Journal of Algorithms and Computation

journal homepage: http://jac.ut.ac.ir

Hybrid Filter-Wrapper Feature Selection using Equilibrium Optimization

Mohammad Ansari Shiri*[1] and Najme Mansouri†[2]

[1,2]Department of Computer Science, Shahid Bahonar University of kerman, Kerman, Iran.

ABSTRACT

The topic of feature selection has become one of the hottest subjects in machine learning over the last few years. The results of evolutionary algorithm selection have also been promising, along with standard feature selection algorithms. For K-Nearest Neighbor (KNN) classification, this paper presents a hybrid filter-wrapper algorithm based on Equilibrium Optimization (EO). With respect to the selected feature subset, the filter model is based on a composite measure of feature relevance and redundancy. The wrapper model consists of a binary Equilibrium Optimization (BEO). The hybrid algorithm is called filter-based BEO (FBBEO). By combining filters and wrappers, FBBEO achieves a unique combination of efficiency and accuracy. In the experiment, 11 standard datasets from the UCI repository

ARTICLE INFO

Article history:
Research paper
Received 07, January 2023
Received in revised form 12, March 2023
Accepted 09 May 2023
Available online 01, June 2023

*Keyword:*Feature selection, Classification, Wrapper , Filter, Equilibrium Optimization.

AMS subject Classification: 05C78.

*mohammadansari7685@gmail.com
†Corresponding author: N. Mansouri. Email: najme.mansouri@gmail.com

1 Abstract continued

were utilized. Results indicate that the proposed method is effective in improving the classification accuracy and selecting the best optimal features subsets with the least number of features.

2 Introduction

Many practical applications of feature selection (FS) have been empirically and theoretically demonstrated, including data mining, machine learning, natural language processing, and face recognition. Classification and clustering learning usually start with this step. The primary goal of feature selection is to find an optimal subset of the original features capable of capturing the approximation of the whole information. As a result, it simplifies the learning model, reduces training time, and decreases learning error rates by keeping relevant features and removing redundant ones [10, 15, 16]. Subset generation, evaluation, stopping conditions, and verification are the steps in feature selection. In order to evaluate the subset of features effectively, the most important step is to determine its effectiveness [26]. The general process of feature selection is illustrated in 1. In terms of evaluation criteria, there are three main classes of feature selection methods: filter, wrapper and embedded. There is no connection between the filter approach and the learning algorithm. This procedure focuses on inputs and outputs without considering classification accuracy. Learning models are used in wrapper methods to select subsets of features. In embedded systems, the search for subsets is incorporated in both the wrapper and filter learning algorithms [2, 19].
As a result of their complexity and time consumption, mathematical algorithms are no longer suitable for feature selection. Instead, metaheuristic algorithms that emulate living creatures' lives and their evolution are becoming increasingly popular. Global optimization can be achieved using metaheuristic methods. Nonlinear and indeterminate problems are solved more quickly by these types of algorithms than classical algorithms. A metaheuristic algorithm did not need to restart whenever new data was entered or the environment changed. When considering their nature, evolutionary algorithms can be used to select features that reduce processing time and produce classification results that are acceptable in accuracy. In addition, they can be integrated with other optimization techniques in order to solve any problem that can be formulated. These algorithms are characterized by their ability to solve problems quickly and accurately, without using common mathematical solutions [1, 9]. There are two important phases that govern the structure of a meta-heuristic algorithm, namely exploration and exploitation. As part of the exploration phase, randomization methods are generally used to efficiently search the search space. In contrast, meta-heuristic algorithms focus on the most promising part of a search space in the exploitation phase [3]. An approach to hybrid feature selection is presented in this paper. Then, we introduce a feature selection algorithm based on the equilibrium optimizer, which explicitly considers relevance and redundancy. This study includes the following contributions:

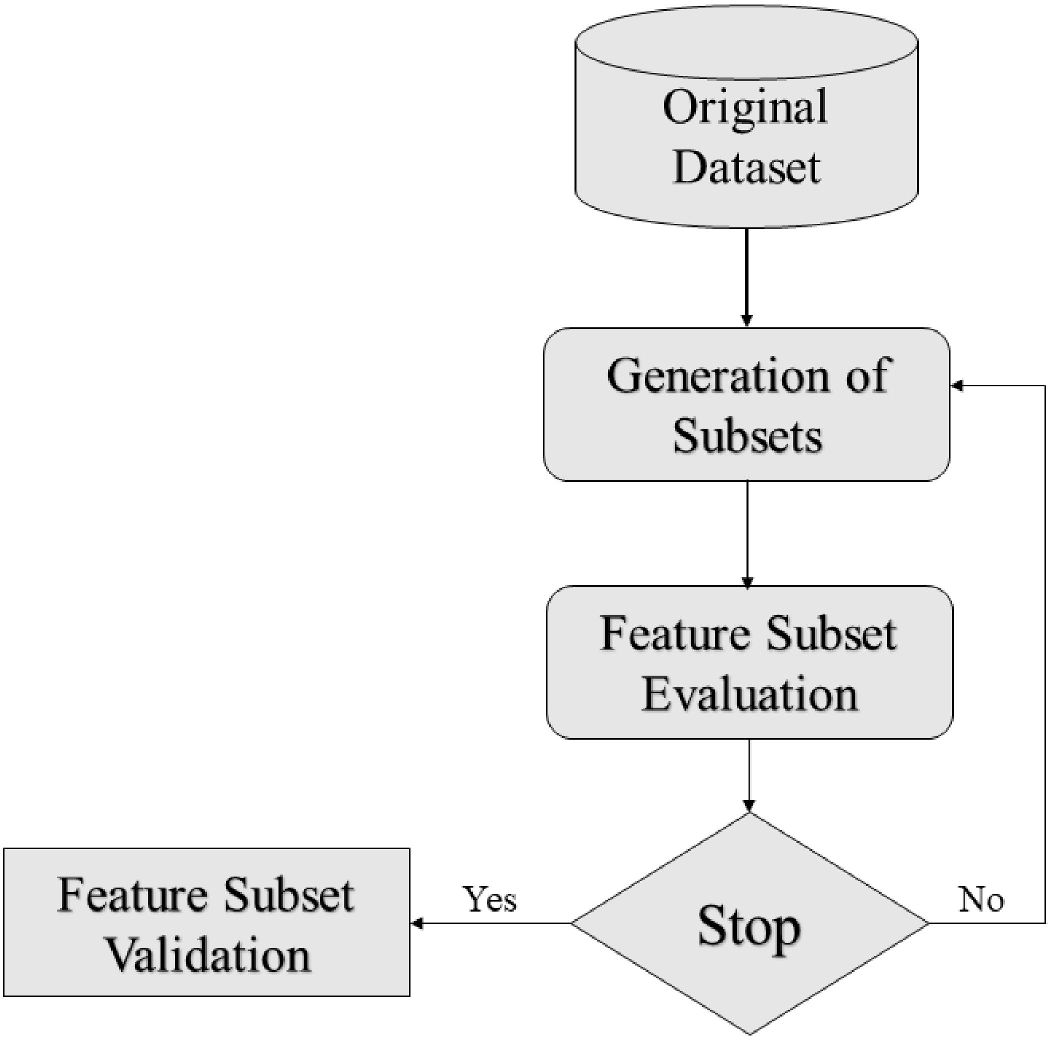

Figure 1: General feature selection process. [1]

1. Avoiding overly related features by examining redundancy.

2. Identifying features that are relevant to the class label feature, then ignoring other features that are unrelated.

3. Introducing an equilibrium optimization algorithm for feature selection.

4. Comparison of the proposed method with five feature selection methods based on 11 UCA datasets.

3 Background

3.1 Filter method

As a principal criterion for variable selection by ordering, filter methods use variable ranking techniques. In practical applications, ranking methods have proven to be successful due to their simplicity. An appropriate ranking criterion is used for scoring variables and a threshold is used for removing variables below it. The ranking method is a filter method because it removes the less relevant variables before classification. This section discusses the two filter methods used in the proposed method [5].

