

## RESEARCH ARTICLE

# Population Initialization Factor in Binary Multi-Objective Grey Wolf Optimization for Features Selection

NUR LYANA SHAHQIQA ALBASHAH<sup>1b</sup> AND HELMI MD RAIS<sup>1b</sup>

Institute of Health and Analytics, Universiti Teknologi Petronas, Seri Iskandar 32160, Malaysia

Corresponding author: Nur Lyana Shahfiqa Albashah (yanashahfiqa88@gmail.com)

The research is financially supported by Yayasan Universiti Teknologi Petronas (YUTP) grant.

**ABSTRACT** Features selection methods not only reduce the dimensionality, but also improve significantly the classification results. In this study, the effect of the initialization population using the population factor has been explored. There are twenty wolves obtained by the population initialization method in binary multi-objective grey wolf optimization for features selection. There are two objectives function that will be minimized i.e. number of features and error rate. The proposed method has been compared with the previous study Binary Multi-Objective Grey Wolf Optimization (BMOGWO-S) using UCI datasets, oil and gas datasets. The results reflect that the proposed method outperforms all existence methods in terms of reducing feature numbers and error rates.

**INDEX TERMS** Grey wolf optimizer, features selection, multi-objective, optimization, classification.

## I. INTRODUCTION

Swarm intelligence is one of the topics that have excessive impact on the optimization area, because of the capability to solve many optimization problems in a variety of fields such as medicine, engineering, and others. It helps to escape from the local optima area that is usually present in optimization problems, and the algorithm is simple and efficient. Most algorithms are based on natural processes such as hunting strategy, shown in the grey wolf optimization algorithm. This algorithm was founded by Seyedi Mirjalili [1]. Seyedi Mirjalili also proposed several extension algorithms, for example, solving a multi-objective optimization problem. Based on his research, many other research works have been discovered to improve and explore the use of various types of data to solve a specific problem. For example, the features selection problem with Grey Wolf optimization focuses on how to decrease the features while improving classification performance. In this area, problems can be divided into two categories which are for single-objective problems and multiple-objective problems. A single objective problem only focuses on solving one

specific problem. For example, to decrease the error rate of the classification problems. While, multi-objective problems focused on two or more than two objectives, for example, to minimize the number of features and the error rate.

In this study, multi-objective optimization problems for features selection and classification using Grey Wolf Optimizer (GWO) by introducing an initialization population using the population factor have been focused. Furthermore, how initialization parameter affects searching the Pareto front in different datasets with varying numbers of features.

The grey wolf optimization process consists of exploration and exploitation. The exploration process is a condition when the algorithm is able to find other directions from the previous searches path, specifically detecting prey. While, the exploitation process relates to find the optimal solution in more detail on the original track. When the step of exploration is too small tends to cause convergence in local optima and the algorithm will converge too slowly with insufficient exploitation which also happens in GWO. Other disadvantages of GWO are lack of population diversity and imbalance in exploration and exploitation. There are various studies to overcome this problem, one of them is to introducing a new

The associate editor coordinating the review of this manuscript and approving it for publication was Sabah Mohammed<sup>1b</sup>.

hunting search mechanism using dimension learning-based hunting by Nadimi-shahraki et al. [2]. GWO can obtain diversity in its searching population by randomly mutating some of the solutions in the population. This can assist the algorithm in discovering new regions of the search space that it might not have discovered yet [3]. Additionally, GWO can also use a crossover to combine solutions from different parts of the search space, which can also help to broaden the diversity of the search population [4]. Finally, GWO can also use to choose the specific solutions from the search population which leads to a successful solution. Combine, these various methods help GWO to explore different areas of the search space and find solutions that are more likely lead to a successful solution. Overall, the diversity of the search population is important in GWO because it helps to ensure that the algorithm is exploring all possible solutions. By randomly mutating some of the solutions in the search population using crossover to select specific solutions, GWO can broaden its search space and find more likely solutions.

The dataset used in this research is based on thirteen benchmark datasets and one dataset from the oil and gas area. In this paper, we investigated the following objectives:

1. To propose the initialization population obtaining using population factor with binary multi-objective grey wolf optimization.
2. To compare the proposed method against the previous method Binary Multi-Objective Grey Wolf Optimization (BMOGWO-S), and indicate which approach gives better performance in minimizing error rate and features number.

The structure of the paper, followed by the GWO concept, features selection study areas located in section II. The proposed method will be explained in section III. In section IV, the experimental design will be illustrated and described in detail. The result and discussion will be elaborate in section V, and the final section contains a conclusion and future work

## II. LITERATURE REVIEW

Swarm intelligence is related to nature, it behaves like an animal to solve their problems for them to survive. They work together to find the solution and the crucial part is getting the optimal solution to fulfill the constraints. There are various swarm models, for example, obtained from a colony of bees, fish schooling, bird flocking, ant foraging, and grey wolf hunting mechanism. All these models are applied in various fields, especially in solving real-time problem, such as computers and robotics. This is essential in the development of high technology equipment and facilities. Grey Wolf Optimizer (GWO) algorithm has been used as a features selection technique in various domains, including face expression identification, EMG signal categorization, disease diagnosis, gene selection, and intrusion detection systems, as evidenced by the literature [5]. Here, the concept of grey wolf optimization will explain in term of applications i.e. how it works in solving multi-objective optimization. Following that, a literature review on feature selections is

explained to provide a better understanding of the various methods discovered and proposed in the literature related to current research study.

### A. GREY WOLF OPTIMIZER (GWO)

Like a grey wolf, they are living in a group. Grey wolf optimizers (GWO) have a hierarchy known as alpha, beta, delta, and omega. Each of them has its responsibility. Like alpha is a leader of the group. While beta and delta are being assisted by alpha wolves. The omega wolf is a lower hierarchy of the group. Usually, they will eat last and be bullied by another grey wolf member. In mathematical concepts, the four wolves represent as below:

- The alpha wolf will obtain the best solution.
- The beta wolf gives the second-best solution.
- The delta wolf gave the third-best solution.
- The omega wolf gave all other solutions.

The hunting mechanism contains three important steps, which are encircling, hunting, and attacking. Below are the mathematical equations representing on encircling mechanism:

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) \right|, \quad (1)$$

$$\vec{X}(t+1) = \vec{X}_p(t) + \vec{A} \cdot \vec{D}, \quad (2)$$

where,  $X_p$  is the position of the prey and  $t$  is the iteration numbers. Here,  $\vec{A}$  and  $\vec{D}$  are the coefficient vectors. Below is the formulation for both coefficient vectors:

$$\vec{A} = 2 \cdot \vec{a} \cdot \vec{r}_1 - \vec{a}, \quad (3)$$

$$\vec{C} = 2 \cdot \vec{r}_2, \quad (4)$$

where,  $\vec{r}_1$  and  $\vec{r}_2$  are random number vectors ranging from 0 to 1 and value  $\vec{a}$  has a value between 2 and 0. For the hunting mechanism, the position of three-wolf alpha, beta, and delta is formulated as below:

$$\vec{D}_\alpha = \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right|, \quad (5)$$

$$\vec{D}_\beta = \left| \vec{C}_1 \cdot \vec{X}_\beta - \vec{X} \right|, \quad (6)$$

$$\vec{D}_\delta = \left| \vec{C}_1 \cdot \vec{X}_\delta - \vec{X} \right|, \quad (7)$$

$$\vec{X}_1 = \vec{X}_\alpha \cdot A_1(\vec{D}_\alpha), \quad (8)$$

$$\vec{X}_2 = \vec{X}_\beta \cdot A_2(\vec{D}_\beta), \quad (9)$$

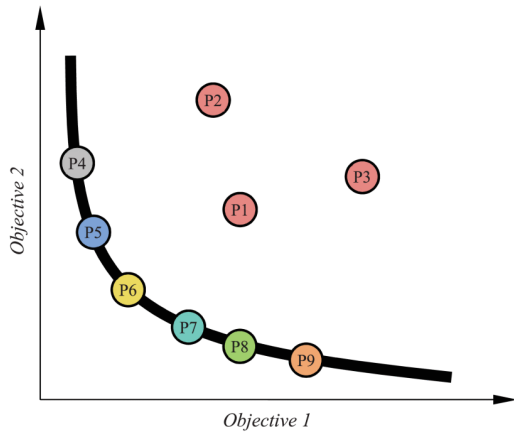
$$\vec{X}_3 = \vec{X}_\delta \cdot A_3(\vec{D}_\delta), \quad (10)$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3}. \quad (11)$$

The last step is the attacking mechanism, where the vector  $\vec{a}$  has a value between 0 and 2, and the value decreases linearly for each iteration. The formula for  $\vec{a}$  is shown below:

$$\vec{a} = 2 - \frac{t \cdot 2}{maxiter}, \quad (12)$$

where  $maxiter$  is the maximum iteration value, and  $t$  is the value of iteration.



**FIGURE 1.** Pareto front for two objective functions with minimization problem.

### B. MULTI-OBJECTIVE OPTIMIZATION

Multi-objective problems occur when there are two or more objective functions that need to be optimized according to the decision-maker. For example, for minimizing and maximizing all objectives function, or in some cases, the objective needs to minimize and maximize the objective's functions. For this study, we would like to minimize the two objectives function, and the mathematical formula can be shown below:

$$\text{Minimize } F(y) = [f_1(y), f_2(y)], \quad (13)$$

$$\text{Subject to } g_k(y) > 0, \quad i = 1, 2, \dots, m, \quad (14)$$

$$h_k(y) = 0, \quad i = 1, 2, 3 \dots l, \quad (15)$$

where the value of  $g_k(y)$  and  $h_k(y)$  is represented by constrained for the objective function,  $k$  is the number of the objective function, and the value is greater than 1. In this study,  $k$  is equal to 2. Meanwhile,  $f_1$  and  $f_2$  refer to functions that we want to optimize. For example, here, we have the function of error rate and the number of features. The  $i$ -objective minimization problem, for example, has two solutions, which we call  $c$  and  $d$ . If the following requirements are met, it is safe to believe that  $c$  has dominated over  $d$  or that  $d$  has dominated over  $c$ . Figure 1 shows an example of a minimization problem with two objectives to understand better what might be regarded as an optimal solution for a particular optimization problem: Both P2 and P3 were dominated by P1, P2, and P3 cannot, however, overpower each other. This is important because it provides the optimal solutions for the whole process. To keep the non-dominated solution during the search and to modify it, when a new, superior solution is found, an external achieve is added to MOGWO [6]. A Pareto optimum solution is not governed by any other solution. Figure 1 shows the Pareto optimum as a black line containing P4, P5, P6, P7, P8, and P9. The equilibrium has been reached at this point. This is understandable if you want to reduce objective 2 in point P6. For example, we need to sacrifice by maximizing objective 1 to get into points P7, P8, or P9. In this situation, we cannot reduce both at the same time.

