



A hybrid filter-wrapper feature selection using Fuzzy KNN based on Bonferroni mean for medical datasets classification: A COVID-19 case study



Amukta Malyada Vommi^{a,*}, Tirumala Krishna Battula^b

^a ECE Department, Jawaharlal Nehru Technological University, Kakinada 533005, A.P., India

^b ECE Department, Jawaharlal Nehru Technological University, Kakinada 533005, A.P., India

ARTICLE INFO

Keywords:

Relieff
Fuzzy entropy measures
Binary equilibrium optimizer
Feature selection
Medical datasets
COVID-19 case study

ABSTRACT

Several feature selection methods have been developed to extract the optimal features from a dataset in medical datasets classification. Creating an efficient technique has become a challenge because of the high dimensions, noise, and redundant information. In this paper, we propose a hybrid filter-wrapper approach for feature selection. An ensemble of filter methods, Relieff and Fuzzy Entropy (RFE) is developed, and the union of top-n features from each method are considered. The Equilibrium Optimizer (EO) technique is combined with Opposition Based Learning (OBL), Cauchy Mutation operator and a novel search strategy to enhance its capabilities. The OBL strategy improves the diversity of the population in the search space. The Cauchy Mutation operator enhances its ability to evade the local optima during the search, and the novel search strategy improves the exploration capability of the algorithm. This enhanced form of EO is integrated with eight time-varying S and V-shaped transfer functions to convert the solutions into binary form, Binary Enhanced Equilibrium Optimizer (BEE). The features from the ensemble are given as input to the Binary Enhanced Equilibrium Optimizer to extract the essential features. Fuzzy KNN based on Bonferroni mean is used as the learning algorithm. Twenty-two benchmark datasets and four microarray datasets are used to test the algorithm's efficiency. This method is also applied to a COVID-19 case study. The results demonstrate the superiority of the proposed approach, RFE-BEE, among other methods in terms of fitness values, accuracy, precision, sensitivity, and F-measure, among several other state-of-the-art algorithms. RFE-BEE can be applied to various biomedical, computer vision and engineering applications such as electromyography pattern recognition, COVID-19 diagnosis, face recognition and fault diagnosis.

1. Introduction

Data mining has been at the forefront of biological data analysis. With rapid development in the biological sciences, many datasets with high complexity and dimensions are created, which makes data analysis difficult. The traditional classifiers alone are unable to handle these issues. Hence the FS approach has become necessary to remove the noise and unwanted attributes from the original data. These approaches not only reduce the number of features but also enhance the model's performance.

FS is mainly used as a pre-selector to reduce the number of features and discard the redundant and irrelevant data from the given data (Chen and Chen, 2015). These methods help to find the significant features for early identification and diagnosis of diseases. Efficient FS techniques are required to improve the quality of the feature subset obtained.

FS methods can be classified into two types, namely, the ranking methods and the subset selection methods. Ranking or feature weighting methods are used to assess the information contained in each feature, and weights are allocated depending on their degree of relevance (Dessì & Pes, 2015). Generally, evaluation measures like chi-square statistic, correlation coefficient, signal-to-noise statistic and information gain are evaluated for each feature to acquire their ranks. Features with the highest ranks are chosen with a threshold cut-off to obtain an optimal subset (Dessì & Pes, 2015). The main disadvantage of this method is that it can eliminate only the irrelevant features and not the redundant features. These are the univariate methods. They are simple and competitive compared with the multivariate methods (Drotar et al., 2015).

Subset selection methods use heuristic search strategies to obtain a small-dimension feature subset. Various optimization algorithms have been proposed by researchers, which are used in heuristic search

* Corresponding author.

E-mail address: malyada1212@gmail.com (A.M. Vommi).

strategies. Some of the optimization algorithms include Equilibrium Optimizer (EO) (Faramarzi et al., 2020), Particle Swarm Optimization (PSO) (Eberhart & Kennedy, 1995), VOMMI Algorithm (Vommi & Vemula, 2018), Slime Mould Algorithm (SMA) (Li et al., 2020), Genetic Algorithm (GA) (Ghosh et al., 2019), Grey Wolf Optimizer (Mirjalili et al., 2014), Whale Optimization Algorithm (Mafarja & Mirjalili, 2018) and Ant Colony Optimization (ACO) (Dorigo et al., 2006). These methods involve generating and evaluating the subset which accomplishes the good measure. This operation is continued until the stopping condition is achieved (Yu & Liu, 2004). These are the multivariate techniques which detect the correlations among the features related to the data. These methods can remove redundant and irrelevant attributes but are computationally expensive (Spolaor et al., 2013).

FS methods are mainly classified into four types (Yu & Liu, 2004):

Filter-based approaches: A metric is selected to find the redundant data in this approach. These are known as classifier-independent approaches, as they extract attributes without using a specific learning algorithm. These approaches use the data's intrinsic qualities for ranking the attributes. Several evaluation metrics like information, consistency, distance, and dependency are used, and the best features are selected depending on the feature scores. These methods are economical, fast, and computationally simple and hence primarily used for solving problems of datasets with high dimensions.

Wrapper-based approaches: Wrapper methods iteratively search the entire feature space and generate an optimal feature subset. These techniques use a specific classifier as a learning algorithm. As the feature space grows, the computational complexity increases which is one of the limitations of this technique. As a specific learning algorithm is used, these approaches produce better results than filter-based approaches.

Hybrid approach: Hybrid approaches utilize filter and wrapper techniques to improve the model's performance and produce superior results.

Embedded techniques: These techniques include the FS method as a part of the optimization objective of the classification. Hence, the classifier's learning algorithm performs feature selection and classification simultaneously. Some of these techniques include Random Forest, Lasso and Extra Tree. These techniques are less prone to overfitting.

No individual approach, either filter or wrapper-based, can effectively tackle high-dimensional datasets. Due to this fact, researchers utilize hybrid methods to attain optimal results. Even though these methods work effectively, determining a practical algorithm for FS is still an open issue. According to the No-Free-Lunch theorem (Wolpert & Macready, 1997), a single optimization technique cannot solve all FS problems. This encouraged many researchers to present new algorithms with enhanced performances for years.

The limitations mentioned above motivated us to propose a novel hybrid feature selection algorithm integrated with Fuzzy KNN based on Bonferroni-Mean classifier (BMFK) as an induction algorithm. The proposed method combines ReliefF and Fuzzy Entropy methods as a filter method and the Binary Enhanced Equilibrium Optimizer as a wrapper technique. The Equilibrium Optimizer is a physics-based meta-heuristic technique inspired by mass balance models. Many researchers used the EO algorithm to solve various problems like image segmentation problems (Abdel-Basset et al., 2020), FS problems (Vommi & Battula, 2023) and engineering design problems (Gupta et al., 2020).