- A) Relevance: According to intuition, a given feature is relevant when it provides information about the class label feature (C) alone or when combined with other variables. As shown in Table 1, there are various definitions of relevance in the literature, including weakly relevant, strongly relevant, and irrelevant features. A strong relevance feature provides information that cannot be replaced by another feature about C. In contrast, weakly relevant features provide information about C, but can be replaced by other features without removing the information they provide. Information about C can be lost if irrelevant features are discarded since they don't provide any information about C [25, 27].

Table 1: Levels of relevance for feature f_i, according to probabilistic framework and mutual information [25, 27].

Relevance level	**Condition**	**Probabilistic approach**	**Mutual information**
Strongly relevant	$\nexists$	$p(C \mid f_i, \neg f_i) \neq p(C \mid \neg f_i)$	$I(f_i; C \mid f_i) \succ 0$
Weakly relevant	$\exists S \subset \neg f_i$	$p(C \mid f_i, \neg f_i) \neq p(C \mid \neg f_i)$ $\wedge$ $p(C \mid f_i, S) \neq p(C \mid S)$	$I(f_i; C \mid f_i) \succ 0$ $\wedge$ $I(f_i; C \mid S) \succ 0$
Irrelevant	$\exists S \subset \neg f_i$	$p(C \mid f_i, S) \neq p(C \mid S)$	$I(f_i; C \mid S) \succ 0$

- B) Redundancy: Redundancy refers to the level of interdependency among two or more features. A feature can be measured in terms of its dependency on a feature subset , by simply using the MI, . Its properties include symmetry, nonlinearity, nonnegative, and not diminishing with the addition of features. As a result of this measure, however, it is difficult to determine concretely which features of S are redundant with one another. Therefore, more elaborate criteria of redundancy, such as Markov blankets and total correlations, should be developed [25, 27].

- C) Spearman correlation coefficient: Feature correlation coefficients must be greater than a predetermined threshold in order to qualify as highly correlated. By using Spearman's Correlation Coefficients, we estimated the correlation coefficients between two features. Each correlation analysis method has different trigger conditions, as indicated in Eq. 3.1. Correlations between two features can be measured using the SCC (both linear and nonlinear correlations). The SCC is proportional to the degree of correlation between two features. Stronger the correlation, the higher the SCC value. When f_i and f_j are highly correlated, the absolute value of SCC between them should exceed k_1 [13].

$$\operatorname{corr}(f_i, f_j) = |\operatorname{SCC}(f_i, f_j)|, n \geq k_1 \tag{3.1}$$

There are two distinct features in dataset that are f_i and f_j, and predefined threshold k_1 represents them.

3.2 Wrapper method

Using a wrapper method, the variable subset is evaluated using the predictor performances as the objective function. Due to the NP-hardness of evaluating 2N subsets, heuristic search algorithms are used to find suboptimal subsets. In order to maximize classification performance, a subset of variables can be selected using a variety of search algorithms.
The Wrapper methods can be divided into Sequential Selection Algorithms and Heuristic Search Algorithms. Feature additions (feature removals) are performed sequentially until a maximum objective function is achieved. As the objective function increases incrementally, the minimum number of features is selected until the maximum is reached. An objective function can be optimized using heuristic search algorithms by examining different subsets. In either case, subsets are generated by searching around in the search space or by solving the optimization problem. Among the e searches, we discuss the equilibrium optimizer here.

3.2.1 Equilibrium Optimization Algorithm

The original equilibrium optimizer algorithm will be briefly described in this section.

- Inspiration: In the original equilibrium optimization (EO), equilibrium and dynamics were estimated by mass control balances. Nonreactive constituent concentration in the control volume is described by its sink mechanisms and various sources. Using the mass balance equation, the mass entering is conserved, the underlying physics is supported, and energy is generated by a control volume. Mass in time is determined by subtracting the amount of mass entered in the system from the amount that leaves the system, based on the generic mass balance equation [6, 20]. Mathematical model: Particles are treated as solutions in the EO, and their concentrations indicate their position. By optimizing, particles update their concentration based on the best solutions, known as equilibrium candidates. Particles (search agents) repeat the process until they reach an equilibrium state, which is the optimal solution. To update each particle, three different terms are used, each representing a rule. Equilibrium concentration, on the other hand, is a solution selected from a pool of equilibrium solutions by chance. In the second term, we have the difference in concentration between the equilibrium state and the particle. Each particle's exploration capability (global search) is improved by this term. The last term is used as a refiner or exploiter of small steps in the solution. The generation rate is referred to by this term [6, 20].

$$X_{k+1} = X_{eq} + (X_k - X_{\mathrm{eq}}) \cdot E + \frac{G}{\lambda V}(1 - E) \tag{3.2}$$

This equation consists of V as the control volume, X_{eq} as the concentration at an equilibrium state, $\wedge$ as the residence time, and k as the current iteration number. The mathematical formula of $\wedge$is defined in Eq. 3.3 [6, 20].

$$\lambda = \frac{Q}{V} \tag{3.3}$$

A volumetric flow rate is defined as Q out and into a control volume. As a result of the exponential term E, the main concentration rule is updated. Exploration and exploitation phases are balanced using this term. The definition is as follows [6, 20]:

$$E = a_1 \operatorname{sign}(r - 0.5) \left[e^{-\lambda t} - 1\right] \tag{3.4}$$

An exploration ability is controlled by constant parameter a_1, while r is a random parameter between 0 and 1. $Sign(0.5)$ specifies the direction of a local search and controls exploitation [6, 20].

$$t = \left(1 - \frac{k}{Max_k}\right)^{a_2 \frac{k}{Max_k}} \tag{3.5}$$

In this case, a_2 is a constant. Using it, exploitation ability can be controlled. One of the most important terms in EO algorithm is the mass generation rate within the volume control G. It is defined as follows [6, 20]:

$$G = G_0 E \tag{3.6}$$

Where:

$$G_0 = GCP\left(X_{eq} - \lambda X\right) \tag{3.7}$$

$$GCP = \begin{cases} 0.5r_1 \times ones(1, \mathrm{dim}), r_2 \geq GP \\ 0, r_2 < GP \end{cases} \tag{3.8}$$

The generation rate control parameter GCP is a constant parameter and the generation probability GP is a constant parameter. In this case, $ones(1, dim)$ is a vector with a length equal to $1dim$ which is initialized with ones. Dim is the number of dimensions in the vector. There are three random parameters r_1, r_2, and r_3 generated between 0 and 1. A comparison of two approaches to feature selection can be seen in Fig. 2.