### C. FEATURE SELECTION

By definition, feature selection is a competing minimization problem with two objectives i.e. minimizing the number of features and the classification error rate. However, the competing relationship between these two purposes in the context of feature selection implies that the two objectives are not equally vital for practitioners, and non-dominated solutions with a smaller number of features are chosen for this multi-objective problem.

The features are obtained from the data experiment. Typically, prior to performing any analysis, much information about features is gathered during an experiment. Because conducting experiments necessitates a significant amount of time and effort. Most of the time, the researcher will gather as many features as possible in order to produce a high-quality result. Unfortunately, all the features are not excellent, and this will be revealed in detail during the analysis phase. Among practitioners, feature selection is a widely used strategy for reducing the dimensionality. It seeks to select a small subset of relevant features from the original set based on a relevance assessment criterion, which usually results in improved learning performance (e.g., higher classification learning accuracy), cheaper processing costs, and better model interpretability. Feature selection processes can be categorized as a filter, wrapper, and embedded models. The filter model separates feature selection from classifier learning, ensuring that a learning computation bias does not conflict with a feature selection algorithm's bias. It is based on distance, consistency, dependency, information, and correlation as generic attributes of the training data. In comparison, the wrapper model is designed to decouple search algorithms from the intrinsic statistical characteristics of the datasets, thereby paying more attention to searchability. The search process can be classified as a traditional or metaheuristic approach. The traditional search process further contains sequential forward and sequential backward processes where sequential forward is a process of adding several features one by one to the model, while sequential backwards is removing one by one feature from the model. It is less efficient, easily traps local optima, and has scalability issues when dealing with high-dimensional datasets. Many studies, such as [7], [8], and [9], demonstrated that metaheuristic outperform the traditional method. The embedded model has overcome the limitation of the filter and wrapper model by executing it before creating the classifier.

A variety of swarm intelligent algorithms have been used in research studies to support and enhance the present approach, providing a wide range of options for researchers across all disciplines to choose from and a better outcome for their particular research topic. Table 1 gives an overview of the metaheuristics research study. Abutarboush et al. [10] provide a comprehensive research study from 2009 to 2019. Based on this summary, we can conclude that most algorithms perform better when implemented in binary form. There are various binary transform techniques such as V-shaped transfer function, sigmoid function, binary binomial, and

**TABLE 1. Summary of literature review.**

METAHEURISTICS ALGORITHM	AUTHOR	YEAR	DATASET	CONTRIBUTION
Cuckoo (CS)	Tiwari [10]	2012	Face Recognition	Application on face recognition using CS
	Rodrigues et al.[12]	2013	Theft detection and industrial dataset	Proposed binary version of the cuckoo search
	Gunavathi and Premalatha [13]	2015	Microarray data	Applied cuckoo search algorithm into the cancer research study
	Salesi and Cosma [14]	2017	Biomedical datasets	Proposed pseudobinary mutation neighbourhood search in BCS
	Pandey et al. [15]	2020	UCI datasets	Proposed binary binomial cuckoo search algorithm
Bats (BA)	Nakamura et al [16]	2012	UCI datasets	Development of a binary version of BA (BBA) with a sigmoid function
	Laamari and Kamel [17]	2014	UCI datasets	Introduced V-shaped transfer function with SVM
	Rodrigues et al. [18]	2014	UCI datasets	Development of hyperbolic tangent transfer function for binary version of BA.
	Enache and Sgârciu [19]	2015	UCI datasets	Improvement in changing method of classifier
	Kaur et al. [20]	2018	UCI datasets	Hybrid Fisher with bat parameter
	Naik et al. [21]	2020	Microarray data	Developed BBA with neural network classifier
Firefly (FFA)	Emary et al. [22]	2015	UCI datasets	First introduced a binary version of FFA
	Kanimozhi and Latha [23]	2015	Image dataset	Introduced application of FFA in image retrieval
	Subha and Murugan [24]	2016	cardiotocogram data	Introduced application of FFA in disease detection
	Zhang et al. [25]	2017	public data	Introduced return cost based FFA
	Xu et al. [26]	2018	UCI datasets	Proposed combination of binary FFA with opposition based on learning algorithm
Flower pollination (FPA)	Rodrigues et al. [27]	2015	UCI datasets	Developed binary version of FPA
	Zawbaa and Emary [28]	2018	UCI datasets	Introduced binary version of FPA and KNN for feature selection in multi-objective problem
	Rajamohana et al. [29]	2017	UCI datasets	Proposed different value of lambda parameter
	Majidpour et al. [30]	2018	UCI datasets	Proposed BFPA with Ada-boost algorithm
	Sayed et al. [31]	2016	UCI datasets	Proposed clonal selection algorithm for exploitation and FPA for exploration to form the binary clonal FPA
	Yan et al. [32]	2019	Biomedical datasets	Improving FPA using adaptive transfer function for binary encoding and gaussian mutation strategy for exploitation
Grey Wolf (GWO)	Emary et al. [33]	2016	UCI datasets	First binary version of GWO
	Sharma et al. [34]	2019	Parkinson's disease dataset	Identifying symptoms using GWO with KNN and decision tree
	Pathak et al. [35]	2019	Bosbase ver 1.01 dataset	Introduced levy flight GWO
	Devanathan et al. [36]	2019	Indirect Immunofluorescence image	Applied for diagnosis of cardiovascular disease and obtain another version of binary GWO
	Al-Tashi et al. [37]	2018	UCI datasets	Proposed Combination GWO with SVM as a classifier
	Al-Tashi et al. [6]	2020	UCI datasets	Proposed binary version of GWO uses sigmoidal functions and ANN as the classifier for application to a multi-objective problem
	B Ahmadi et al.[38]	2022	UCI datasets	proposed new transfer functions and a new updating scheme for parameters called advanced GWO

hyperbolic tangent. Most of the algorithms experiment using UCI datasets. The study also discovered that there is still room for improvement in overcoming slow convergence rates because of random generation movement, avoiding traps into local optima, improving in finding the direction of exploration space, and adjusting the parameter to overcome premature convergence.

The multi-objective study also has contributed to proposing various methods. For example, Al-Tashi et al. [6]’s research work focuses on multi-objective feature selection problems using binary versions. The study also shows that the sigmoid function obtains good results for all 15 benchmark datasets. ANN classifiers were used in this study that will first be introduced with GWO. He also implements the proposed method in oil and gas datasets [7]. As a result, it obtains good results compared with MOPSO and NSGA-II methods. Furthermore, a study published in [8] offered a reinforced memory strategy and coupled mutations to improve multi-objective PSO with two competing goals: classification performance and reliability, and this method was compared to

two state-of-the-art systems. For face emotion recognition systems, another study [9] used a wrapper-based feature selection method based on a multi-objective differential evolution algorithm. As a classifier, a linear SVM was used. The proposed method has two goals: the number of feature characteristics and the accuracy of classification. Three facial recognition databases were used to test the suggested approach.

A MOPSO was proposed in [10] and [11] to address the issue of feature selection, with two distinct implementations investigated: 1) introduced the notion of non-dominated sorting, and 2) utilized the principles of dominance, mutation, and crowding to find Pareto’s front alternatives.

**D. BINARY MULTI-OBJECTIVE GREY WOLF OPTIMIZATION (BMOGWO-S)**

The proposed method is the primary focus of this study because it outperforms the Non-dominated Sorting Genetic Algorithm (NSGA-II), MGWO with tanh transfer function, and Multiple-objective Particle Swarm



Optimization (MPSO) in most cases, as demonstrated by Al-Tashi et al. [6].

MOGWO was designed to solve continuous optimization issues. Hence, it cannot be utilized to solve multi-objective feature selection problems directly. As a result, a new component is added to the preceding two components. These are the prerequisite for the algorithm to be effective in extracting features. As a result, a binary version is created to solve the selected problems. The sigmoid activation function is used, as shown in Equation (16).

$$x_d^{t+1} = \begin{cases} 1, & \text{if } \text{sigmoid} \left( \frac{x_1 + x_2 + x_3}{3} \right) \geq r \text{ and} \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

where  $x_d^{t+1}$  is a new position in binary form in the d dimension at t iteration. A random number is selected between 0 and 1 taken from a uniform distribution, and sigmoid (a) is explained as below:

$$\text{sigmoid}(a) = \frac{1}{1 + e^{-10(x-0.5)}} \quad (17)$$

### III. PROPOSED METHOD

In this part, the proposed method is explained, which is an extension of Binary Multi-Objective Grey Wolf Optimization (BMOGWO-S).

#### A. INITIALIZATION OF GREY WOLF POPULATION

In this section, we introduced new initialization in order to help the grey wolf to obtain the nearest initial Pareto front as earlier as possible. This strategy helps the wolf have an enough time to get updated on a better Pareto front. Based on a prior work, we found the data for the target initial population. We have founded that the optimum numbers of features are less than half of the total features in all datasets. Based on the observations, the initial population with several features ranging from 1 to half of the total features in each dataset has created. The feature's position is determined, based on the population factor strategy as introduced by [38]. The initial population is calculated as below:

$$x_{ij} = \begin{cases} 1 & \text{if } r \text{ and } > \psi_k \quad i = 1, 2, \dots, 4, \\ 0 & \text{otherwise, } j = 1, 2, \dots, n, \quad k = 1, 2, \dots, 5, \end{cases} \quad (18)$$

where n is the total number of features. In this study we choose  $\Psi$  value as 0.1, 0.2, 0.3, 0.4 and 0.5. Each value of  $\Psi$  is given four members, and the total population is 20 grey wolves. Figure 2 shows the entire process for the proposed algorithm, and its pseudo code can be found in Algorithm 1.