Due to its simplicity, most previous feature selection works have used KNN as a learning algorithm. The KNN classifier works on the principle of majority voting. Hence, if a dataset is imbalanced, the classes with a large number of instances dominate the prediction of the sample from the test set. A solution to this problem is to use the mean values calculated from the classes. Then, the prediction of the class is

based on mean values rather than the nearest neighbours. By following this procedure, the classes with the highest number of instances will not dominate those with the lowest number of instances (Kumbure et al., 2020). The advantage of the Fuzzy KNN classifier is that it performs classification through weighted voting using a membership degree. These advantages encouraged us to use the BMFK classifier as a learning algorithm.

The contributions of the paper are as follows:

- A novel hybrid FS technique (RFE-BEE) is proposed, which combines the filter and wrapper-based FS techniques.
- An ensemble of filter methods, ReliefF and Fuzzy Entropy is developed to remove irrelevant features and reduce the search space.
- The Binary Enhanced EO method is developed using opposition-based learning, mutation and a novel search strategy which improves the performance of EO (wrapper method).
- Unlike other FS approaches, Fuzzy KNN based on Bonferroni mean classifier has been used as the learning method.

The RFE-BEE method is evaluated by conducting experiments on 22 standard medical datasets and four microarray datasets. Furthermore, this method is applied to a COVID-19 case study. Six measures have been used to test the efficiency of the RFE-BEE algorithm: fitness values, accuracy, precision, sensitivity, F-measure, and the number of features selected. The proposed algorithm achieved outstanding results compared with three metaheuristic algorithms and seven other existing methods in the literature.

The remaining paper is organized as follows: Some of the feature selection works are discussed in Section 2. The preliminaries are explained in Section 3. The proposed RFE-BEE approach is presented in Section 4. A comparison of the RFE-BEE feature selection approach with numerous FS algorithms is given in Section 5. Section 6 tests the RFE-BEE approach on microarray datasets and a COVID-19 case study. Finally, Section 7 concludes the work.

2. Related work

Feature Selection (FS) is mainly used as a pre-selector to determine the optimal subset of relevant features. Efficient FS methods are required to improve the quality of the feature subset obtained. This section reviews recent FS works, including the wrapper and hybrid FS algorithms. Over the last several decades, many metaheuristic algorithms have been developed to solve FS problems. These algorithms are mainly used as wrapper-based FS techniques.

(Faris et al., 2019) introduced a binary salp swarm algorithm in which the number of leaders and followers vary with time. The Random Weight Network classifier is used as a learning algorithm. This method (TVBSSA) is evaluated using 20 standard UCI datasets, and the results show that TVBSSA combined with RWN outperformed other similar methods. (Al-Tashi et al., 2019) presented a binary form of hybrid GWO and PSO methods for feature selection (BGWOPSO). This method is tested using eighteen UCI datasets. The results show that BGWOPSO obtained outstanding results compared with BPSO, BGWO, BGA and WOA with SA methods. Similarly, (Emary et al., 2016) introduced a binary form of GWO using two techniques. The first technique obtains the new GWO position by stochastic crossover for the first three solutions. In the second, the sigmoid function has been used to convert GWO to binary form. The proposed methods obtained superior results compared with the GA and PSO algorithms.

(Elmanakhly et al., 2021) provided an improved binary EO method for FS. Opposition-based learning is used at the initialization stage to improve the exploration by enhancing the diversity of the population in the defined space. The exploitation stage of the EO algorithm is enhanced by using the local search algorithm. The proposed algorithm is evaluated using KNN and SVM classifiers as induction algorithms. The IBEO method obtained superior accuracy results compared with GOA,

Table 1

Summary of recent FS methods.

Author	Algorithm	W/ H	Contribution	Classifier	Evaluation Measures
(Mafarja et al., 2018)	BDA	W	Eight time-varying transfer functions are integrated into BDA. Applied on 18 UCI datasets and three high-dimensional datasets.	KNN	Accuracy, Fitness, # Features, Specificity, Sensitivity, AUC
(Faris et al., 2019)	TVBSSA	W	Dynamically changes the number of leaders and followers in SSA. Applied on 20 UCI datasets. Better classification accuracies compared with GA, PSO and GWO.	RWN	Accuracy, Fitness, # Features
(Al-Tashi et al., 2019)	BGWOPSO	W	Binary version of hybrid GWO and PSO. Applied on 18 UCI datasets.	KNN	Accuracy, Fitness, # Features, Time
(Emary et al., 2016)	bGWO1 bGWO2	W	Two binary versions of GWO. Applied on 18 UCI datasets. Compared with only two methods-GWO and PSO.	KNN	Accuracy, Fitness, # Features, F-score
(Gao et al., 2020)	BEO-V2	W	Integrated S and V-shaped TFs with EO. Tested on 19 UCI datasets. Inferior performance on low-dimensional datasets.	KNN	Accuracy, Fitness, # Features
(Too and Mirjalili, 2020)	GLEO	W	General learning strategy is combined with EO. Applied on 16 biological datasets	KNN	Accuracy, Fitness, # Features, Time
(Meenachi and Ramakrishnan, 2021)	GATFRO ACTFRO	W	Combined GA with tabu search and ACO with tabu search algorithm. Applied on four medical datasets and one non-medical dataset	Fuzzy Rough Nearest Neighbour classifier	Accuracy, Specificity, Sensitivity, F-measure, ROC, Positive Predictive Value
(Elmanakhly et al., 2021)	IBEO	W	Combined OBL to enhance the diversity Implemented a local search strategy to enhance exploitation Applied on 25 datasets	KNN	Fitness values
(Awadallah et al., 2022a)	BHOA	W	Combined binary version of HHO with crossover operators Tested on 24 datasets	KNN	Accuracy, Specificity, Sensitivity, Fitness, # Features, Time
(Awadallah et al., 2022b)	BRSO BERSO BERSOC	W	Combined binary version of RSO with local best concepts of PSO and crossover operators Tested on 24 datasets. Compared with 25 metaheuristics and five other filter methods.	KNN	Accuracy, Specificity, Sensitivity, Fitness, # Features, Time
(Sayed et al., 2022)	CEO	W	Chaos theory is combined with binary version of EO. Applied on 15 datasets and four NLP datasets.	KNN	Accuracy, Specificity, Sensitivity, Error rate
(Tsai & Hsiao, 2010)	PCA + GA + CART	H	Multi-intersection of PCA, GA and CART. Used for stock market prediction.	BPNN	Accuracy, #Features
(Got et al., 2021)	FW-GPAWOA	H	Multi-Objective optimization algorithm based on WOA. Applied on 12 datasets.	KNN	Accuracy, #Features, Time, HV-metric
(Hammami et al., 2019)	FW-NSGA-II	H	Multi-objective approach combining GA and local search. Applied on 12 datasets and six high dimensional datasets.	KNN	Error rate, #Features
(Ghosh et al., 2019)	Relieff + Chi-square + Symmetrical uncertainty + GA	H	Applied on five cancer datasets.	KNN SVM MLP	Accuracy, #Features
(Unler, et al., 2011)	mr ² PSO	H	Tested on six datasets Combines mutual information and modified discrete PSO	SVM	Accuracy, Kappa statistic, # Features
(Dabba et al., 2021)	MIM-mMFA	H	Combined Mutual Information Maximization with modified MFO algorithm Tested on 16 gene expression datasets	SVM with LOOCV	Accuracy, # Genes, Recall, Precision, F-measure, G-mean
(Varzaneh et al., 2022)	BIMEO	H	Combined Binary EO algorithm with an Entropy-based operator and levy-flight. Applied on 23 datasets.	KNN	Accuracy, Fitness, #Features, Time, Precision, Recall, F-measure
(Zhang et al., 2020)	IG-MBKII	H	Combined Information gain and Improved binary krill herd algorithm. Tested on nine microarray datasets.	KNN, SVM, NB	Accuracy, # Features
(Ouadfel and Elaziz, 2022)	RBEOL-LS	H	Combined ReliefF and local search strategy with EO. Applied on 16 UCI datasets and ten biological datasets	KNN	Accuracy, # Features, Fitness
(Sun et al., 2019)	RFACO-GS	H	Combined ReliefF and ACO. Tested on six biological datasets.	KNN	Accuracy, # Genes