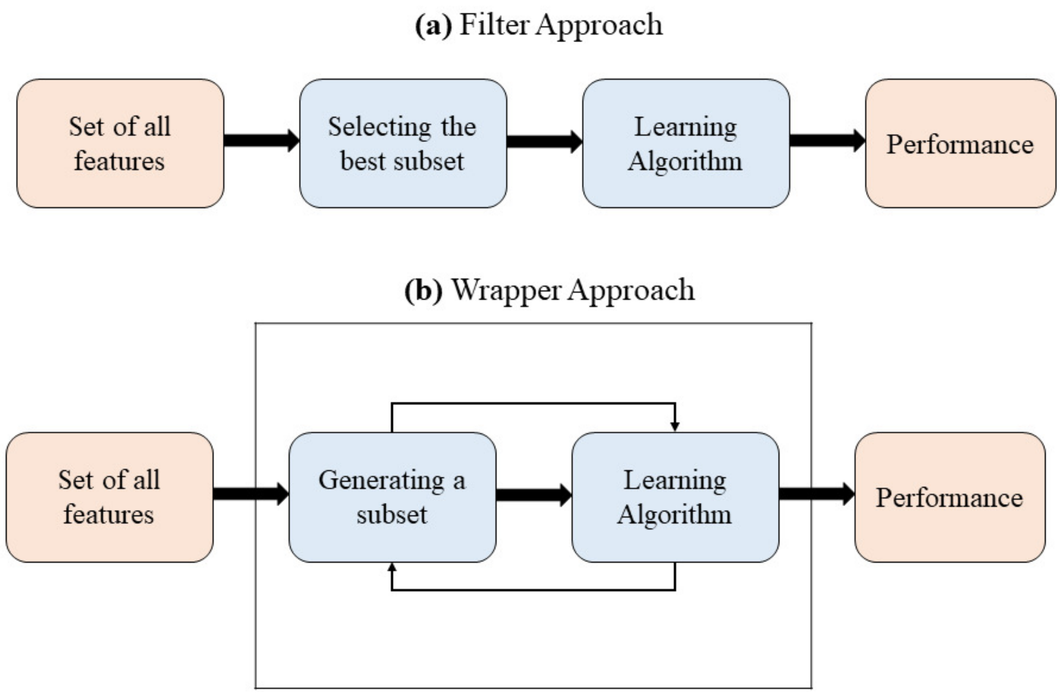

Figure 2: Filter and wrapper approach [19].

Figure 3 shows the flowchart of equilibrium optimization.

4 Related Works

Now, various feature selection methods have been developed including filtering, wrapper, and embedded methods. Based on the filter method, inputs and outputs are taken into account regardless of classification accuracy. A learning method is usually used to select an optimal subset of features in the wrapper method. It is actually on the agenda to combine filter and wrapper methods in embedded methods in order to maximize both methods' advantages. There have also been a number of methods that introduce feature selection techniques using evolutionary algorithms in order to avoid local optima and to reach the optimal solution as quickly as possible. Optimization algorithms have the advantage of balancing exploration and exploitation [3]. For support vector machine (SVM) classification, Unler et al. [26] proposed a hybrid filter–wrapper algorithm using particle swarm optimization (PSO). Based on mutual information, the filter model measures the relevance and redundancy of selected features with respect to their subset. The wrapper model is a modified discrete PSO algorithm. An algorithm that achieves maximum

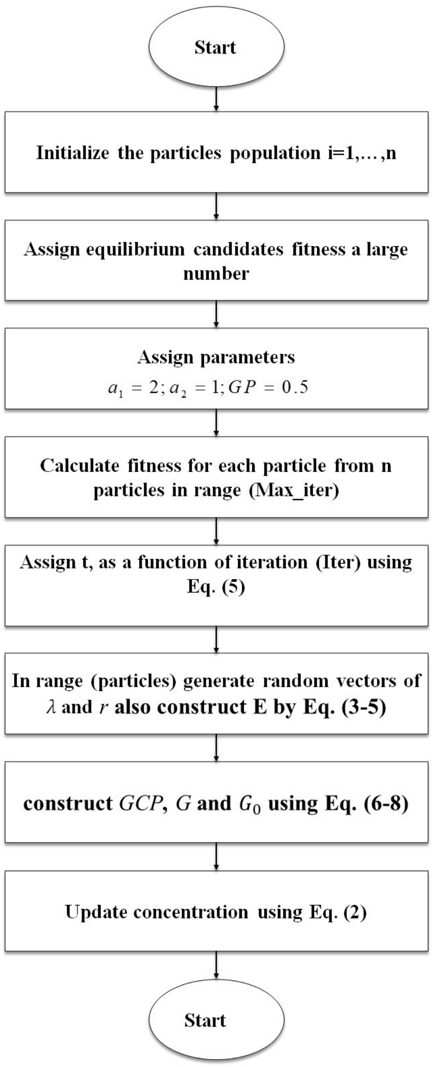

Figure 3: Flowchart of equilibrium algorithm.

relevance with minimum redundancy is called Maximum relevance minimum redundancy PSO (Mr2PSO). Zheng et al. [28] proposed an improved whale optimization algorithm, MPMDIWOA, which is a hybrid feature subset selection algorithm. In the first step, a filter algorithm named maximum Pearson maximum distance (MPMD) is proposed based on Pearson's correlation coefficient and correlation distance. Second, the modified whale optimization algorithm can act as a wrapper algorithm. MVWC and the threshold are introduced before filter and wrapper algorithms are combined. Tubishat et al. [21] proposed two improvements to the WOA algorithm to avoid being trapped in local optima. WOA's initialization phase can be improved by using Elite Opposition-Based Learning (EOBL). Second, each WOA iteration incorporates evolution operators including mutation, crossover, and selection from the Differential Evolution algorithm. Moslehi and Haeri

[18] incorporated evolutionary genetic algorithms (GA) and particle swarm optimization (PSO) into a hybrid filter-wrapper method for feature subset selection. HGP-FS is a method that aims to reduce the computation time and complication required to achieve the best solution to the problem of choosing features from high dimensional datasets. Filters and wrappers are integrated to select datasets with effective characteristics. In Guha et al. [7], whale optimization algorithm (WOA) is an FS technique inspired by humpback whale foraging behavior. WOA embeds its wrapper process into ECWSA, a version of WOA for achieving high classification accuracy using its wrapper process. Using a filter approach, the selected subset is further refined with low computation costs. A hybrid model for feature selection has been suggested by Moorthy and Gandhi [16] for categorizing different datasets of heart disease (HD). Initially, the relevant feature sets were selected by means of a filter method, in particular an analysis of variance. After that, a whale optimization-based evolutionary wrapper will be used to determine the optimal feature sets from the previous feature selection will be introduced. This method is called ANOVA-WO. According to Got et al. [8], a new hybrid filter-wrapper feature selection approach can be achieved using whale optimization algorithm (WOA). Multi-objective optimization is used in the proposed algorithm to optimize filter and wrapper fitness functions simultaneously. A novel multi-population-based particle swarm optimization algorithm based on this theory was presented by Kılıç et al. [12]. Using this method, both random and relief-based initializations are used to generate initial solutions and search solution space simultaneously. A comparison of the number of features selected shows that MPPSO is better than other algorithms. Finch and Bertolini [4] explored the stability of filter feature selection methods in data pipelines using simulations. When cross-validating three feature selection algorithms using six metrics, Borda's method and Borda's method for determining features associated with a binary target outcome were used. We addressed some of the challenges of feature selection by removing irrelevant and repetitive features and accelerating convergence to the optimal solution. Using Dynamic Butterfly Optimization Algorithms based Interaction Maximization (IFS-DBOIM), Tiwari and Chaturvedi [22] proposed a hybrid feature selection approach. Their aim was to overcome poor trade-offs between exploration and exploitation phases and get stuck into an optimal local solution. DBOA and FIM are combined to select the optimal feature subset using the proposed Dynamic Butterfly Optimization Algorithm (DBOA). With fewer iterations, DBOA increases convergence rates. Using Boolean variants of Particle Swarm Optimization (BPSO) in combination with Evolutionary Population Dynamics (EPD), Thaher et al. [23] proposed an efficient feature selection approach. By enhancing exploration abilities, the BPSO avoids local optima obstacles. By repositioning the worst half of the solutions around the best half, the BPSO-EPD discards the worst half of the solutions. We use six natural selection mechanisms to choose our guiding solutions: best-based, tournaments, roulette wheels, stochastic universal sampling, linear ranks, and random-based. Table 2 shows related works on feature selection discussed above.