#### B. OBJECTIVE FUNCTION FORMULATION

As in most features selection studies, the multi-objective function is still the same as below:

- a) Minimizing the feature number.
- b) Minimizing the error rate of classification.

The objective function for both problems is calculated as mentioned in Equation (19).

$$\begin{aligned} & \text{Minimize } F(x) \\ & = \begin{cases} f_1(x) = \frac{M}{N} \quad M \in N, \quad N \in R^+ \\ f_2(x) = \frac{FP + FN}{TP + TN + TF + TN} \times 100, \quad (P + N) \in R^+, \end{cases} \end{aligned} \quad (19)$$

where  $f_1$  refers to objective (a), the ratio of the number of features selected and total features. While  $f_2$  refers to the second objective. Moreover, M signifies the selected features, and N denotes the entire dataset's features. Furthermore, TP stands for true positives, TN for true negatives, FP for false positives, and FN for false negatives

### IV. EXPERIMENTAL STRUCTURE

To verify the proposed algorithm, one benchmarking method is used for comparison. We compare the proposed method with Binary Multi-Objective Grey Wolf Optimization (BMOGWO-S). Two approaches will be compared. The first involves converting binary using the crossover function, as shown in Equation (20), known as BMOGWO-C.

$$x_d^{t+1} = \begin{cases} 1, & \text{if } \text{crossover} (x_1 + x_2 + x_3) \geq r \text{ and} \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

The second approach is explained in the previous section. All methods are implemented in MATLAB 2017a.

TABLE 2. Datasets.

DATASET	#FEATURES	CLASSES	SAMPLES
Breastcancer	9	2	699
WineEW	13	3	178
HeartEW	13	2	270
Zoo	16	7	101
Lymphography	18	4	184
SpectEW	22	2	267
Ionosphere	34	2	351
KrvskpEW	36	2	3196
WaveformEW	40	3	5000
SonarEW	60	2	208
Hillvalley	100	2	606
Musk Version 1	166	2	476
Madelon	500	2	4400
Oil and gas	22	4	377

#### A. DATASETS

Thirteen standard datasets from the UC Irvine machine learning repository [39] and one dataset from oil and gas were utilized to verify the proposed multi-objective methods. Table 2 shows an example of the features of the benchmark datasets that have been used, packed with a variety of attributes (ranging from 9 to 500), classes (101 to 5000), and samples (101 to 500) (from 2 to 7).

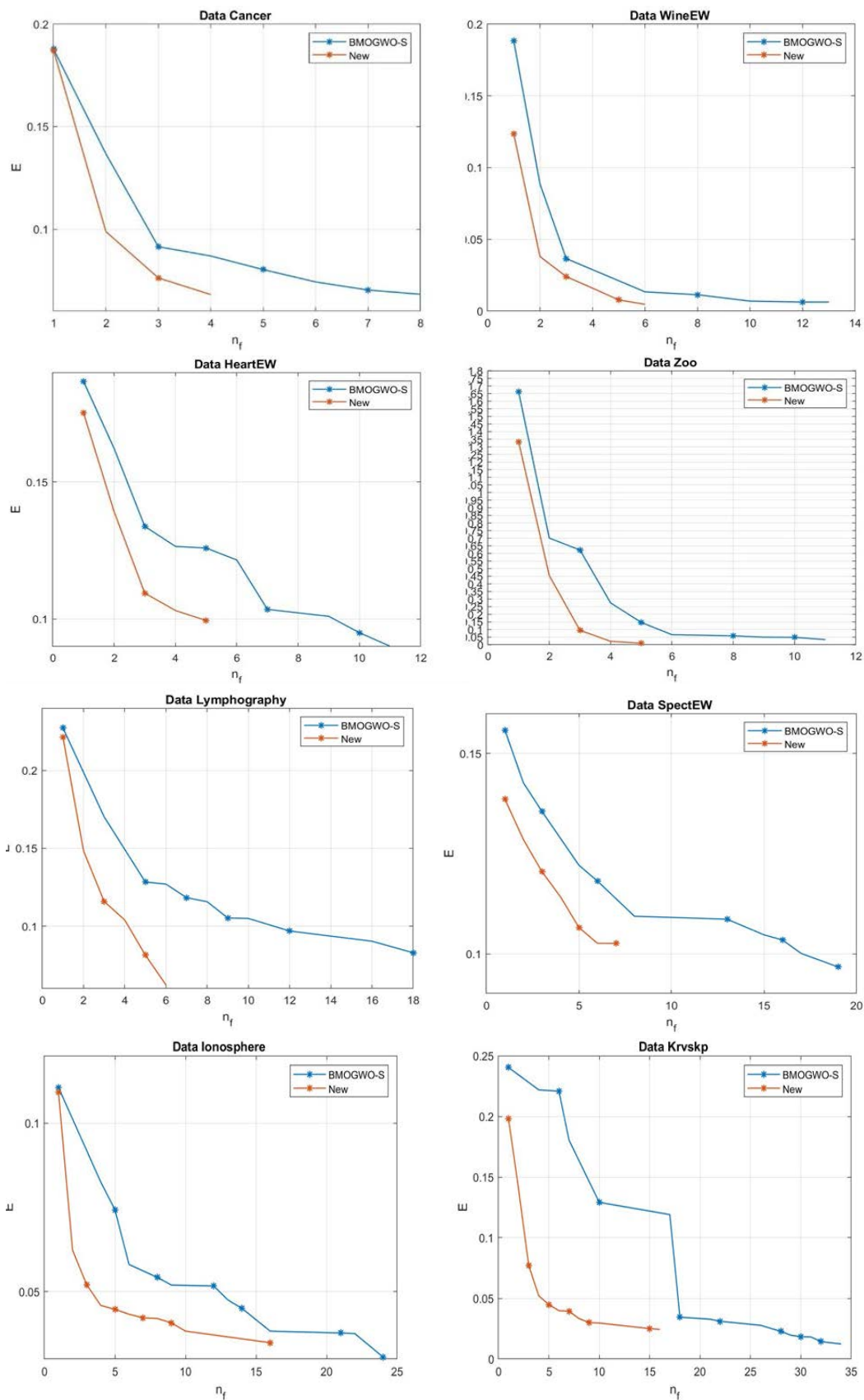


FIGURE 2. Comparison between proposed method and BMOGWO-S. This is result using both methods.

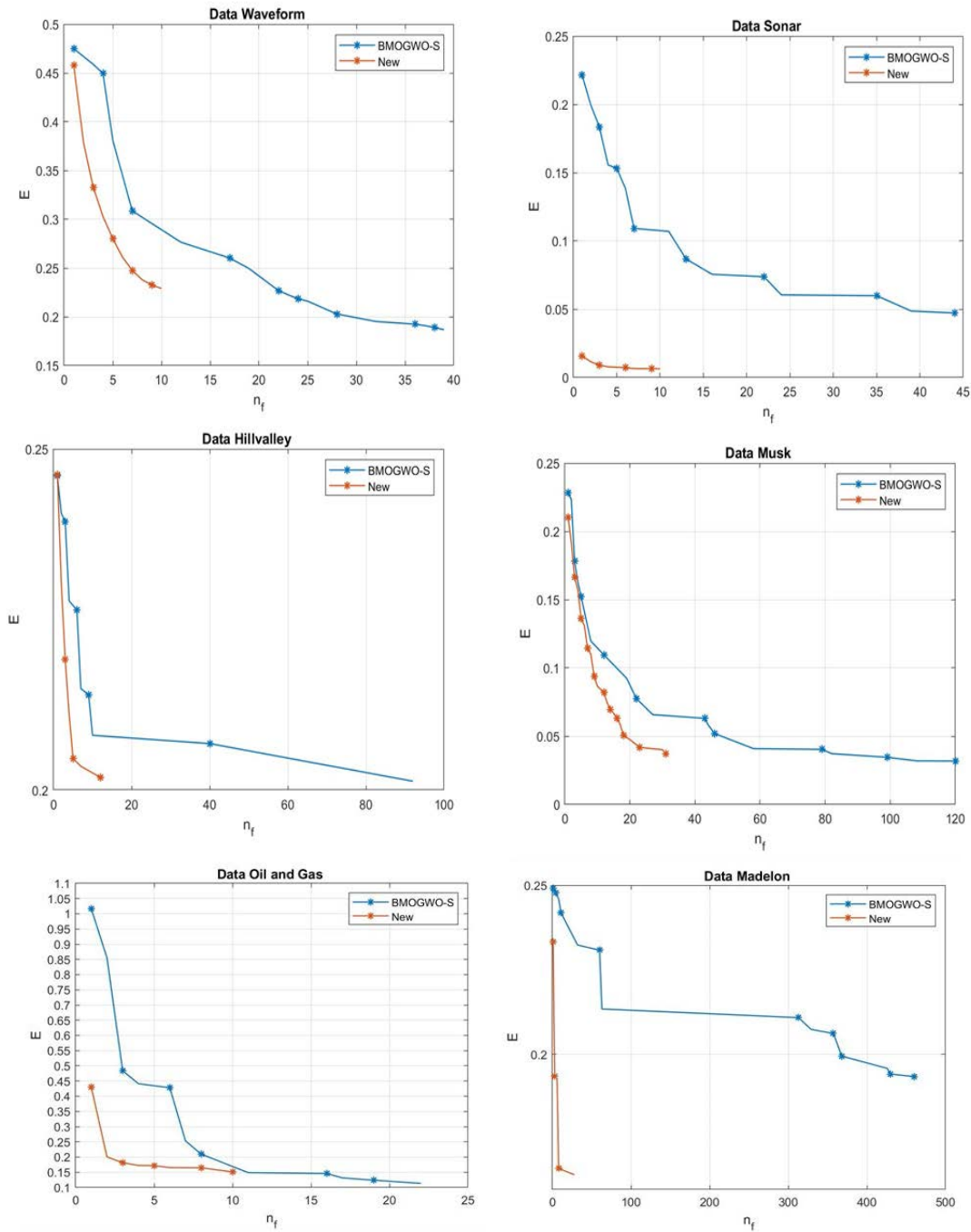


FIGURE 2. (Continued.) Comparison between proposed method and BMOGWO-S. This is result using both methods.

**B. PARAMETER SETTINGS**

Data is partitioned into training and testing randomly, training dataset (70%), and testing and validation (30%).