GWO, PSO, WOA, DA and improved SSA algorithms. Similarly, (Too and Mirjalili, 2020) improved the EO algorithm using a General Learning Strategy (GLEO) to avoid the local areas and enhance its capacity to identify promising regions. The GLEO is evaluated using sixteen biological datasets. It provides better results than BOA, GWO, PSO, SCA and RF.

(Awadallah et al., 2022a) proposed three versions of the Rat Swarm optimizer algorithm (RSO). The enhanced RSO is combined with local concepts of PSO and crossover operators. The local concepts of PSO improve the exploitation ability, and the crossover operators improve the diversity of the algorithms. These versions obtained excellent results compared with 25 metaheuristic algorithms and five filter methods. In (Awadallah et al., 2022b), the authors integrated S, U and V-shaped transfer functions into Horse Herd Optimization Algorithm (HOA) to convert the solutions to binary form. Additionally, three crossover operators are combined to improve the algorithm's efficiency. Of all the versions, the BHOA with S-shape and one-point crossover gave competitive results compared with the other methods.

To deal with high-dimensional datasets, the performance of the metaheuristics decreases inspite of their robustness. Therefore, hybrid methods are used in these cases, i.e., a filter approach is hybridized with a wrapper approach in a single context to boost the feature selection method. The filter methods reduce the search space and hence improve the convergence speed.

(Hammami et al., 2019) proposed a multi-objective hybrid method for FS. In this work, they used two filter approaches and one wrapper approach. Initially, the whole population's performance is evaluated using the filter approach. In the next stage, only the best features have been improved using a local search method. The results show that the proposed method gives outstanding results compared with existing methods.

(Got et al., 2021) introduced a hybrid FS approach using the WOA algorithm. It is a multi-objective approach. In this method, the filter and wrapper approaches are optimized simultaneously using WOA. This method has outperformed other methods in terms of accuracy and the number of features selected. (Tsai & Hsiao, 2010) combined three popular feature selection approaches, namely, PCA, GA and Decision Trees, to find more significant features by using the union, intersection, and multi-intersection strategies for stock prediction. BPNN is used as the prediction model.

(Ghosh et al., 2019) presented a bi-stage model for cancerous gene identification. An ensemble of filter techniques is developed by considering the union and intersection of top-n attributes of chi-square, ReliefF and symmetrical uncertainty. This union combines all the information from the three rankings. The authors used GA on the group to get better results. Three classifiers, KNN, SVM and multi-layer perceptron, have been used. This method is tested on five microarray datasets, demonstrating its superiority over other state-of-the-art methods. (Unler et al., 2011) introduced a hybrid filter-wrapper FS method with SVM as the learning algorithm. The filter-based selection method is based on mutual information, and the wrapper-based selection method is based on modified discrete PSO. The information obtained from the filter method is used to weigh the discrete PSO's bit selection probabilities. This method delivered better accuracy outcomes. Table 1 gives an outline of the recent wrapper and hybrid FS works. It shows various FS methods proposed by different authors and their contributions.

3. Theory

3.1. Equilibrium Optimizer (EO)

In the EO method (Faramarzi et al., 2020), a mass balance equation has been used, which is represented by a first-order differential equation described in Eq. (1).

$$V \frac{dZ}{dt} = XZ_{em} - XZ + R \quad (1)$$

Here Z is the concentration in the system, $V \frac{dZ}{dt}$ is the change in mass of the system with time t and volume V, Z_{em} is the concentration of the system in an equilibrium state, X is the volumetric flow rate into and out of the system and R is the mass generation rate in the system. A steady state is reached when $V \frac{dZ}{dt} = 0$. Eq. (1) is rearranged and integrated to obtain the concentration Z as follows:

$$Z = Z_{em} + (Z_0 - Z_{em}) \cdot \vec{F} + \frac{R}{\alpha' V} (1 - \vec{F}) \quad (2)$$

where Z_0 is the initial concentration and F is calculated as follows:

$$F = \exp[-\alpha(t - t_0)] \quad (3)$$

where t_0 denotes the initial time.

3.1.1. Initialization of the particles and Equilibrium pool

The initial concentration of the particles is generated in the defined space using Eq. (4) with random initialization.

$$Z_i = Z_{min} + rand_i(Z_{max} - Z_{min}) \quad i = 1, 2, \dots, n \quad (4)$$

where Z_{max} and Z_{min} are the maximum and minimum values of the dimensions, Z_i is the initial concentration, and $rand_i$ is the random vector in the range [0,1].