Table 2: Related works on feature selection

Ref.	**Year**	**Proposed-Method**	**Compared-Methods**	**Learning-algorithm**	**Dataset-used**	**Advantage(s)**	**Disadvantage(s)**
Unler et al. [22]	2011	Mr2PSO	PSO	SVM	6	High classification accuracy, Best computational performance	Dependence on the feature subset construction sequence
Zheng et al. [26]	2018	MPMDIWOA	WOA	SVM	10	Low computational complexity, High accuracy	Poor performance on high-dimensional data
Tubishat et al. [9]	2019	WOA	WOA	SVM	4	High performance in sentiment analysis classification accuracy, Minimum selected feature	Limited performance, the number of datasets is low
Moslehi and Haeri [18]	2020	HGP-FS	GA, PSO	KNN	5	More accurate classification, Remove unsuitable and unessential characteristics	The number of datasets is low, Low accuracy in some datasets
In Guha et al. [7]	2020	ECWSA	WOA	KNN	18	Better performance on most datasets	High computational complexity
Moorthy and Gandhi [16]	2021	ANOVA-WO	WOA	KNN, SVM, NB	5	Better classification accuracy with extensively fewer features	High computational complexity
Got et al. [8]	2021	GPAWOA	WOA	KNN	12	Unique ability to achieve optimal feature set	The crowding distance calculation and the used filter function require more running time
Kılıç et al. [12]	2021	MPPSO	PSO	KNN	29	Optimal performance for high-dimensional datasets	Less efficient for low-dimensional datasets
Finch and Bertolini [4]	2022	simulation study	-	-	-	provide greater insight into the stability of filter feature selection techniques	A specific field is discussed
Tiwari and Chaturvedi [22]	2022	IFS-DBOIM	DBOA	SVM, NB, DT	20	Keeping exploration and exploitation phases in balance, Avoiding redundancy and irrelevancy	Applied dataset characteristics often dictate its generalizability
Thaher et al. [23]	2022	BPSO-EPD	PSO, EPD	KNN, DT	22	Assigns equal chances to each individual (particle) in guiding the poor solutions in order to maintain the diversity of the population	High computational complexity

5 Methodology

A description of the proposed method can be found in this section, which also includes the transfer function, fitness function, and general scheme of the method.

5.1 Convert to Probability

The sigmoid function was used to convert the algorithm to a binary version according to Eq. 5.1. As a result, FS can only be solved for binary values between 0 and 1. Binary vectors represent solutions, where 1 indicates that the corresponding feature has been selected, and 0 indicates that it has not been selected.

$$T(x) = \frac{1}{1 + e^{-x}} \tag{5.1}$$

An example of converting continuous search space to binary can be seen in Fig.4.

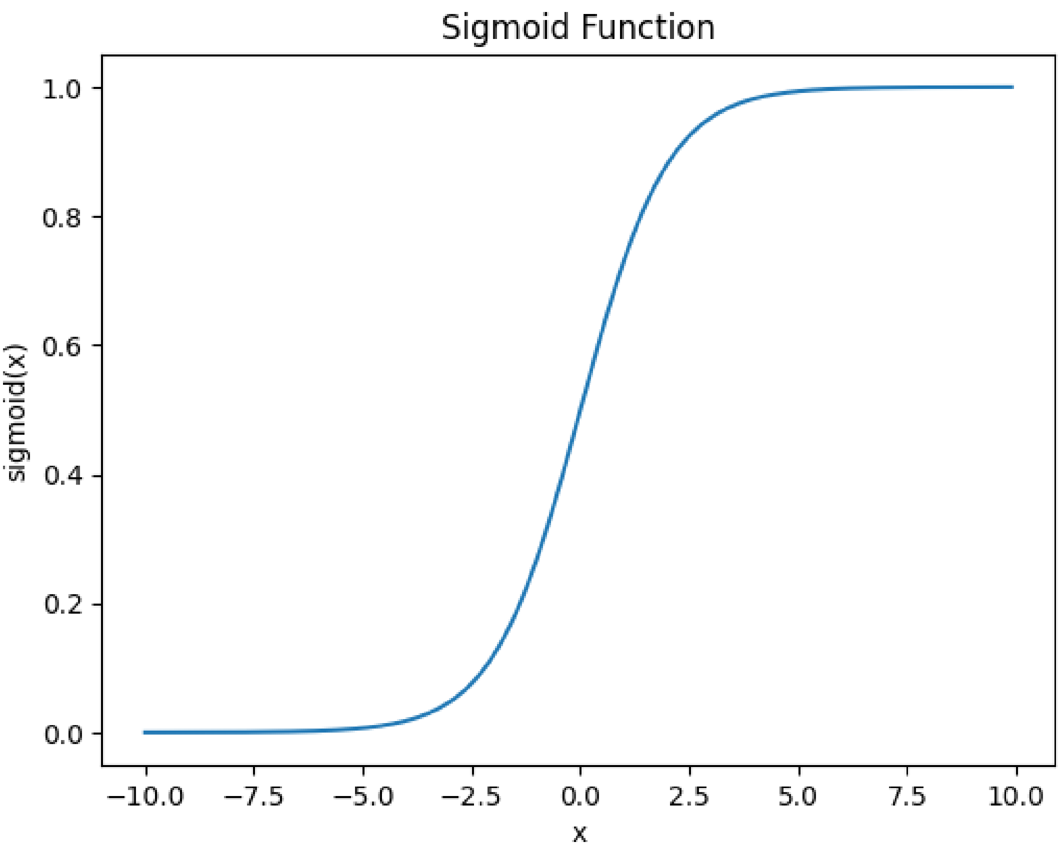

Figure 4: Transfer function for converting continuous search space to binary.

5.2 Fitness Function

The general goal of FS is to maximize classification accuracy and minimize the number of features selected. Both of these objective's conflict with each other. These two objectives can be combined into a single objective problem using Eq. 5.2.

$$\text{Fitness } = \omega\gamma(F) + (1-\omega) \times \text{ Feature ratio} \tag{5.2}$$

In this case, Feature ratio refers to the ratio between the number of selected features and the original dimension of the dataset. is classification error rate of the subset of features selected, weights (w) are represented by the values 0 and 1.

5.3 Proposed Filter Based Binary Equilibrium Optimization (FBBEO)

In order to measure redundancy and remove highly correlated features, we first used Spearman's correlation coefficient with a correlation limit of 0.8 [13]. Those features that provided no useful information about the class label feature were ignored because they were irrelevant or weakly related. To select the optimal subset of features, we then used the unique properties of the equilibrium optimizer, such as avoiding local optima, achieving optimal solutions quickly, and balancing exploration and exploitation. An overview of the proposed method can be seen in Fig. 5.

5.4 Example

Figure 6 illustrates an example of how the FBBEO algorithm is used for feature selection using the proposed approach. A feature subset (i.e., a solution) can be represented by a one-dimensional binary vector containing 11 elements. A '1' indicates that this corresponding feature can be selected, while a '0' indicates the opposite. As a first step, all elements are assigned one, followed by the filter phase and calculating the Spearman correlation. A feature is removed if the correlation exceeds a certain level. Following this phase, the best solution (the best feature set) is found using equilibrium optimization search processes (see section 3.2).

6 Experimental setup and results

The results of the comparison of the proposed method with five binary optimization methods including Genetic Algorithm (GA), Particle Swarm Optimization (BPSO), Golden Rate (BGR), Social Mimic Optimization (BSMO) and Atom Search Optimization (BASO) are given. The proposed method is evaluated using 11 UCA datasets, box plots, and fitness diagrams. For the purpose of evaluating classification accuracy, we applied KNN classifiers [17] to a subset of the entire dataset that was chosen by applying the FS method to the entire dataset. According to [14], K= 5 is the recommended value for the KNN classifier. There are 0.8 of instances in a training dataset, while 0.2 are in a testing dataset.