For classification, an ANN Feed forward network with ten hidden layers and a learning rate of 0.8 is used. ANN is chosen for classification because it is a powerful classifier model in terms of error rate and feature number reduction. The overall algorithm is based on the wrapper-based method, which needs a classifier to get the value of error rate and number of features during the evolutionary training process.

Another parameter setting for the feature selection method is shown below:

- Total number of iterations is 100.
- Number of grey wolves is 20.
- Achieve size is 50.
- Alpha parameter set up as 0.1.
- Number of the grid is 10.
- Value of the beta parameter is 4.
- Value of the gamma parameter is 2.

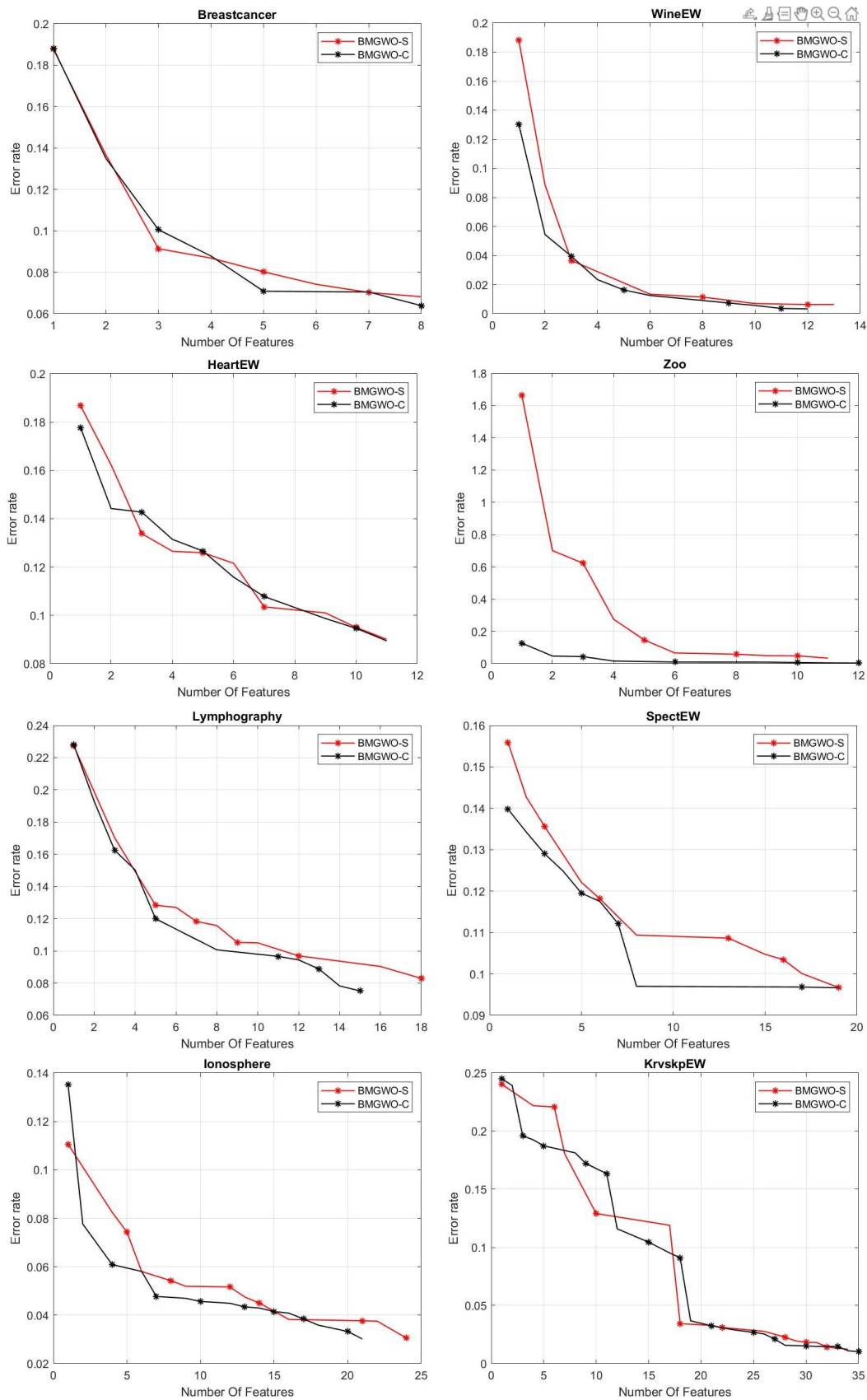


FIGURE 3. Comparison between BMOGWO-C and BMOGWO-S.



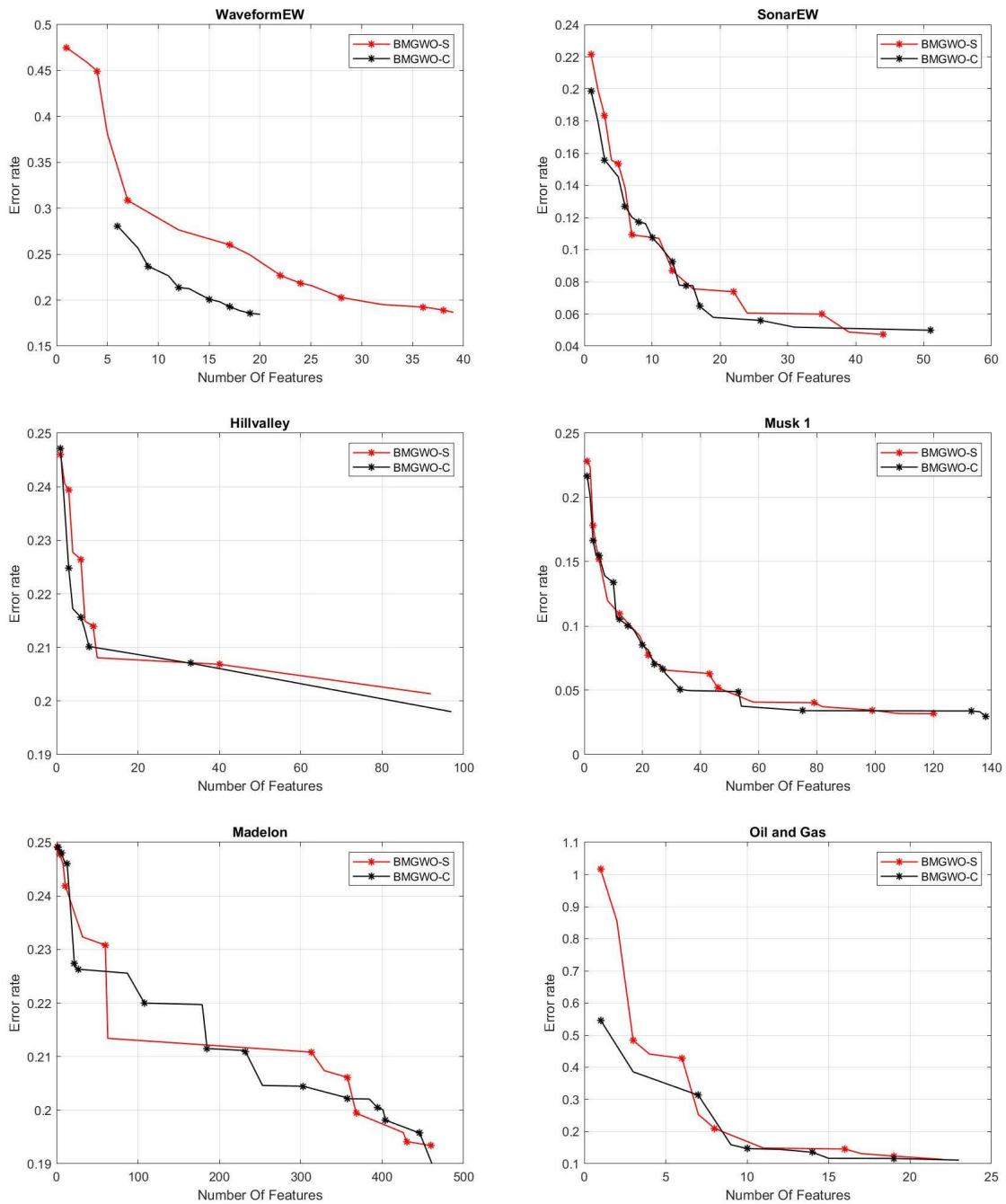


FIGURE 3. (Continued.) Comparison between BMOGWO-C and BMOGWO-S.

The values of the parameters are chosen based on a study shown by Al-Tashi et al. [16] which produced optimistic results, and the proposed study’s solely focused on the cause and effect of initial population into BMOGWO-S algorithms.

V. RESULT AND DISCUSSION

A. COMPARISON OF THE PROPOSED METHOD AND BMOGWO-S

Figure 2 shows the experimental result of our proposed method i.e. BMOGWO-S and the name of datasets are stated on each graph. The selected features are represented by

horizontal lines, while the error rate is represented by a vertical line. The blue color is represented by BMOGWO-S in the graph. In contrast, the new approach represents the red line. In this study we are dividing the data into small, medium, and large datasets, where the small datasets contain on data related to breast cancer, such as HeartEW, WineEW, Zoo, and Lymphography. We have SpectEW, Ionosphere, KrvskpEW, waveform, and oil and gas in the medium dataset. The big dataset group contains HillvalleyEW, Musk 1, and Madelon. This division is based on total number of features for each dataset. Figure 3 shows that in the majority of the datasets,

