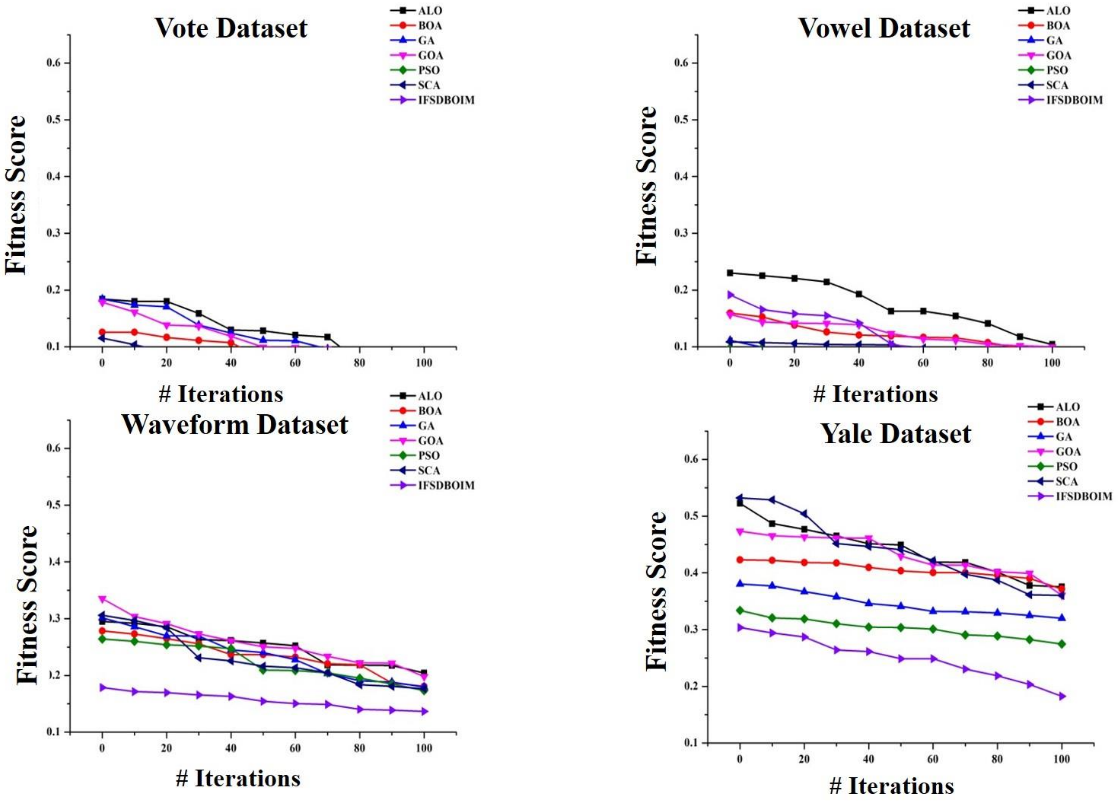


Figure 7.5. Comparison of convergence curves of all competing algorithms over the 20 datasets.

7.6.2.7. Implications of the Study

Search space optimization is an indispensable part of machine learning that aims to locate plausible input points with acceptable solution quality. In this research work, we developed a hybrid feature selection model, namely, Iterative Feature Selection using Dynamic Butterfly Optimization-based Interaction Maximization (IFS-DBOIM) algorithm to eliminate irrelevant and redundant data points from high-dimensional space. Although baseline metaheuristic algorithms (ALO, BOA, GA, GOA, PSO, and SCA) realize good classification accuracy when used for solving dimension reduction, still most of them suffer from two major problems: (1) poor ability to escape from local optima, which lead to premature convergence problems, and (2) low solution diversity.