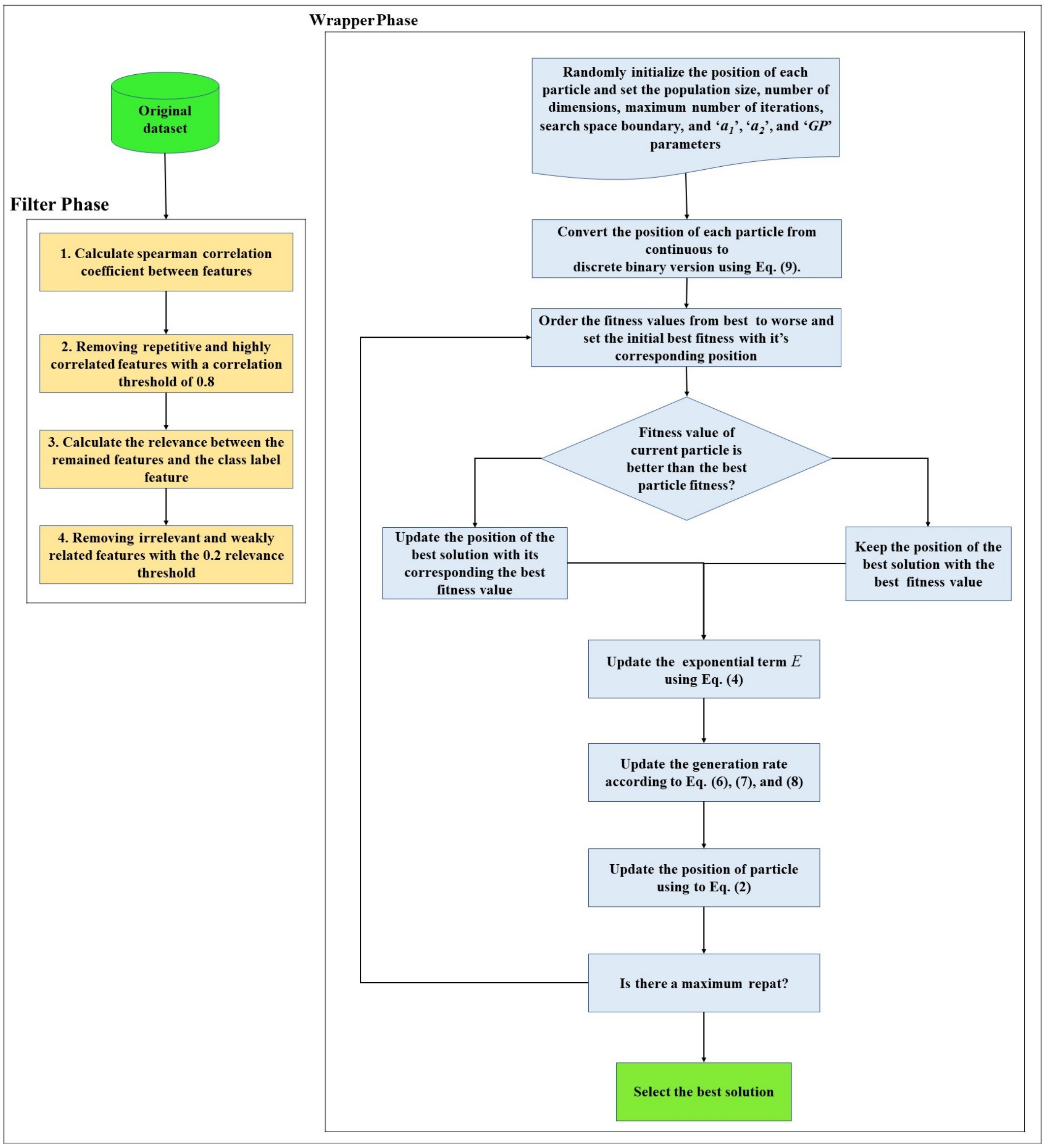

Figure 5: Flowchart of FBBEO.

6.1 Control parameters

Table 3 shows the control parameters used in proposed method and the other methods. Based on this table in the proposed method FBBEO, Omega is weightage for number of features and accuracy, Max-Iter shows the number of iterations also Pop-Size shows the initial population size. a_1 is a constant control parameter for controlling exploration ability also a_2 is a constant that is used to control the exploitation ability. Generation probability (GP) is a constant parameter.

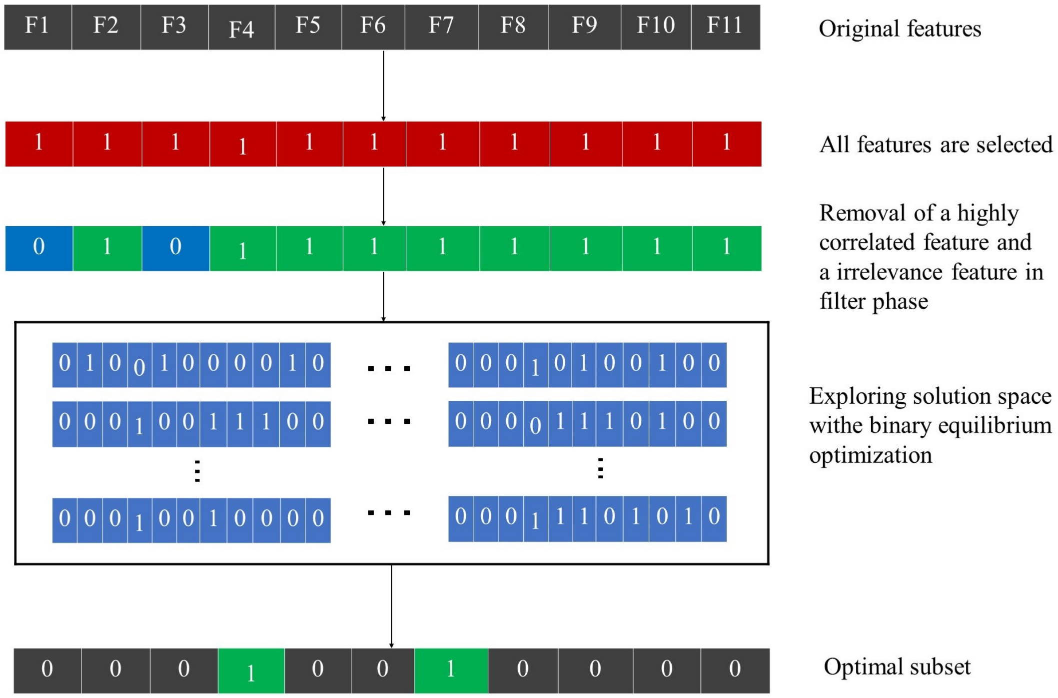

Figure 6: Example of Filter Based Binary Equilibrium Optimization (FBBEO) on Breast Cancer dataset.

Table 3: Control parameters.

Algorithm	**Parameters**
FBBEO	Omega = 0.9, Max-Iter = 30, Pop-Size = 10, a_1 = 2, a_2 = 1, GP = 0.5
BGA	Pop-Size = 10, Max-Iter = 20, Crossover-prob = 0.6, Muprob-min = 0.01, Muprob-max = 0.3
BPSO	Pop-Size = 20, Max-Iter = 30, C1, C2 = 2, WMAX = 0.9, WMIN = 0.4
BGR	Golden = (1 + 5 ** 0.5) / 2, pop-Size = 10, max-Iter = 10
BSMO	Pop-Size = 10, Max-Iter = 20
BASO	Pop-size = 10, Max-Iter = 30, α = 50, β = 0.2

6.2 Dataset description

The evaluation datasets are listed in Table 4. It is possible to see how many features each dataset contains before and after filtering. During the filter phase, it determines whether there is redundancy between features so duplicate features can be excluded, as well as the correlation between features and category features so that generally unrelated features can be ignored. In addition to being more accurate in selecting the subset of features, the proposed method is also less computationally complex.

Table 4: Description of datasets before and after filter phase.