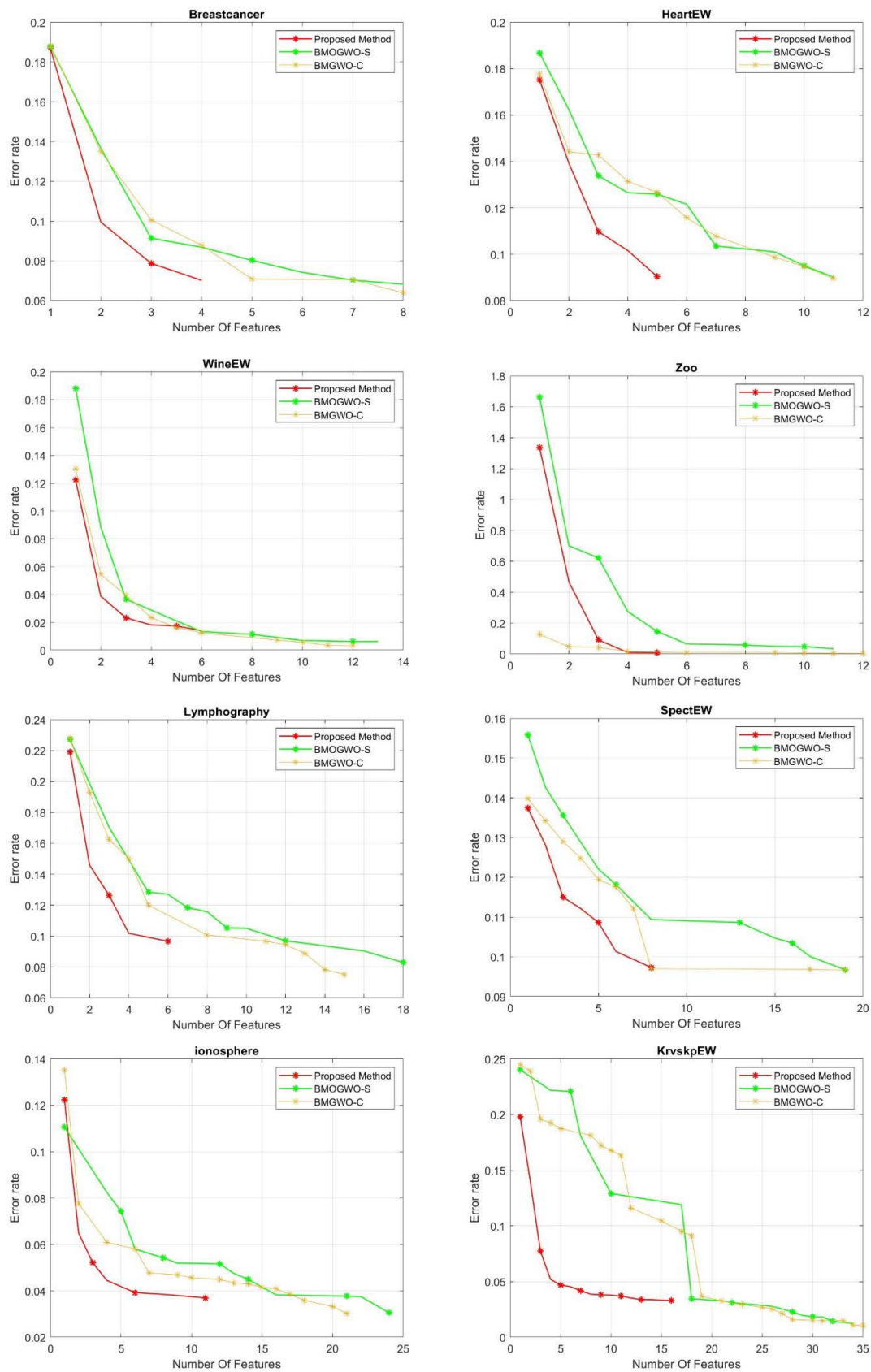


FIGURE 4. Comparison proposed method, BMOGWO-S and BMOGWO-C.

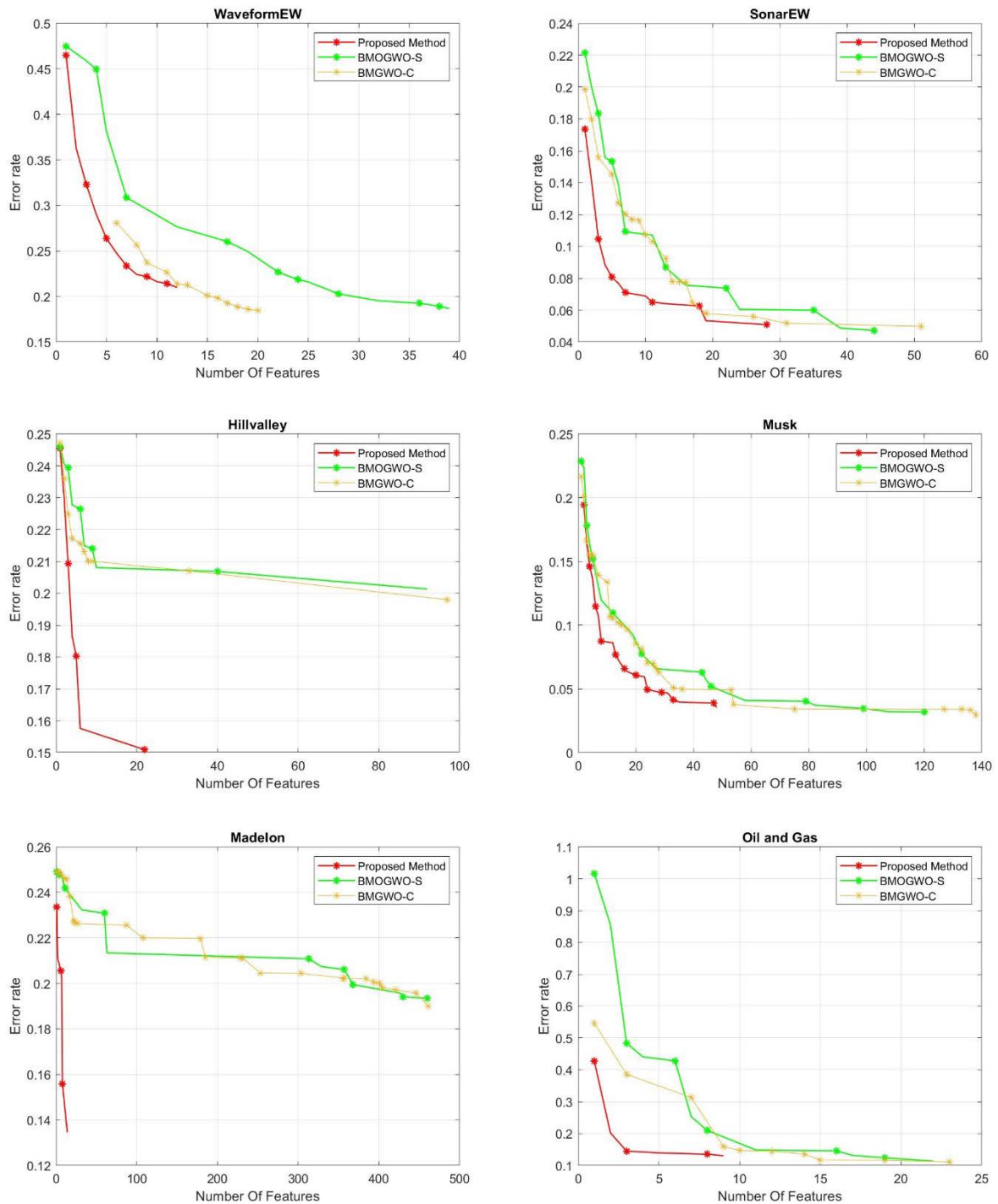


FIGURE 4. (Continued.) Comparison proposed method, BMOGWO-S and BMOGWO-C.

the proposed method yielded a non-dominated solution. In the breast cancer dataset the proposed method, choose 4 from 9 features around 40% and choose 6 out of 13 features in WineEW datasets. For dataset HeartEW, 5 from 13 features were selected and 5 from 16 for the zoo dataset. In the Lymphography dataset, the proposed method chooses 6 out of 18 features. All these are small datasets. In the SpectEW dataset, 8 out of 22 features were selected. In the Ionosphere

dataset, 11 features were selected from 34 of the total, 16 from 36 features selected for the KrvskpEW dataset, and for the waveform dataset new approach, choose 12 features from 40 features. The medium dataset has a selected range of 20% to 40 % features and an error rate of less than 0.45. For the Hillvalley dataset, the proposed method selects 22 from 100 features and 48 from 166 features selected for dataset Musk 1 and all the result is shown in Table 4. Using all

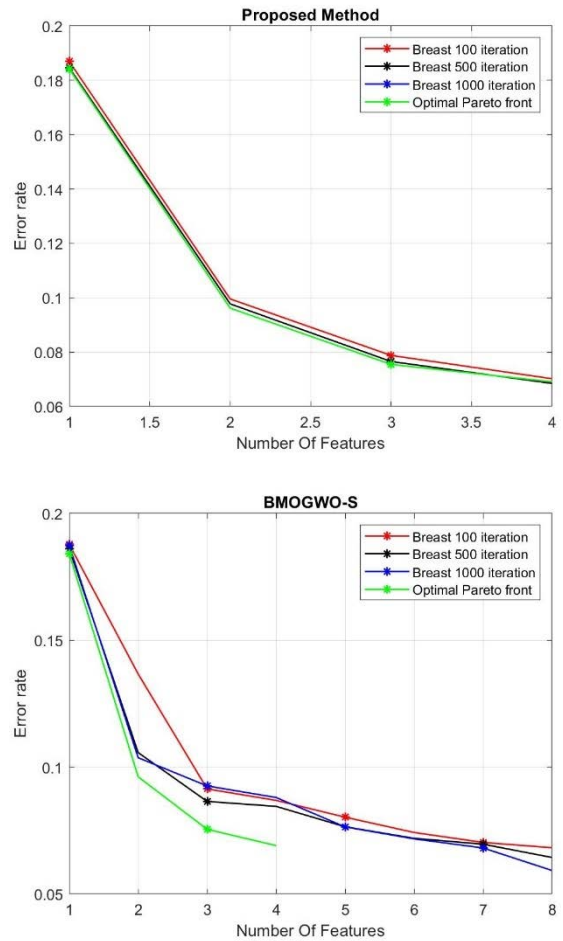
**Algorithm 1** Pseudo Code Proposed Method

Separate dataset into training and testing  
 Initialization of parameter and population as shown in Equation (18)  
 Fitness value which is the error rate and feature numbers selected for search member  
 Obtain a non-dominated solution & initialize the archive  
 Select the finest leader  $X_\alpha$ ,  $X_\beta$ , and  $X_\delta$  from archive  
 $t = 1$ ; iteration = 100  
**While** ( $t < \text{iterations}$ )  
**For** each search member  
 Update the current positions of the search member by Equations (5-11)  
 Binarize the updated positions using sigmoid activation function in Equations (17)  
**End For**  
 Update the parameters  
 Compute the fitness values (Features and Error rate) for search members  
 Determine the non-dominated solutions in the current population  
 Modify the archive with the attained non-dominated solutions  
**If** (there is no more space in the archive)  
 Delete a solution using the grid mechanism  
 Insert the new obtained solution to the archive  
**End If**  
 Choose alpha from the archive:  $X_\alpha$   
 Avoid selecting  $X_\alpha$ , by temporarily excluding it from the archive  
 Choose beta from the archive:  $X_\beta$   
 Avoid selecting  $X_\beta$  by temporarily excluding it from the archive  
 Choose delta from the archive:  $X_\delta$   
 Add fitness leader  $X_\delta$  and second fitness leader ( $X_\delta$ ) back to the archive  
 $t = t + 1$   
**End While**  
**Return Archive**

the mentioned datasets the propose methods outperforms the BMOGWO-S methods in terms of both the number of features and error rate. In some cases i.e. large dataset, the BMOGWO-S obtained a solution with a similar error rate to the proposed method. For example, in Musk 1 dataset, two solutions have a similar error rate. Finally, the oil and gas dataset, the proposed method outperforms the BMOGWO-S and helps to reduce both the number of features and the error rate. We discovered a significant difference between the proposed method and BMOGWO-S in all datasets using t-test analysis, as shown in Table 6.