This algorithm constructs a vector consisting of five particles at an equilibrium state. The first four particles are the best-so-far particles which improve the exploration ability of the algorithm, and the fifth particle is the arithmetic mean of the first four particles, enhancing the exploitation capability of the algorithm. These equilibrium particles are used to construct an equilibrium pool vector:

$$Z_{em,pool} = \{Z_{em(1)}, Z_{em(2)}, Z_{em(3)}, Z_{em(4)}, Z_{em(avg)}\} \quad (5)$$

$$Z_{em(avg)} = \frac{Z_{em(1)} + Z_{em(2)} + Z_{em(3)} + Z_{em(4)}}{4} \quad (6)$$

3.1.2. Exponential term (F) and generation rate (R)

A fine balance between exploration and exploitation ability of the particles is obtained in EO using the exponential term.

$$F = b_1 sign(\vec{r} - 0.5)[e^{-\vec{\alpha}'t} - 1] \quad (7)$$

$$t = (1 - \frac{Itr}{Itr_{max}})^{(b_2 \times \frac{Itr}{Itr_{max}})} \quad (8)$$

where b_1 and b_2 are constants with values 2 and 1, respectively. r denotes the random vector in the range [0,1]. Itr indicates the current iteration value and, Itr_{max} indicates the maximum no. of iterations.

Generation Rate (R) is used to find a better solution by improving the exploitation capability of the algorithm. The author used the first-order exponential term to define the generation rate. The final set of equations defining the generation rate is as follows:

$$\vec{R} = \vec{R}_0 e^{-\vec{\alpha}'(t-t_0)} = \vec{R}_0 \vec{F} \quad (9)$$

Here R_0 indicates the initial value of generation rate and α' is the decay constant

$$\vec{R}_0 = \vec{C}_p (Z_{em} - \vec{\alpha}' Z) \quad (10)$$

$$\vec{C}_p = \begin{cases} 0.5r_1, & r_2 \geq G_p \\ 0, & r_2 < G_p \end{cases} \quad (11)$$

where r_1 and r_2 indicate the random numbers in the range [0,1] and \vec{C}_p is the Generation rate Control Parameter. A fine balance between

exploration and exploitation ability is achieved with $G_p = 0.5$.

Finally, the equation for updating concentration of the particles by EO is defined as:

$$Z \left(Itr + 1 \right) = Z_{em} + (Z - Z_{em}) \cdot \vec{F} + \frac{R}{\alpha' V} \left(1 - \vec{F} \right) \quad (12)$$

Here V is set to 1. The first term in the above equation is the equilibrium concentration. The second term is responsible for the exploration ability of the algorithm, and the third term enhances the exploitation capability, making the solution more accurate. Algorithm 1 presents the EO algorithm in better detail.

Algorithm 1: EO method

```

Initialize the population of the particles using Eq. (1)
Define the constants b1 = 2, b2 = 1 and Gp = 0.5
for Itr = 1:Itrmax
    for i = 1: n
        Calculate the objective function of the particles
        Acquire the four best-so-far candidates, Zem(1), Zem(2), Zem(3), Zem(4)
    end for
    Compute the average of four candidates, Zem(avg) using Eq. (6)
    Construct the equilibrium pool, Zem,pool using five candidates as in Eq. (5)
    Achieve memory saving (if Itr > 1)
    Assign time, t using Eq. (8)
    for i = 1: n
        Select one candidate from Zem,pool randomly
        Compute exponential term F using Eq. (7)
        Compute R using Eq. (9)
        Update the concentrations, Z using Eq. (12)
    end for
end for
return the best solution

```

3.2. Fuzzy entropy measures using similarity values

Entropy is defined as a “measure of the degree of fuzziness” (De Luca & Termini, 1972). It can be defined as the average information available in the data for decision making, e.g., classifying objects.

The entropy measure proposed by De Luca & Termini has been used in this paper, which is defined as follows:

$$H(B) = - \sum_{i=1}^n (\mu_B(z_i) \log \mu_B(z_i) + (1 - \mu_B(z_i)) \log (1 - \mu_B(z_i))) \quad (13)$$

where $\mu_B(z_i)$ is the membership degree of z_i in the range [0,1] to the fuzzy set B.

These entropy values can be used for classification. They signify whether the data is informative or identified by uncertainty (Luukka, 2007). If the entropy values are small, it indicates regularities and structure in the data, whereas high values indicate randomness (Yao et al., 1999). The fuzzy entropy values can also find out the relevance of features in a dataset.

The procedure for feature removal is described below:

- The data is given as input. An ideal vector is calculated for the training data using generalized means over the samples for specific classes.
- After calculating the ideal vectors of the data, the next step is to calculate the similarity value S of a feature f of sample z_i of the training set with that of the ideal vector v_j of class j. The value of the parameter d is set to 1.

$$S(z_{if}, v_{jf}) = \sqrt[d]{1 - |z_{if}^d - v_{jf}^d|} \quad (14)$$

- The entropy value H for each feature f is calculated using the similarity value S, which is represented by the equation below:

$$H_f = \sum_{j=1}^N \sum_{i=1}^n H(S(z_{if}, v_{jf})) \quad (15)$$

where the similarity values obtained are summed over all samples (i = 1, 2, ..., n) and classes (j = 1, 2, ..., N).

- The features which obtained the highest entropy values are removed.

3.3. ReliefF

The Relief method (Kira & Rendell, 1992) is used to select the attributes depending on how well their values discriminate from instances which are close to each other. The attribute score decreases if there is a difference in feature value in the nearest sample with the same class. Alternatively, the score increases if there is a difference in feature value in the nearest sample with a different class. This method can effectively measure the quality of the features with dependencies and can handle nominal and numerical attributes. The disadvantage of this method is that it is limited to problems with two classes.

This method has been extended to ReliefF (Robnik-Sikonja & Kononenko, 2003), which is a robust method and can deal with noisy data. It can be used for multi-class classification problems. The Relief method’s single nearest hit and miss are insufficient and may cause noisy data. Hence, ReliefF finds k nearest hits and k nearest misses and finds the average to find the weights of the attributes.

3.4. Fuzzy KNN based on Bonferroni mean

3.4.1. Fuzzy KNN

KNN (Fix et al., 1951) is a non-parametric supervised learning method which determines the class of a sample through a majority voting of nearest neighbours. In a Fuzzy KNN (Keller et al., 1985) method, a membership degree is calculated for each class, and the highest membership degree determines the classification decision. Let Z = {z₁, z₂, ..., z_n} be a training set where n is the no. of samples. The membership degree assigned to the test sample y in each class i is expressed as below:

$$\mu_i(y) = \frac{\sum_{j=1}^k u_{ij} (1 / ||y - z_j||^{\frac{2}{m-1}})}{\sum_{j=1}^k (1 / ||y - z_j||^{\frac{2}{m-1}})} \quad (16)$$

where u_{ij} is the membership of the jth observation in the ith class and m ∈ (1, ∞).