Datasets	**No-of sample**	**No-of features before applying filter**	**No-of features after applying filter**	**No-of class**	**Domain**
Algerian-forest-fires	244	14	8	2	Life
BreastCancer	698	11	9	2	Life
BreastEW	568	31	15	2	Life
dataR2	116	10	4	2	Life
lung-cancer	32	57	28	2	Life
pd-speech-features	756	755	82	2	Life
PenglungEW	73	326	148	7	Physical
sobar-72	72	20	16	2	Life
Sonar	208	61	18	2	Physical
Wholesale-customers-data	440	8	4	2	Business
Zoo	101	17	10	2	Life

6.3 Evaluation of fitness value

The fitness value metric is one of the evaluation metrics used to determine which method performs better than others. According to this metric, if the fitness value of the algorithm is lower than the other methods and the algorithm reaches the optimal fitness faster, it performs better than the other methods. Table 5 shows that FBBEO is preferred over other comparison methods for determining fitness value. In all tables, bold number denotes the best performance. Comparatively, the proposed method obtained a better fitness value for 6 datasets (55%) than other methods, compared to 3 datasets (27%) for BGR, 18% for BSMO, and 0% for other methods. Table 6 presents the average fitness value. BPSO performed better in only one dataset (0.09%), while the FBBEO and BSMO performed better in five datasets (45%). Additionally, Table 7 shows the worst fitness value. Based on comparisons with the proposed method and other methods, BSMO performes relatively better in six datasets (55%), followed by FBBEO in four datasets (36%), and then PSO with 0.9%. Other methods are often stuck in the local optimum, so they have higher fitness values. In Fig. 7, the proposed method and other comparison methods are shown in 30 iterations. There are six datasets where the proposed method has lower fitness value and reached the optimal solution faster.

Table 5: Comparison of the best fitness value.

Datasets	**FBBEO**	**BGR**	**BPSO**	**BGA**	**BSMO**	**BASO**
Algerian-forest-fires	**0.0142**	0.0404	0.0346	0.0635	0.0219	0.0230
BreastCancer	**0.0378**	0.0542	0.0482	0.0571	0.0499	0.0503
BreastEW	0.0372	0.0374	0.0350	0.0771	**0.0350**	0.0396
dataR2	0.1458	**0.1020**	0.1541	0.2437	0.1319	0.1509
lung-cancer	**0.0133**	0.0455	0.1535	0.1535	0.1553	0.0499
pd-speech-features	0.0925	0.1096	0.0993	0.1401	**0.0795**	0.1110
PenglungEW	**0.0081**	0.0475	0.0240	0.1470	0.0393	0.0181
sobar-72	**0.0066**	0.0236	0.0394	0.1857	0.0421	0.0315
Sonar	**0.0722**	0.1157	0.0879	0.1307	0.1083	0.0952
Wholesale-customers-data	0.1458	**0.0718**	0.1125	0.1318	0.0779	0.0837
Zoo	0.0984	**0.0468**	0.0967	0.1154	0.0624	0.0908

Table 6: Comparison of the average fitness value.

Datasets	FBBEO	BGR	BPSO	BGA	BSMO	BASO
Algerian-forest-fires	**0.0164**	0.0644	0.0369	0.0818	0.0219	0.0230
BreastCancer	**0.0384**	0.0716	0.0489	0.0783	0.0499	0.0514
BreastEW	0.0377	0.0596	0.0385	0.1007	**0.0350**	0.0396
dataR2	0.1480	0.1654	0.1541	0.4260	**0.1319**	0.1509
lung-cancer	**0.0148**	0.1274	0.1614	0.1614	0.1553	0.0499
pd-speech-features	0.0951	0.1261	0.0993	0.1631	**0.0795**	0.1110
PenglungEW	**0.0082**	0.1173	0.0278	0.1733	0.0393	0.0181
sobar-72	**0.0082**	0.0576	0.0410	0.2428	0.0421	0.0315
Sonar	0.0882	0.1648	**0.0879**	0.1872	0.1083	0.0952
Wholesale-customers-data	0.1458	0.1016	0.1125	0.1792	**0.0779**	0.0872
Zoo	0.1150	0.0997	0.0967	0.1702	**0.0624**	0.0937

Table 7: Comparison of the worst fitness value.

Datasets	FBBEO	BGR	BPSO	BGA	BSMO	BASO
Algerian-forest-fires	0.0469	0.0923	0.0865	0.1212	**0.0219**	0.0230
BreastCancer	**0.0439**	0.0939	0.0692	0.1292	0.0499	0.0567
BreastEW	0.0522	0.0823	0.0698	0.1073	**0.0350**	0.0396
dataR2	0.1791	0.2958	0.1541	0.4875	**0.1319**	0.1509
lung-cancer	**0.0199**	0.3178	0.2723	0.2723	0.1553	0.0499
pd-speech-features	0.1085	0.1441	0.0993	0.1800	**0.0795**	0.1110
PenglungEW	**0.0102**	0.1871	0.0811	0.1904	0.0393	0.0181
sobar-72	**0.0133**	0.0868	0.0882	0.2819	0.0421	0.0315
Sonar	0.1092	0.2319	**0.0879**	0.2116	0.1083	0.0952
Wholesale-customers-data	0.1458	0.1394	0.1125	0.2263	**0.0779**	0.0898
Zoo	0.1301	0.1465	0.0967	0.1964	**0.0624**	0.0971

6.4 Evaluation of classification accuracy and selected features

A comparison of the classification accuracy of the proposed method with that of five other methods is shown in Table 8. Among 11 datasets, the proposed method achieved the best accuracy in 7 datasets (64%); this value is also 64% for BGR, BPSO in 4 datasets (36%), BGA in 3 datasets (27%), and BSMO in 3 datasets (27%). A total of four datasets (36%) and five datasets (45%) showed the best accuracy obtained by BASO, indicating its superiority. A comparison of the number of selected features by the proposed method versus the other method is shown in Table 9. Based on these tables, it can be seen that the proposed method achieved a minimum of selected features in all datasets (100%), which is impressive. With the proposed FBBEO, redundant and irrelevant features have been removed from the data set, which significantly reduces the number of selected features. Compared to BGR, BPSO, BGA, the proposed FBBEO method was equal in only 0.9% of the data set, whereas the other two methods were equal in zero, proving the proposed method's absolute superiority. In other words, this method is more accurate than others since it reduces the computational cost and complexity, as well as avoiding the local optimum and prominent features of the balance optimizer.

A boxplot of all datasets of the proposed FBBEO (with KNN) is shown in Fig. 8. In data analysis, a boxplot is a popular way to visualize qualitative and quantitative data summaries. A box plot is another method to evaluate the performance of a method. Box plots display the quartiles, minimums, maximums, and medians of a dataset. Each method behaves differently according to the dataset [11]. A comparison of the proposed