**B. COMPARISON BMOGWO-C AND BMOGWO-S**

Figure 3 shows that the black color represents BMOGWO-C and the red color represents BMOGWO-S. In the Zoo, SpectEW and Waveform dataset, method BMOGWO-C



**FIGURE 5.** Comparison sensitivity analysis between proposed method and BMOGWO-S.

clearly dominates in all points compared to other datasets. Most datasets for both methods do not clearly show dominance over each other in this comparison.

**C. COMPARISON OF THE PROPOSED METHOD, BMOGWO-S AND BMGWO-C**

Figure 4 shows the experimental results of our proposed method versus BMGWO-S and BMGWO-C for 13 datasets. On top of each graph, the names of the datasets are specified. Each x axis represents the number of features, while the y axis represents the classification error rate. The yellow line shows the BMGWO-C, the green one shows the BMOGWO-S, and the red line shows the suggested algorithm. In most cases, our proposed method outperforms the state-of-the-art method in terms of both error rates and feature selection. For example, in small size datasets (Breastcancer, WineEW, Zoo and Lymphography) the proposed method dominates both BMGWO-S and BMGWO-C in terms of both features number and error rate. In the Zoo datasets, the BMGWO-C dominates in terms of features and classification error rate.

When comparing MGWO-S and BMGWO-C based on medium-sized datasets like (SpectEW, IonosphereEW,

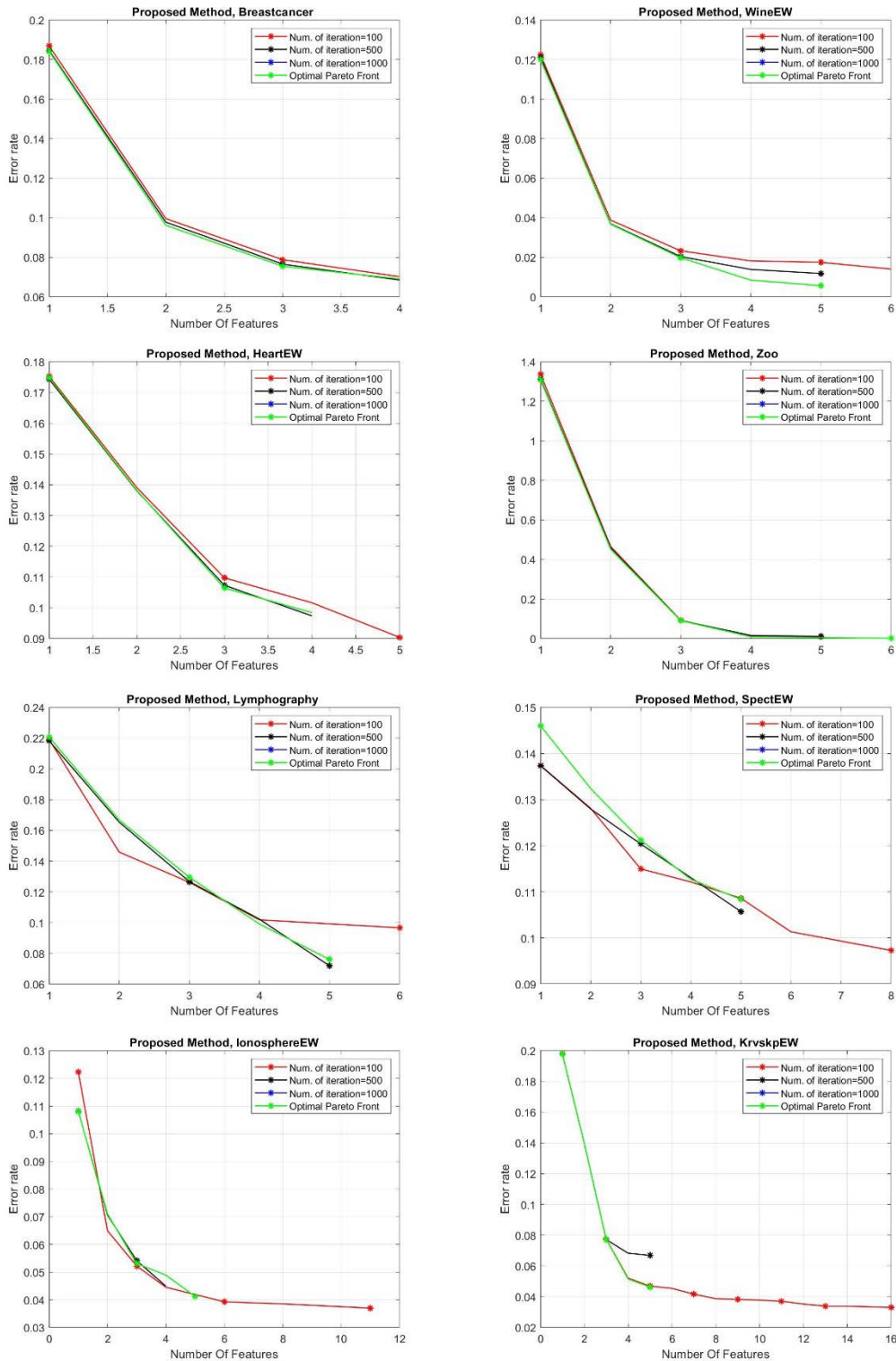


FIGURE 6. Sensitivity analysis using proposed method for all dataset.

KrvskpEW, and Waveform), it was founded that our proposed method performed better than the BMGWO-S, BMGWO-C in terms of feature reductions and improved classification accuracy.

The proposed method performs better in larger datasets, particularly in Hillvalley, Musk, and Madelon in terms of the number of features that are reduced. For instance, in the Hillvalley dataset, the suggested technique has selected 22% of all



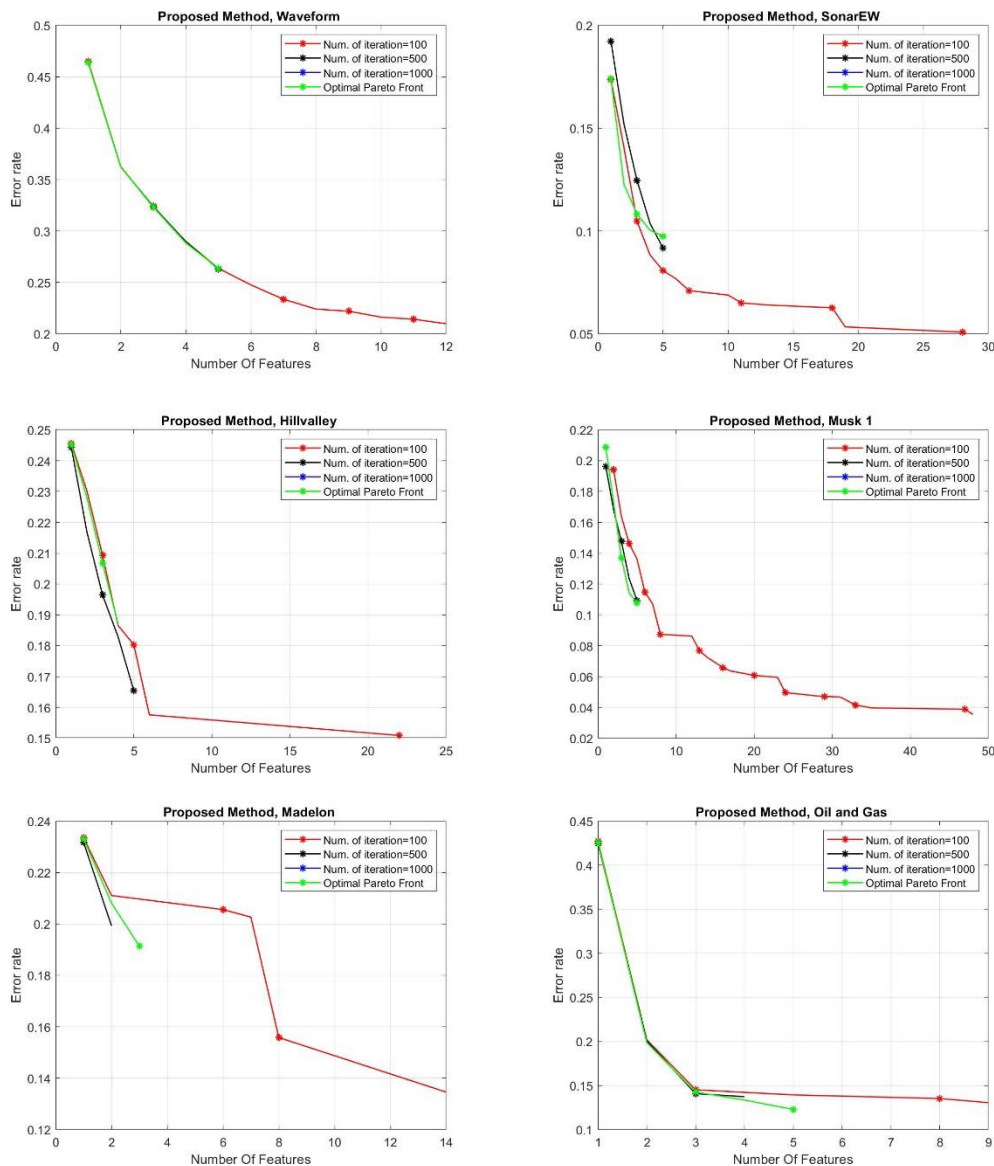


FIGURE 6. (Continued.) Sensitivity analysis using proposed method for all dataset.

features, compared to BMGWO-S i.e. 92% and BMOGWO-C, 97%. The results showed that BMGWO-C can achieve better outcomes than BMGWO-S and BMGWO-C in the Musk and Madelon dataset. Additionally, the datasets for oil and Gas exhibit superior results.