3.4.2. Bonferroni mean operator

The Bonferroni mean operator is one of the aggregation operators developed by Bonferroni (Bonferroni, 1950). It can view inter-relationships and allow multiple comparisons between the input arguments. The Bonferroni mean can be defined as follows:

Let Z = {z₁, z₂, ..., z_n}, z_i ∈ [0,1] ∀ i = 1, 2, ..., n be a vector with atleast one z_i ≠ 0 and p, q ≥ 0 be parameters. The Bonferroni mean of z_i is expressed as:

$$B^{p,q}(Z) = \left(\frac{1}{n} \sum_{i=1}^n z_i^p \left(\frac{1}{n-1} \sum_{i,j=1, j \neq i}^n z_j^q \right) \right)^{\frac{1}{p+q}} \quad (17)$$

The Algorithm 2 describes the procedure for BMFK classifier

Table 2

Time Varying STFs and VTFs.

Time Varying S-shaped family		Time Varying V-shaped family	
Name	TF	Name	TF
TV_{S1}	$TF(z, \tau) = \frac{1}{1 + e^{\frac{-2z}{\tau}}}$	TV_{V1}	$TF(z, \tau) = \begin{cases} 1 - \frac{2}{1 + e^{\frac{-2z}{\tau}}} & z \leq 0 \\ \frac{2}{1 + e^{\frac{-2z}{\tau}}} - 1 & z > 0 \end{cases}$
TV_{S2}	$TF(z, \tau) = \frac{1}{1 + e^{\frac{-z}{\tau}}}$	TV_{V2}	$TF(z, \tau) = \begin{cases} 1 - \frac{2}{1 + e^{\frac{-z}{\tau}}} & z \leq 0 \\ \frac{2}{1 + e^{\frac{-z}{\tau}}} - 1 & z > 0 \end{cases}$
TV_{S3}	$TF(z, \tau) = \frac{1}{1 + e^{\frac{-z}{2\tau}}}$	TV_{V3}	$TF(z, \tau) = \begin{cases} 1 - \frac{2}{1 + e^{\frac{-z}{2\tau}}} & z \leq 0 \\ \frac{2}{1 + e^{\frac{-z}{2\tau}}} - 1 & z > 0 \end{cases}$
TV_{S4}	$TF(z, \tau) = \frac{1}{1 + e^{\frac{-z}{3\tau}}}$	TV_{V4}	$TF(z, \tau) = \begin{cases} 1 - \frac{2}{1 + e^{\frac{-z}{3\tau}}} & z \leq 0 \\ \frac{2}{1 + e^{\frac{-z}{3\tau}}} - 1 & z > 0 \end{cases}$

(Kumbure et al., 2020):

Algorithm 2: BMFK

Input: $\{z_j, c_j\}$ (labeled set), y (test or query sample), k (no. of nearest neighbours), p, q ($p, q > 0$)

Output: Class label for query sample (c^*)

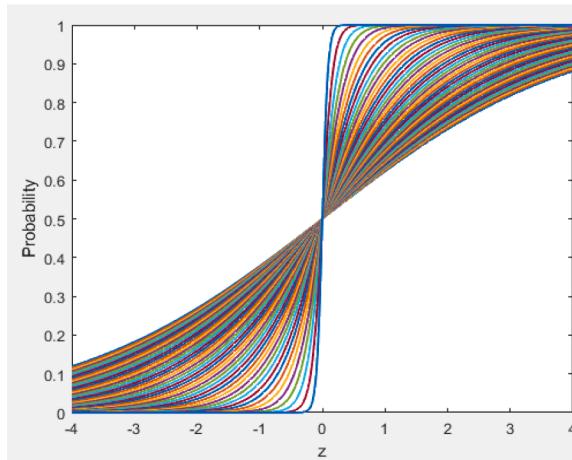
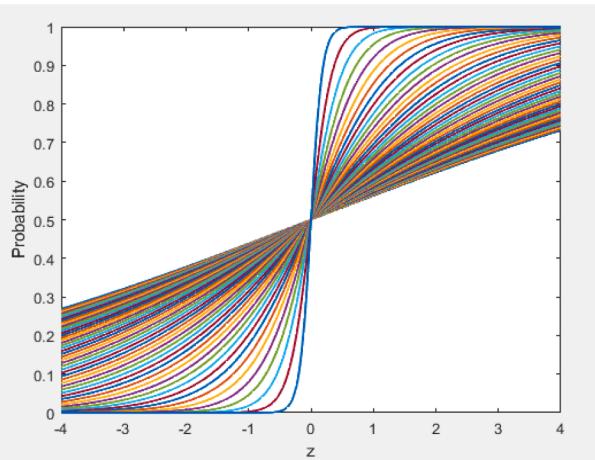
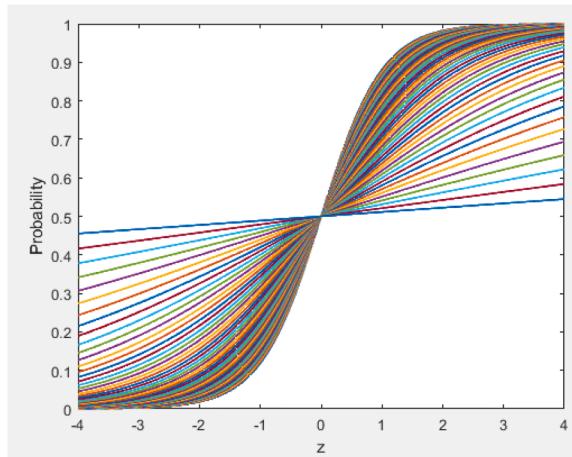
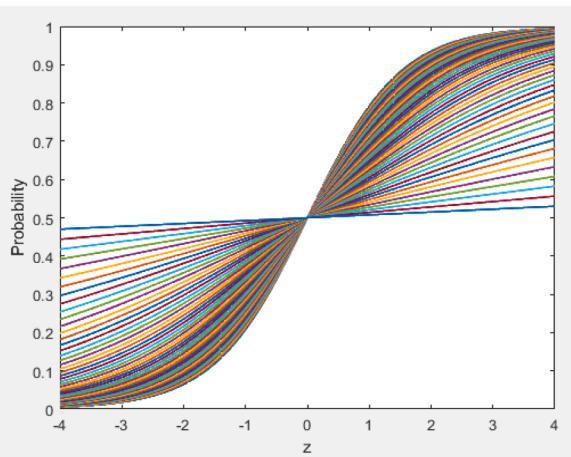
```

for  $j = 1: N$ 
    Calculate Euclidean distances ( $d_{eu}(y, z_j)$ ) from  $y$  to  $z_j$ 
    if  $j < k$  then
        Add  $z_j$  to the nearest neighbour set  $nn^k(y)$ 
    else if  $z_j$  is closer to the query than any of the neighbors in the  $nn^k(y)$ 
        Remove the farthest neighbour and add  $z_j$  in  $nn^k(y)$ 
    end if
end for
for  $r = 1: t$  (no. of included classes in  $nn^k(y)$ )
    Find  $B_r$  in the nearest neighbour set  $nn^k(y)$  using Eq. (17) and set the corresponding class  $c_r$ 
    Calculate Euclidean distances ( $d_{eu}(y, B_r)$ ) from  $y$  to  $B_r$ 
    Assign membership degree  $\mu_r(y)$  to  $c_r$  in terms of weighted distance
end for
return  $c^*$ 