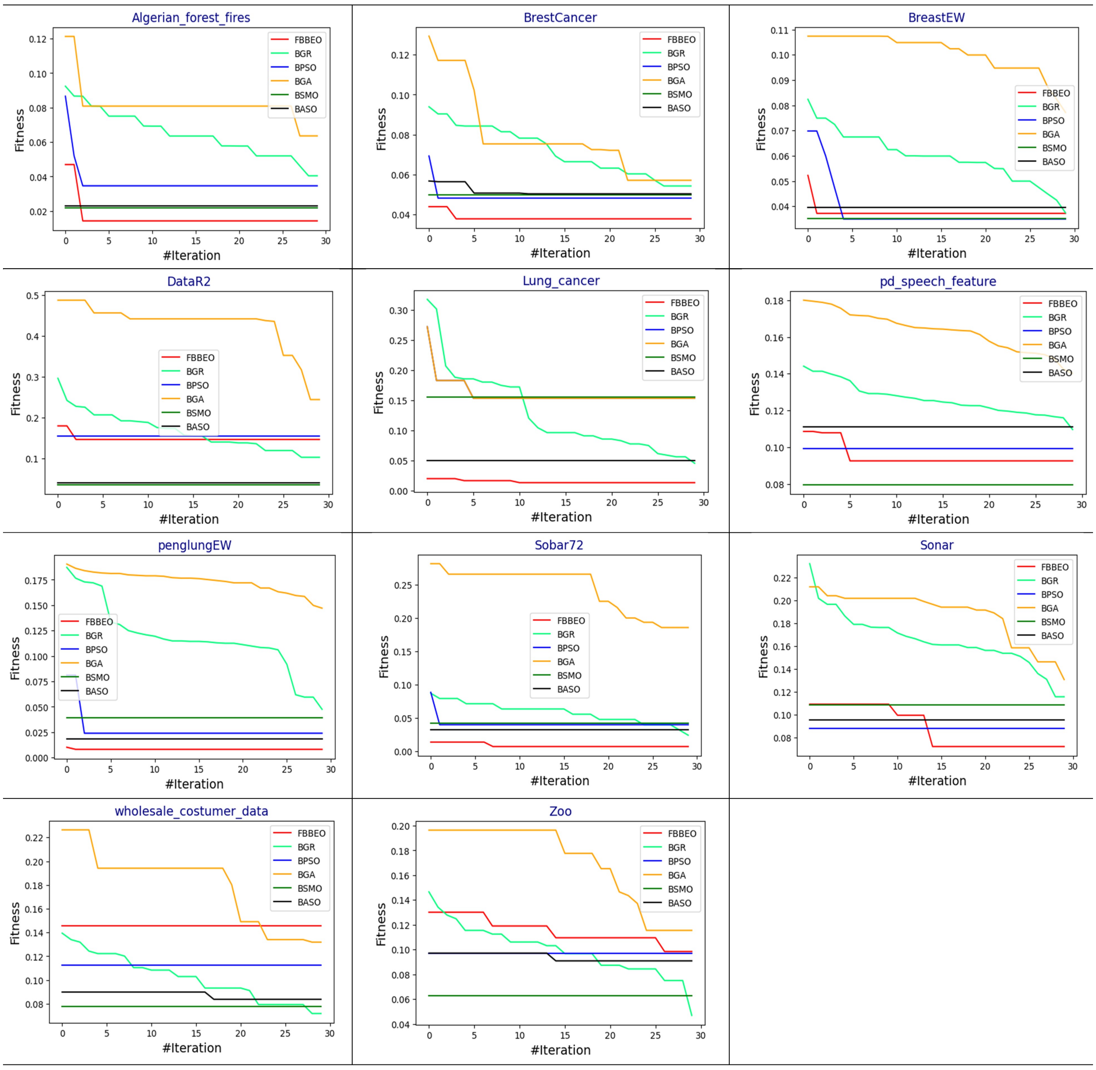

Figure 7: Fitness values using FBBEO and other methods (KNN classifier).

Table 8: Classification accuracy obtained by FBBEO and other methods.

Datasets	**FBBEO**	**BGR**	**BPSO**	**BGA**	**BSMO**	**BASO**
Algerian-forest-fires	**1.0**	**1.0**	**1.0**	0.9795	0.9795	**1.0**
BreastCancer	0.9857	0.9785	0.9785	**0.9857**	**0.9857**	0.9771
BreastEW	**1.0**	**1.0**	**1.0**	0.9649	0.9736	0.9929
dataR2	0.8750	**0.9583**	0.9166	0.7916	0.8750	0.9310
lung-cancer	**1.0**	**1.0**	0.8571	0.8571	0.8571	**1.0**
pd-speech-features	**0.9342**	0.9276	0.9078	0.8947	0.9078	0.9206
PenglungEW	**1.0**	**1.0**	**1.0**	0.8666	**1.0**	**1.0**
sobar-72	**1.0**	**1.0**	**1.0**	0.8666	**1.0**	**1.0**
Sonar	0.9523	0.9523	0.9523	**0.9761**	0.9285	0.9615
Wholesale-customers-data	0.8750	0.9431	0.9431	0.9431	0.9090	**0.9636**
Zoo	0.9523	**1.0**	0.9523	**1.0**	**1.0**	**1.0**

Table 9: The number of selected features by FBBEO and other methods.

Datasets	**FBBEO**	**BGR**	**BPSO**	**BGA**	**BSMO**	**BASO**
Algerian-forest-fires	**1**	2	3	**1**	2	3
BreastCancer	**2**	**2**	**2**	3	4	3
BreastEW	**4**	6	7	5	5	10
dataR2	**1**	4	5	4	4	8
lung-cancer	**4**	15	12	15	30	28
pd-speech-features	**27**	257	106	311	114	299
PenglungEW	**11**	113	41	73	22	59
sobar-72	**1**	4	5	2	5	6
Sonar	**5**	20	19	28	35	19
Wholesale-customers-data	**1**	2	3	3	**1**	3
Zoo	**4**	6	6	8	10	6

FBBEO with other methods can be seen in Fig. 8. Consequently, FBBEO has a better average accuracy in 7 (64%) datasets. BSMO's performance was poor in four data sets (36%), while BGA's accuracy was lower in six (55%). The proposed method can be considered superior to other methods in solving FS problems and selecting the optimal subset of features after this analysis and evaluation.

7 Conclusion and future work

A significant role will be played by the feature selection, which will increase the algorithm's efficiency by removing unrelated features. In this paper we proposed a hybrid filter-wrapper stratgey based on Equilibrium Optimization (EO). In order to select the feature subset, a composite measure of feature relevance and redundancy is used. The wrapper model is based on binary equilibrium optimization (BEO). This proposed method achieves efficiency and accuracy by combining filters and wrappers. The experiment used 11 UCI standard datasets and five feature selection approaches. For 55% of the data sets, the proposed method showed superiority in terms of fitness, while for 64% of data sets, it showed superiority in classification accuracy. Due to the fact that the proposed method has selected a much smaller number of features, this shows the method's superiority. Feature selection methods can be improved in the future by using other ranking-based feature selection or additional constraints. In addition, we are interested in using different learning algorithms with evolutionary computation.