**D. DISCUSSION**

The experiment showed that the proposed method gives superior performance compared to BMOGWO-S. Based on Figure 5 is the result of sensitivity analysis on the number of iterations for proposed method and BMOGWO-S. From figure, by increasing the number of iterations, the optimal pareto front is obtained at 1000 iteration for proposed method. But, for BMOGWO-S after 1000 iteration still not achieve the optimal pareto front. Figure 6 is sensitivity analysis for all dataset. From the graph, several dataset such as

Zoo, KrvskepEW, Ionosphere and Waveform obtain optimal solution at 100 iteration. While for dataset Breastcancer, WineEW, Musk 1 and oil and gas achieve optimal solution at 1000 iteration. Introducing wolf, which has the target position characteristic, assists the wolf in obtaining the correct leader at the same time while still maintaining the technique to sustain the diversity of the leader selection. We still maintain the good attributes of BMOGWO-S in our proposed method, such as the ability to maintain the balancing exploration and exploitation. The results also show that the crossover method has no discernible effect on the BMGWO-S methods.

Based on literature review, there are a few study using same dataset which are Breast Cancer and Ionosphere. From [16], for Breast Cancer dataset, method Bat Algorithm, Harmony Search and Fire Fly Algorithm obtain 7 features, while Particle Swarm Optimization and Gravitational Search Algorithm

**TABLE 3.** Statistics result on the number selected features produces by two algorithms.

		<i>PF-BMOGWOS</i>	<i>BMOGWO-S</i>	<i>BMOGWO-C</i>
<b>Breastcancer</b>	Min	1	1	1
	Max	4	8	8
	Range	3	7	7
	STD	1.291	2.4495	2.5635
	Mean	2.5	4.5	4.2857
<b>WineEW</b>	Min	1	1	1
	Max	6	13	12
	Range	5	12	11
	STD	1.8708	4.6117	3.9455
	Mean	3.5	6.875	6.3
<b>HeartEW</b>	Min	1	1	1
	Max	5	11	11
	Range	4	10	10
	STD	1.5811	3.4254	3.4254
	Mean	3	5.8	5.8
<b>Zoo</b>	Min	1	1	1
	Max	5	11	12
	Range	4	10	11
	STD	1.5811	3.4785	4.1567
	Mean	3	5.9	6.4444
<b>Lymphography</b>	Min	1	1	1
	Max	6	18	15
	Range	5	17	14
	STD	1.9235	5.1821	5.1962
	Mean	3.2	8.6364	8
<b>SpectEW</b>	Min	1	1	1
	Max	8	19	19
	Range	7	18	18
	STD	2.4103	6.6085	6.1065
	Mean	4.1429	9.5455	7.2
<b>Ionosphere</b>	Min	1	1	1
	Max	11	24	21
	Range	10	23	20
	STD	3.559	7.3082	6.2819
	Mean	5	11.9231	11.5625
<b>KrvskpEW</b>	Min	1	1	1
	Max	16	34	35
	Range	15	33	34
	STD	4.5898	11.0785	10.9671
	Mean	8.0663	19.75	17.88
<b>WaveformEW</b>	Min	1	1	6
	Max	12	39	20
	Range	11	38	14
	STD	3.6056	12.966	4.5594
	Mean	6.5	20.667	13.6667
<b>SonarEW</b>	Min	1	1	1
	Max	28	44	51
	Range	27	43	50
	STD	7.9494	14.2421	11.9159
	Mean	9.7692	15.4667	13.8947
<b>Hillvalley</b>	Min	1	1	1
	Max	22	92	97
	Range	21	91	96
	STD	7.1979	28.5431	29.5672
	Mean	6.1429	17.4	17
<b>Musk 1</b>	Min	2	1	1
	Max	48	120	138
	Range	46	119	137
	STD	13.7407	39.2616	44.0276
	Mean	19.3182	41.1579	38.2593
<b>Madelon</b>	Min	1	1	2
	Max	14	460	461
	Range	13	459	459

**TABLE 3. (Continued.) Statistics result on the number selected features produces by two algorithms.**

	STD	4.6761	191.4435	170.8419
	Mean	6.333	215.1176	204.6923
<b>Oil and gas</b>	Min	1	1	1
	Max	9	22	23
	Range	8	21	22
	STD	3.266	7.1901	6.8158
	Mean	4.667	9.667	11.3

**TABLE 4. T-test statistic result between new approach and BMOGWO-S.**

DATASET	T-VALUE	P-VALUE
Breastcancer	-40.9583	8.255E-188
WineEW	-24.0689	1.0353E-57
HeartEW	-29.0992	4.4605E-85
Zoo	-18.4331	6.2204E-34
Lymphography	-37.2426	2.4313E-87
SpectEW	-41.9471	2.871E-119
Ionosphere	-23.9676	8.2208E-76
KrvskpEW	-148.688	0
WaveformEW	-354.212	0
SonarEW	-44.8655	1.323E-108
Hillvalley	-83.0896	0
Musk Version 1	-73.0181	2.314E-260
Madelon	-1355.03	0
Oil and gas	-33.6708	1.056E-109

obtained 8 features. This shows that our proposed method gets better result which is 4 features. However, there is still insufficient comparison because the majority of studies use single objective methods. It is impossible to compare non-dominated solutions. Other methods with multiple objectives for the feature selection problem do not use UCI datasets as references, so no comparison can be made. We assume that the proposed method is better than BMOGWO-S, which means that also better than MOPSO using tanh transfer function, NGSa-II and MOGWO.

**VI. CONCLUSION AND FUTURE WORK**

From the previous research, we found the binary type of multi-objective Grey Wolf Optimizer (GWO) is efficient in handling feature selection problems. We maintain all the parts mentioned in the previous study therefore, we can still get beneficent. Initialization population using population factor as prior information into the algorithm, as a factor that will lead the exploration and exploitation in the correct direction. As a result, it will help the algorithm obtain the correct position and select the correct wolf. It shows that our proposed method obtains a non-dominated solution in all datasets. It also demonstrates in selected fewer features number and error rate compared to Binary Multi-Objective Grey Wolf Optimization (BMOGWO-S).

Despite the fact that there are various approaches, most of them have a single goal: to enhance the accuracy of classification. There is not enough research in the literature to explore feature selection as a multi-objective problem. As a result, designing a multi-objective grey wolf to solve the challenge of feature selection still has space.

The main limitation of our study is there no detail study on effect of choosing different values of parameter  $\Psi$  into pareto front which may be show different result for each datasets. For future work, we found there is still space for improvement in terms of attained in the Pareto front, and we will continue finding it. Other than that, we want to compare the filter method based with our proposed method. We also want to know the effect of modification in the binary part by replacing it using a mutation algorithm and comparing it with the result with state-the-art methods.

**CONFLICT OF INTEREST**

All authors declare that they have no conflict of interest.

**REFERENCES**

- [1] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014, doi: 10.1016/j.advengsoft.2013.12.007.
- [2] M. H. Nadimi-Shahraki, S. Taghian, and S. Mirjalili, "An improved grey wolf optimizer for solving engineering problems," *Expert Syst. Appl.*, vol. 166, Mar. 2021, Art. no. 113917, doi: 10.1016/j.eswa.2020.113917.
- [3] O. A. Alomari, S. N. Makhadmeh, M. A. Al-Betar, Z. A. A. Alyasseri, I. A. Doush, A. K. Abasi, M. A. Awadallah, and R. A. Zitar, "Gene selection for microarray data classification based on gray wolf optimizer enhanced with TRIZ-inspired operators," *Knowl.-Based Syst.*, vol. 223, Jul. 2021, Art. no. 107034, doi: 10.1016/j.knsys.2021.107034.
- [4] E.-S. 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, doi: 10.1109/ACCESS.2020.3001151.
- [5] Q. Al-Tashi, H. Rais, S. J. Abdulkadir, S. Mirjalili, and H. Alhussian, "A review of grey wolf optimizer-based feature selection methods for classification," in *Evolutionary Machine Learning Techniques*. Singapore: Springer, 2020, pp. 273–286.
- [6] Q. Al-Tashi, S. J. Abdulkadir, H. M. Rais, S. Mirjalili, H. Alhussian, M. G. Ragab, and A. Alqushaibi, "Binary multi-objective grey wolf optimizer for feature selection in classification," *IEEE Access*, vol. 8, pp. 106247–106263, 2020, doi: 10.1109/ACCESS.2020.3000040.
- [7] Q. Al-Tashi, E. A. P. Akhri, S. J. Abdulkadir, S. Mirjalili, T. M. Shami, H. Alhussian, A. Alqushaibi, A. Alwadain, A. O. Balogun, and N. Al-Zidi, "Classification of reservoir recovery factor for oil and gas reservoirs: A multi-objective feature selection approach," *J. Mar. Sci. Eng.*, vol. 9, no. 8, p. 888, Aug. 2021, doi: 10.3390/jmse9080888.
- [8] Z. Yong, G. Dun-Wei, and Z. Wan-Qiu, "Feature selection of unreliable data using an improved multi-objective PSO algorithm," *Neurocomputing*, vol. 171, pp. 1281–1290, Jan. 2016, doi: 10.1016/j.neucom.2015.07.057.
- [9] U. Mlakar, I. Fister, J. Brest, and B. Potočnik, "Multi-objective differential evolution for feature selection in facial expression recognition systems," *Expert Syst. Appl.*, vol. 89, pp. 129–137, Dec. 2017, doi: 10.1016/j.eswa.2017.07.037.
- [10] V. Tiwari, "Face recognition based on cuckoo search algorithm," *Indian J. Comput. Sci. Eng.*, vol. 3, pp. 1–9, Jul. 2012.
- [11] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms," *Appl. Soft Comput.*, vol. 18, pp. 261–276, May 2014, doi: 10.1109/TSMCB.2012.2227469.