```

3.5. Time-Varying transfer functions

There are mainly-three types of transfer functions (TF): S-shaped Transfer Functions (STF), V-shaped Transfer Functions (VTF) and linear normalized transfer functions. The first transfer function belongs to the S-shaped family. This was proposed by Kennedy and Eberhart (Kennedy

a) TV_{S1} b) TV_{S2} c) TV_{S3} d) TV_{S4} **Fig. 1.** Time-Varying S-shaped family Transfer Functions when $\tau_{mn} = 0.01$ and $\tau_{mx} = 4$ for 100 iterations.

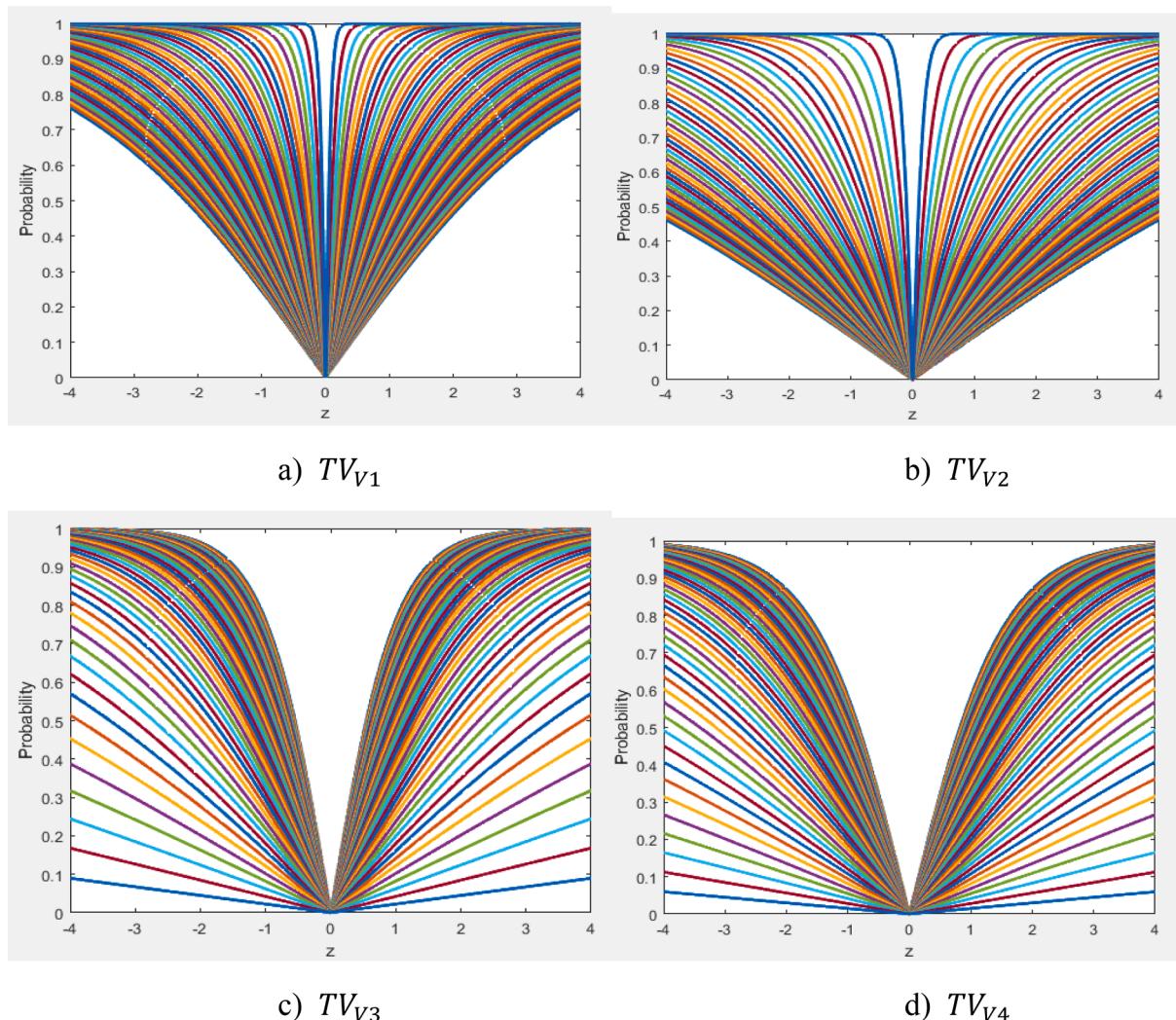


Fig. 2. Time-Varying V-shaped family Transfer Functions when $\tau_{mn} = 0.01$ and $\tau_{mx} = 4$ for 100 iterations.

1	0	0	1	1	0	0	1
---	---	---	---	---	---	---	---

Fig. 3. Feature subset sample.

Table 3
ReliefF and Fuzzy Entropy diversity study.

Dataset	Spearman's rank correlation coefficient
Heart Statlog	0.1483
Breast Tissue	0.1833
Mammographic	0.3000
Appendicitis	-0.1071
E. coli	0.0714
ILPD	0.3454
Pima Indian Diabetes	0.0714
Cleveland	-0.0989
Lymphography	0.0113
Hepatitis	0.3140
New Thyroid	0.2000

& Eberhart, 1997). It is represented as:

$$TF(z_i^d(t)) = \frac{1}{1 + e^{-z_i^d(t)}} \quad (18)$$

Table 4
Results of the robustness of the two filter methods (ReliefF and Fuzzy Entropy) and the ensemble using Spearman correlation coefficient.