The DBOA effectively resolves both problems by introducing a Local Search Algorithm based on Mutation (LSAM) that works in two ways: (1) improving the quality of the current best solution by introducing mutations that further help to solve local optima problems; (2) random solutions selection with each iteration to enhance the diversity of the final set. In addition, baseline algorithms can only monitor the newly selected feature's relevance and redundancy level. At the same time, the IFS-DBOIM rigorously considers the interaction of the features with the previously selected ones. If the new feature is important and non-redundant, it must share minimum mutual information with the already selected features. Henceforth, the final feature subset becomes more relevant but with fewer features. Since the performance of the IFS-DBOIM is outstanding on high-dimensional datasets, it indicates that its local search strategy is better than that of the remaining algorithms. Therefore, it can be effectively used to solve similar research problems such as EEG-based Channel Selection (CS) in the Brain-Computer Interface (BCI) modeling [1], design parameter selection in engineering models [146], and gene clustering [147].

7.7. Limitations of the Study

In general, the proposed feature selection method uses two selection criteria: relevancy and redundancy. This method maintains a good balance between relevancy maximization and redundancy minimization while designing the optimal feature subset. Despite realizing higher classification accuracy rates on multiple datasets using the IFS-DBOIM approach, the proposed method has some limitations that need to be addressed.

The proposed approach calculates the redundancy in terms of sharable information between the previously selected features and the new ones without considering the class label. The feature may share information, but it does not imply that they are redundant; they may share valuable information with the class attribute. In addition, measuring mutual information from finite data is difficult and non-reliable to determine the redundancy of the features. Therefore, this criterion may not compute the absolute non-redundant set of features for achieving better classification accuracy. Another problem that all metaheuristic-based feature selection methods share is their non-deterministic nature to determine the optimal feature subset. In contrast with exact algorithms whose final result is always fixed, metaheuristics do not provide that kind of bound. They can be very effective on a given instance of a problem but may provide the worst results on another, highlighting that the proposed method is problem-specific and hard to generalize. In practice, the significance of each of the above-mentioned problems depends on the data and properties of each dataset.

7.8. Conclusion & Future Scope

This paper proposes a hybrid feature reduction algorithm, Iterative Feature Selection using Dynamic Butterfly Optimization-based Interaction Maximization (IFS-DBOIM). It employs three objective functions to evaluate the fitness of each solution. Firstly, the DBOA is used as a supervised learning method to generate a set of solutions. Then, an information theory-based three-way interaction mechanism is adopted to extract the best optimal feature subset with maximum classification accuracy and reduced model complexity. In the experiment section, twenty standard datasets from the UCI repository are used to assess the performance of the IFS-DBOIM approach. Experimental results showed that the proposed method produced more promising results than the earlier ones on most datasets, especially those with high dimensionality but an extremely small sample size.

The simulation results confirmed that the proposed method achieved better classification accuracy (90. 43%) and feature reduction rate than the other baseline algorithms. The statistical significance of the reported results is further confirmed by a paired Wilcoxon rank-sum test. Moreover, the proposed method balances better the trade-off between classification accuracy and stability. The performance of the IFS-DBOIM is also compared in terms of classification accuracy, feature reduction rate, fitness score, sensitivity, specificity, time complexity, and convergence rate with other popular methods. The results indicate the superiority of the proposed hybrid method with few limitations. Since it is a non-deterministic algorithm, it suffers from a lack of generalization and relies on the characteristics of applied datasets.

Dimensionality is one of the crucial factors that directly affect the performance of expert and intelligent systems. Compared with the research in the related field, our method realized higher comprehensive performance as its key contributions include: (1) the application of information theory and (2) the maintenance of a balance between the exploration and exploitation phase in DBOA for solving complex feature reduction problems. Therefore, the IFS-DBOIM method can be used in preprocessing to find the minimum but relevant features for improving the classification accuracy.

In the future, the IFS-DBOIM may be merged with other recently developed optimization strategies such as the EarthWorm Optimization Algorithm [148], Elephant Herding Optimization [149], and Slime Mould Algorithm [150] to find a more robust feature subset. To avoid the overfitting issue in the IFS-DBOIM method, a pruning scheme [151] can be incorporated to develop a new parsimonious classification model.