- [12] D. Rodrigues, L. A. M. Pereira, T. N. S. Almeida, J. P. Papa, A. N. Souza, C. C. O. Ramos, and X.-S. Yang, "BCS: A binary cuckoo search algorithm for feature selection," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 465–468, doi: [10.1109/ISCAS.2013.6571881](https://doi.org/10.1109/ISCAS.2013.6571881).
- [13] C. Gunavathi and K. Premalatha, "Cuckoo search optimisation for feature selection in cancer classification: A new approach," *Int. J. Data Mining Bioinf.*, vol. 13, no. 3, pp. 248–265, 2015, doi: [10.1504/ijdm.2015.072092](https://doi.org/10.1504/ijdm.2015.072092).
- [14] S. Salesi and G. Cosma, "A novel extended binary cuckoo search algorithm for feature selection," in *Proc. 2nd Int. Conf. Knowl. Eng. Appl. (ICKEA)*, Oct. 2017, pp. 6–12, doi: [10.1109/ICKEA.2017.8169893](https://doi.org/10.1109/ICKEA.2017.8169893).
- [15] A. Pandey, D. Rajpoot, and M. Saraswat, "Feature selection method based on hybrid data transformation and binary binomial cuckoo search," *J. Ambient Intell. Humanized Comput.*, vol. 11, pp. 719–738, Feb. 2020, doi: [10.1007/s12652-019-01330-1](https://doi.org/10.1007/s12652-019-01330-1).
- [16] R. Y. M. Nakamura, L. A. M. Pereira, K. A. Costa, D. Rodrigues, J. P. Papa, and X.-S. Yang, "BBA: A binary bat algorithm for feature selection," in *Proc. 25th SIBGRAPI Conf. Graph. Patterns Images*, Aug. 2012, pp. 291–297, doi: [10.1109/SIBGRAPI.2012.47](https://doi.org/10.1109/SIBGRAPI.2012.47).
- [17] M. A. Laamari and N. Kamel, "A hybrid bat based feature selection approach for intrusion detection," in *Bio-Inspired Computing—Theories and Applications*, vol. 472. Berlin, Germany: Springer, 2014, pp. 230–238.
- [18] D. Rodrigues, L. A. M. Pereira, R. Y. M. Nakamura, K. A. P. Costa, X.-S. Yang, A. N. Souza, and J. P. Papa, "A wrapper approach for feature selection based on bat algorithm and optimum-path forest," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2250–2258, 2014, doi: <https://doi.org/10.1016/j.eswa.2013.09.023>.
- [19] A.-C. Enache and V. Sgarciu, "A feature selection approach implemented with the binary bat algorithm applied for intrusion detection," in *Proc. 38th Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2015, pp. 11–15, doi: [10.1109/TSP.2015.7296215](https://doi.org/10.1109/TSP.2015.7296215).
- [20] T. Kaur, B. S. Saini, and S. Gupta, "A novel feature selection method for brain tumor MR image classification based on the Fisher criterion and parameter-free Bat optimization," *Neural Comput. Appl.*, vol. 29, pp. 193–206, Feb. 2017.
- [21] A. K. Naik, V. Kuppili, and D. R. Edla, "Efficient feature selection using one-pass generalized classifier neural network and binary bat algorithm with a novel fitness function," *Soft Comput.*, vol. 24, no. 6, pp. 4575–4587, Mar. 2020, doi: [10.1007/s00500-019-04218-6](https://doi.org/10.1007/s00500-019-04218-6).
- [22] E. Emary, H. M. Zawbaa, K. K. A. Ghany, A. E. Hassanien, and B. Parv, "Firefly optimization algorithm for feature selection," in *Proc. 7th Balkan Conf. Informat. Conf.*, Sep. 2015, pp. 1–7, doi: [10.1145/2801081.2801091](https://doi.org/10.1145/2801081.2801091).
- [23] T. Kanimozhi and K. Latha, "An integrated approach to region based image retrieval using firefly algorithm and support vector machine," *Neurocomputing*, vol. 151, pp. 1099–1111, Mar. 2015, doi: [10.1016/j.neucom.2014.07.078](https://doi.org/10.1016/j.neucom.2014.07.078).
- [24] V. Subha and D. Murugan, "Opposition based firefly algorithm optimized feature subset selection approach for fetal risk anticipation," *Mach. Learn. Appl., Int. J.*, vol. 3, no. 2, pp. 55–64, Jun. 2016, doi: [10.5121/mlaj.2016.3205](https://doi.org/10.5121/mlaj.2016.3205).
- [25] Y. Zhang, X.-F. Song, and D.-W. Gong, "A return-cost-based binary firefly algorithm for feature selection," *Inf. Sci.*, vol. 418, pp. 561–574, Dec. 2017, doi: [10.1016/j.ins.2017.08.047](https://doi.org/10.1016/j.ins.2017.08.047).
- [26] H. Xu, S. Yu, J. Chen, and X. Zuo, "An improved firefly algorithm for feature selection in classification," *Wireless Pers. Commun.*, vol. 102, no. 4, pp. 2823–2834, Oct. 2018, doi: [10.1007/s11277-018-5309-1](https://doi.org/10.1007/s11277-018-5309-1).
- [27] D. Rodrigues, X.-S. Yang, A. Souza, and J. Papa, "Binary flower pollination algorithm and its application to feature selection," *Stud. Comput. Intell.*, vol. 585, pp. 85–100, Jan. 2015, doi: [10.1007/978-3-319-13826-8\\_5](https://doi.org/10.1007/978-3-319-13826-8_5).
- [28] H. Zawbaa and E. Emary, "Applications of flower pollination algorithm in feature selection and knapsack problems," in *Nature-Inspired Algorithms and Applied Optimization*. Cham, Switzerland: Springer, 2018, pp. 217–243.
- [29] S. P. Rajamohana, K. Umamaheswari, and B. Abirami, "Adaptive binary flower pollination algorithm for feature selection in review spam detection," in *Proc. Int. Conf. Innov. Green Energy Healthcare Technol. (IGEHT)*, Mar. 2017, pp. 1–4.
- [30] H. Majidpour and F. S. Gharechopogh, "An improved flower pollination algorithm with AdaBoost algorithm for feature selection in text documents classification," *J. Adv. Comput. Res.*, vol. 9, no. 1, pp. 29–40, 2018. [Online]. Available: [http://jacr.iausari.ac.ir/article\\_653945.html](http://jacr.iausari.ac.ir/article_653945.html)
- [31] S. A.-F. Sayed, E. Nabil, and A. Badr, "A binary clonal flower pollination algorithm for feature selection," *Pattern Recognit. Lett.*, vol. 77, pp. 21–27, Jul. 2016, doi: [10.1016/j.patrec.2016.03.014](https://doi.org/10.1016/j.patrec.2016.03.014).
- [32] C. Yan, J. Ma, H. Luo, G. Zhang, and J. Luo, "A novel feature selection method for high-dimensional biomedical data based on an improved binary clonal flower pollination algorithm," *Hum. Heredity*, vol. 84, no. 1, pp. 1–13, 2019, doi: [10.1159/000501652](https://doi.org/10.1159/000501652).
- [33] E. Emary, W. Yamany, A. E. Hassanien, and V. Snasel, "Multi-objective gray-wolf optimization for attribute reduction," *Proc. Comput. Sci.*, vol. 65, pp. 623–632, Jan. 2015, doi: [10.1016/j.neucom.2015.06.083](https://doi.org/10.1016/j.neucom.2015.06.083).
- [34] P. Sharma, S. Sundaram, M. Sharma, A. Sharma, and D. Gupta, "Diagnosis of Parkinson's disease using modified grey wolf optimization," *Cogn. Syst. Res.*, vol. 54, pp. 100–115, May 2019.
- [35] Y. Pathak, K. V. Arya, and S. Tiwari, "Feature selection for image steganalysis using levy flight-based grey wolf optimization," *Multimedia Tools Appl.*, vol. 78, pp. 1473–1494, Jun. 2018.
- [36] K. Devanathan, N. Ganapathy, and R. Swaminathan, "Binary grey wolf optimizer based feature selection for nucleolar and centromere staining pattern classification in indirect immunofluorescence images," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 7040–7043, doi: [10.1109/EMBC.2019.8856872](https://doi.org/10.1109/EMBC.2019.8856872).
- [37] Q. Al-Tashi, H. Rais, and S. J. Abdulkadir, "Feature selection method based on grey wolf optimization for coronary artery disease classification," in *Proc. Int. Conf. Reliable Inf. Commun. Technol.*, Jun. 2018, pp. 257–266, doi: [10.1007/978-3-319-99007-1\\_25](https://doi.org/10.1007/978-3-319-99007-1_25).
- [38] B. Ahmadi, S. Younesi, O. Ceylan, and A. Ozdemir, "An advanced grey wolf optimization algorithm and its application to planning problem in smart grids," *Soft Comput.*, vol. 26, no. 8, pp. 3789–3808, Apr. 2022, doi: [10.1007/s00500-022-06767-9](https://doi.org/10.1007/s00500-022-06767-9).
- [39] C. Dua and D. Graff. (2019). *UCI Repository of Machine Learning Databases*. [Online]. Available: <http://archive.ics.uci.edu/ml>



**NUR LYANA SHAHFIQA ALBASHAH** received the B.Sc. (Hons.) and M.Sc. degrees in statistics from Universiti Kebangsaan Malaysia, in 2010 and 2012, respectively. She is currently pursuing the Ph.D degree with the Department of Fundamental and Applied Science, Universiti Teknologi Petronas, Malaysia. She is also working as a Research Officer at Universiti Teknologi Petronas. Her research interests include brain stroke, machine learning, metaheuristic algorithm, and fuzzy programming and classification.



**HELMI MD RAIS** received the B.Sc. degree in business administration from Drexel University, Philadelphia, PA, USA, the M.Sc. degree in information technology from Griffith University, Australia, and the Ph.D. degree in science and system management from Universiti Kebangsaan Malaysia, Malaysia. He is currently a Senior Lecturer with Universiti Teknologi Petronas. His research interests include ant colony algorithm, optimization, swarm intelligence, and database technologies.

• • •