Datasets	ReliefF (RF)	Fuzzy Entropy (FE)	Ensemble
Colon	0.51	0.73	0.88
DLBCL	0.51	0.77	0.87
Lung Cancer	0.40	0.48	0.93
SRBCT	0.47	0.57	0.94
TOX_171	0.48	0.52	1.00
Average	0.47	0.62	0.92

where $z_i^d(t)$ is the step vector of the i -th individual. t denotes the iteration and d denotes the dimension of the vector.

Based on the probability value, each position's elements will be changed according to the equation below:

$$z_i^d(t+1) = \begin{cases} 1, & \text{rand} < TF(z_i^d(t+1)) \\ 0, & \text{rand} \geq TF(z_i^d(t+1)) \end{cases} \quad (19)$$

Table 5

Description of the medical datasets used.

Dataset	Instances	Features	Classes	Reference
Heart Statlog (HS)	270	13	2	(Dua & Graff, 2019)
Breast Tissue (BT)	106	10	6	(Dua & Graff, 2019)
Mammographic (Mammo)	830	5	3	(Dua & Graff, 2019)
Coimbra	116	10	2	(Dua & Graff, 2019)
Appendicitis (Appen)	106	7	2	(Alcalá-Fdez et al., 2011)
E.coli	336	7	8	(Dua & Graff, 2019)
Hepatitis C (Hep-C)	615	14	5	(Dua & Graff, 2019)
Cleveland (Cleve)	297	13	5	(Dua & Graff, 2019)
Lymphography (Lymph)	148	18	2	(Dua & Graff, 2019)
Haberman (Haber)	306	3	2	(Dua & Graff, 2019)
Post-Operative (PO)	90	8	3	(Dua & Graff, 2019)
New Thyroid (NT)	215	5	3	(Dua & Graff, 2019)
Spectf Heart (Spectf)	267	44	4	(Dua & Graff, 2019)
ILPD	583	10	2	(Dua & Graff, 2019)
Wisconsin	569	30	3	(Dua & Graff, 2019)
Pima Indian Diabetes (PIMA)	768	8	2	(Dua & Graff, 2019)
Thyroid	7200	21	3	(Dua & Graff, 2019)
Hepatitis	155	19	2	(Dua & Graff, 2019)
Dermatology (Derm)	366	34	6	(Dua & Graff, 2019)
HCC Survival (HCC)	165	49	2	(Dua & Graff, 2019)
Colon	62	2000	2	(Dabba et al., 2021)
TOX_171	171	5748	4	(Dua & Graff, 2019)
Leukemia	72	7070	2	(Dabba et al., 2021)

The V-shaped family transfer function is defined as:

$$TF(z) = \begin{cases} 1 - \frac{2}{1 + e^{-2z}}, & z \leq 0 \\ \frac{2}{1 + e^{-2z}} - 1, & z > 0 \end{cases} \quad (20)$$

The position of the current individual is changed according to the equation below:

$$z_i^d(t+1) = \begin{cases} 0, & r \leq TF(z_i^d(t+1)) \text{ and } z_i^d(t+1) \leq 0 \\ 1, & r \leq TF(z_i^d(t+1)) \text{ and } z_i^d(t+1) > 0 \\ z_i^d(t), & r > TF(z_i^d(t+1)) \end{cases} \quad (21)$$

where r is a random number in the range [0,1].

Similar to the STF, the VTF too suffers the problem of getting trapped in the local optima. In order to overcome this problem, time-varying STFs and VTFs have been proposed by (Mafarja et al., 2018). This helps in achieving a balance between exploration and exploitation. The S-shaped time-varying transfer function is defined as follows:

$$TF(z_i^d(t), \tau) = \frac{1}{1 + e^{\frac{-z_i^d(t)}{\tau}}} \quad (22)$$

$$Z\left(Itr + 1\right) = \begin{cases} Z_{em} + (Z - Z_{em}) \cdot \vec{F} + \frac{R}{\vec{\alpha}V} (1 - \vec{F}), & rand() < 0.5 \\ Z_{em1} + \sin(r1) * |Z - Z_{em2}| \cdot \vec{F} + \frac{R}{\vec{\alpha}V} (1 - \vec{F}), & rand() \geq 0.5 \end{cases} \quad (25)$$

where τ is defined as

$$\tau = \left(1 - \frac{t}{Itr_{max}}\right) \tau_{mx} + \frac{t}{Itr_{max}} \tau_{mn} \quad (23)$$

Here τ is a time- varying parameter; τ_{mx} and τ_{mn} are the maximum and minimum values of τ respectively and Itr_{max} defines the maximum no. of iterations. The V-shaped transfer function in Eq. (19) is modified as:

$$TF(z, \tau) = \begin{cases} 1 - \frac{2}{1 + e^{\frac{-z}{\tau}}}, & z \leq 0 \\ \frac{2}{1 + e^{\frac{-z}{\tau}}} - 1, & z > 0 \end{cases} \quad (24)$$

4. Proposed approach

In this section, we present a novel RFE-BEE algorithm for biomedical dataset classification. The RFE-BEE method mainly consists of two stages: filter and wrapper. The ReliefF and Fuzzy Entropy methods (RFE) are combined to form an ensemble. The RFE method reduces the search space by eliminating feature redundancy. Secondly, the Binary Enhanced Equilibrium Optimization (BEE) is applied to the union of top-n features from each filter method. Finally, the optimal feature subset obtained by the RFE-BEE method is evaluated using the BMFK classifier.

4.1. Binary Enhanced EO approach

Although the EO method works well, sometimes it may get trapped in the local optima. Hence, to improve the exploration capability, Eq. (12) in the algorithm is modified using a novel search strategy defined as

where $r1 = 2*\pi*rand()$.

In the above equation, the particles follow the best particles, and the movement of the particles is modified using the sine algorithm. This improves the search ability of the algorithm.