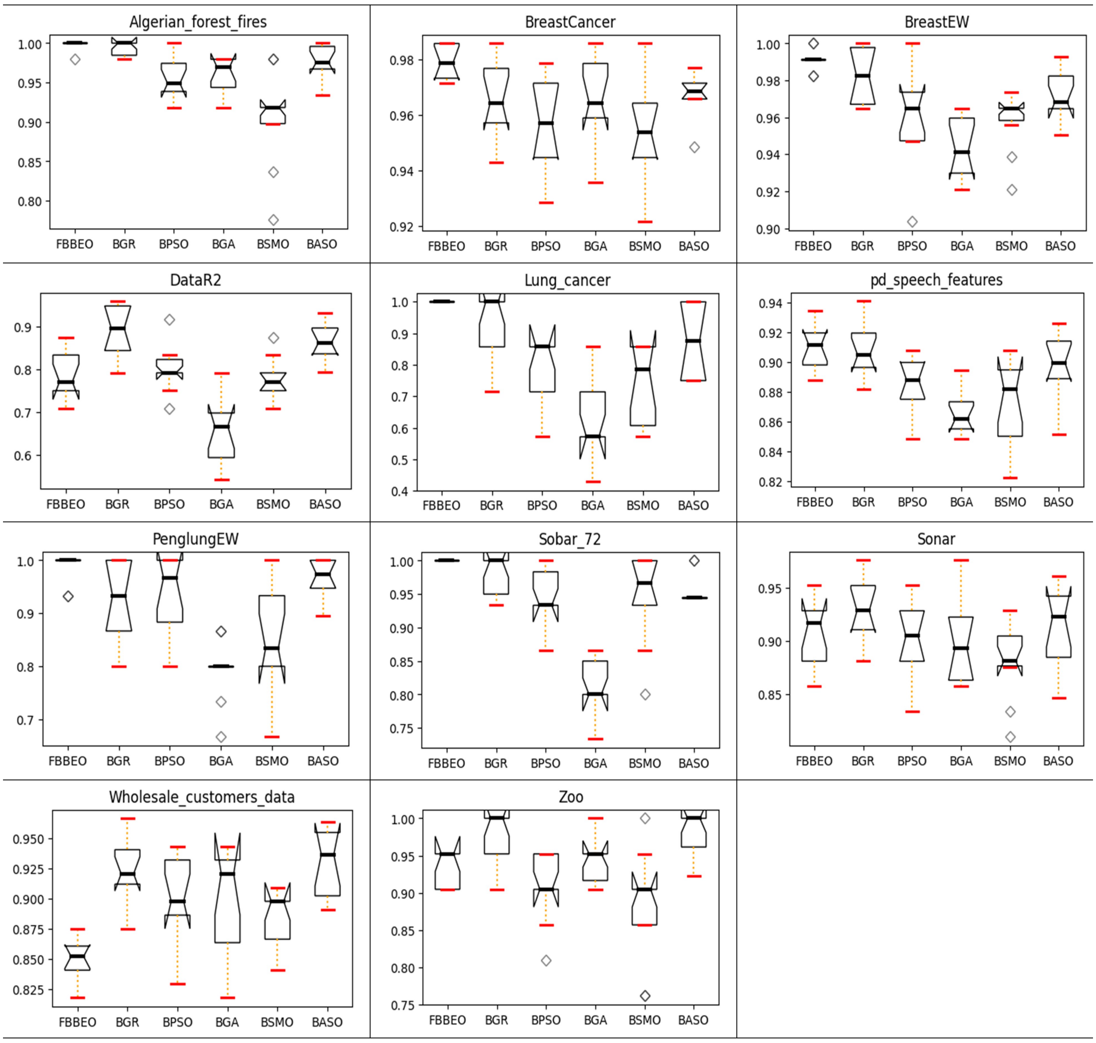

Figure 8: Boxplot of FBBEO and the other methods.

References

[1] Akinola, O.O., Ezugwu, A.E., Agushaka, J.O., Zitar, R.A. and Abualigah, L., Multi-class feature selection with metaheuristic optimization algorithms: a review. Neural Computing and Applications, (2022), 1-40.

[2] Agrawal, P., Ganesh, T. and Mohamed, A.W., Chaotic gaining sharing knowledge-based optimization algorithm: an improved metaheuristic algorithm for feature selection. Soft Computing, 25(14), (2021), 9505-9528.

[3] Anand, P. and Arora, S., A novel chaotic selfish herd optimizer for global optimization and feature selection. Artificial Intelligence Review, 53(2), (2020), 1441-1486.

[4] Bertolini, R. and Finch, S.J., Stability of filter feature selection methods in data pipelines: a simulation study. International Journal of Data Science and Analytics, (2022), 1-24.

[5] Chandrashekar, G. and Sahin, F., A survey on feature selection methods. Computers and Electrical Engineering, 40(1), (2014), 16-28.

[6] Faramarzi, A., Heidarinejad, M., Stephens, B. and Mirjalili, S.,. Equilibrium optimizer: A novel optimization algorithm. Knowledge-Based Systems, 191, (2020), 105190.

[7] Guha, R., Ghosh, M., Mutsuddi, S., Sarkar, R. and Mirjalili, S., Embedded chaotic whale survival algorithm for filter–wrapper feature selection. Soft Computing, 24(17), (2020), 12821-12843.

[8] Got, A., Moussaoui, A. and Zouache, D., Hybrid filter-wrapper feature selection using whale optimization algorithm: A multi-objective approach. Expert Systems with Applications, 183, (2021), 115312.

[9] Hamidzadeh, J., Feature selection by using chaotic cuckoo optimization algorithm with levy flight, opposition-based learning and disruption operator. Soft Computing, 25(4), (2021), 2911-2933.

[10] Hussain, K., Neggaz, N., Zhu, W. and Houssein, E.H., An efficient hybrid sine-cosine Harris hawks optimization for low and high-dimensional feature selection. Expert Systems with Applications, 176, (2021), 114778.

[11] Ibrahim, R.A., Abd Elaziz, M., Ewees, A.A., El-Abd, M. and Lu, S., New feature selection paradigm based on hyper-heuristic technique. Applied Mathematical Modelling, 98, (2021), 14-37.

[12] Kılıç, F., Kaya, Y. and Yildirim, S., A novel multi population based particle swarm optimization for feature selection. Knowledge-Based Systems, 219, (2021), 106894.

[13] Liu, Y., Zou, X., Ma, S., Avdeev, M. and Shi, S., Feature selection method reducing correlations among features by embedding domain knowledge. Acta Materialia, 238, (2022), 118195.

[14] Mafarja, M., Qasem, A., Heidari, A.A., Aljarah, I., Faris, H. and Mirjalili, S., Efficient hybrid nature-inspired binary optimizers for feature selection. Cognitive Computation, 12(1), (2020), 150-175.

[15] Mansouri, N., Khayati, G.R., Mohammad Hasani Zade, B., Khorasani, S.M.J. and Kafi Hernashki, R., A new feature extraction technique based on improved owl search algorithm: a case study in copper electrorefining plant. Neural Computing and Applications, 34(10), (2022), 7749-7814.

[16] Moorthy, U. and Gandhi, U.D., A novel optimal feature selection technique for medical data classification using ANOVA based whale optimization. Journal of Ambient Intelligence and Humanized Computing, 12(3), (2021), 3527-3538.

[17] Murugan, A., Nair, S.A.H. and Kumar, K.P., Detection of skin cancer using SVM, random forest and kNN classifiers. Journal of medical systems, 43(8), (2019), 1-9.

[18] Moslehi, F. and Haeri, A., A novel hybrid wrapper–filter approach based on genetic algorithm, particle swarm optimization for feature subset selection. Journal of Ambient Intelligence and Humanized Computing, 11(3), (2020), 1105-1127.

[19] Naik, A.K., Kuppili, V. and Edla, D.R., Efficient feature selection using one-pass generalized classifier neural network and binary bat algorithm with a novel fitness function. Soft Computing, 24(6), (2020), 4575-4587.

[20] Sayed, G.I., Khoriba, G. and Haggag, M.H., A novel chaotic equilibrium optimizer algorithm with s-shaped and v-shaped transfer functions for feature selection. Journal of Ambient Intelligence and Humanized Computing, 13(6), (2022), 3137-3162.

[21] Tubishat, M., Abushariah, M.A., Idris, N. and Aljarah, I., Improved whale optimization algorithm for feature selection in Arabic sentiment analysis. Applied Intelligence, 49(5), (2019), 1688-1707.

[22] Tiwari, A. and Chaturvedi, A., A hybrid feature selection approach based on information theory and dynamic butterfly optimization algorithm for data classification. Expert Systems with Applications, 196, (2022), 116621.

[23] Thaher, T., Chantar, H., Too, J., Mafarja, M., Turabieh, H. and Houssein, E.H., Boolean Particle Swarm Optimization with various Evolutionary Population Dynamics approaches for feature selection problems. Expert Systems with Applications, 195, (2022), 116550.

[24] Unler, A., Murat, A. and Chinnam, R.B., mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification. Information Sciences, 181(20), (2011), 4625-4641.

[25] Vergara, J.R. and Estévez, P.A., A review of feature selection methods based on mutual information. Neural computing and applications, 24(1), (2014), 175-186.

[26] Wang, D., Chen, H., Li, T., Wan, J. and Huang, Y., A novel quantum grasshopper optimization algorithm for feature selection. International Journal of Approximate Reasoning, 127, (2020), 33-53.

[27] Wang, L., Jiang, S. and Jiang, S., A feature selection method via analysis of relevance, redundancy, and interaction. Expert Systems with Applications, 183, (2021), 115365.

[28] Zheng, Y., Li, Y., Wang, G., Chen, Y., Xu, Q., Fan, J. and Cui, X., A novel hybrid algorithm for feature selection based on whale optimization algorithm. IEEE Access, 7, (2018), 14908-14923.