4.1.1. Cauchy Mutation

Cauchy mutation (Sapre & Mini, 2019) is applied to the EO algorithm to enhance its ability to evade the local optima during the search. The

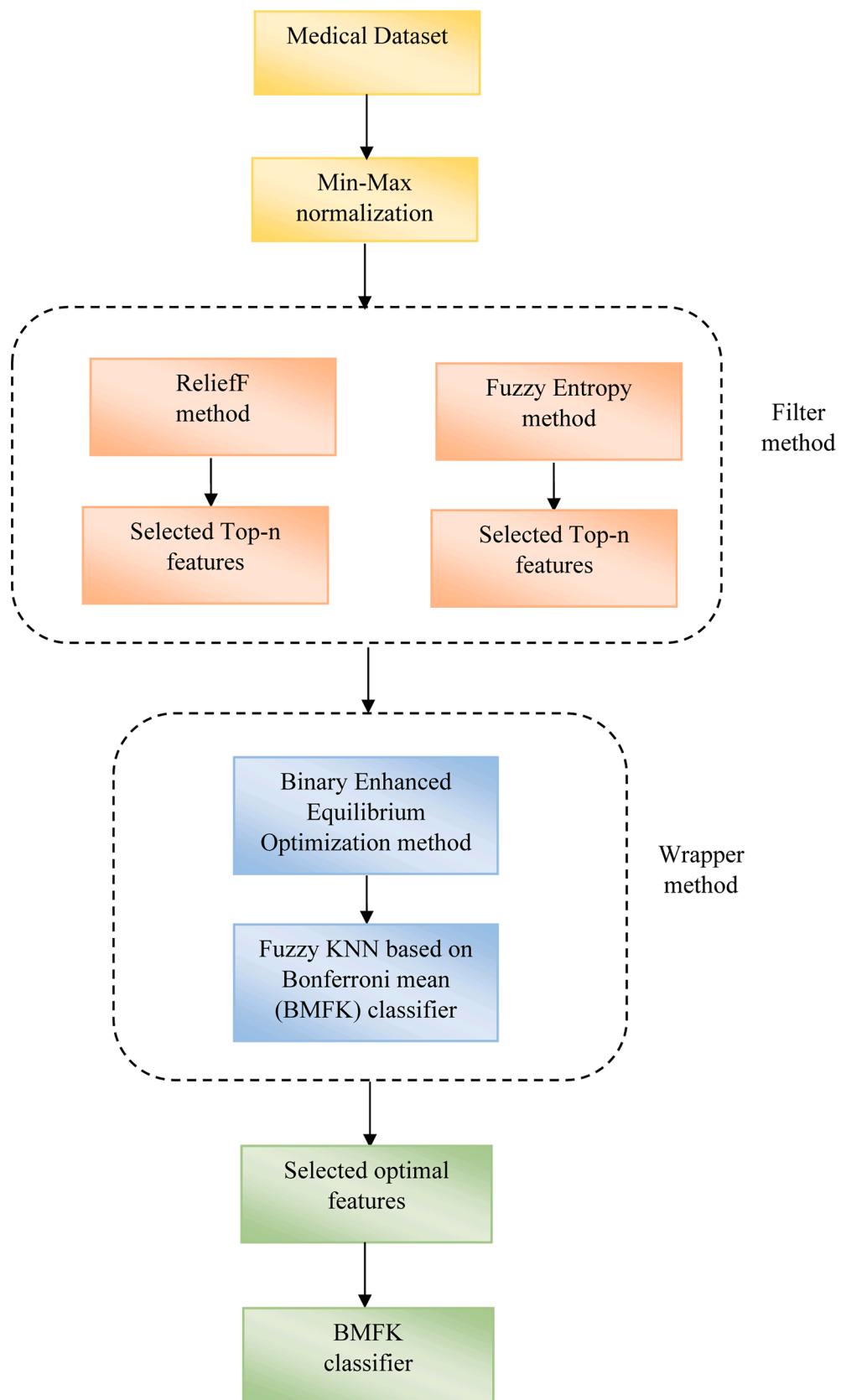


Fig. 4. The proposed RFE-BEE feature selection approach.

Table 6
Parameter values of different algorithms.

Algorithms	Parameter Values
Common parameter settings	N = 10, Maximum number of iterations = 30, value of k in KNN and BMFK = 5, Fuzzy strength parameter in BMFK (m) = 2, α in the fitness function = 0.99.
GA	Mutation Rate = 0.01, Crossover Rate = 0.8.
PSO	Inertia weight, w = 0.9, acceleration constants (c_1 and c_2 = 2).
EO, CEO, BEO, IBEO	Constants (c_1 = 2, c_2 = 1), P = 0.5, Volume v = 1.
GWO, BGWO1, BGWO2, BGWOPSO	Convergence parameter, a = [2,0] (decreases linearly from 2 to 0)
TVBSSA	Constants c_1 = 2
RFE-BEE	Constants (c_1 = 2, c_2 = 1), P = 0.5, Volume v = 1

Cauchy distribution function (y_{cau}) is defined by the equation below:

$$y_{cau} = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{\beta}{g}\right) \quad (26)$$

y_{cau} is the uniformly distributed number defined in the range [0,1], g is taken as unity and $\beta = \tan(\pi(y_{cau} - 0.5))$.

The density function (f_{cau}) is defined as follows:

$$f_{cau}(0, g) = \frac{1}{\pi} \frac{g}{g^2 + \beta^2} \quad (27)$$

After applying the Cauchy random mutation to the i^{th} candidate solution Z_i , the following equation is obtained

$$Z'_i = Z_i(1 + C(\beta)) \quad (28)$$

4.1.2. Opposition based learning (OBL)

The concept of OBL was developed by Tizhoosh (Tizhoosh, 2005). This concept helps the particles of the solution to attain the global optima and improve the convergence rate of the EO algorithm. The present and the opposite populations are generated simultaneously in the search space. Let Z be a real number in the interval [u, l]. The opposite number \bar{Z} can be defined as follows:

$$\bar{Z} = u + l - Z \quad (29)$$

If $l = 0$ and $u = 1$, then

$$\bar{Z} = 1 - Z \quad (30)$$

Table 7
Comparison of the results of EEO approach with EO approach.

Datasets	Best fitness value		Mean fitness value		Mean Classification Accuracy		Number of features selected	
	EEO	EO	EEO	EO	EEO	EO	EEO	EO
Heart Statlog	0.1131	0.0947	0.1427	0.1527	85.73	84.73	5.6	4.25
Breast Tissue	0.0976	0.1908	0.2493	0.2627	75.07	73.73	3.7	3.1
Mammo	0.1292	0.1451	0.1709	0.1736	82.91	82.64	1.8	1.85
Coimbra	0.0464	0.0475	0.1401	0.1497	85.99	85.03	4.1	3.7
Appendicitis	0.0029	0.003	0.0578	0.0664	94.22	93.36	2.5	1.9
E. coli	0.1106	0.0958	0.1557	0.1536	84.43	84.64	5.05	5.15
Hepatitis-C	0.0448	0.0358	0.057	0.0596	94.3	94.04	6.5	5.35
Cleveland	0.0706	0.1013	0.1328	0.1386	86.72	86.14	5.3	4.25
Lymph	0.005	0.0044	0.057	0.0498	94.3	95.02	10.5	6.65
Haberman	0.169	0.1981	0.2283	0.24	77.17	76	1.3	1.15
PO	0.1138	0.1675	0.2225	0.2451	77.75	75.49	4.2	2.5
New Thyroid	0.004	0.0045	0.0302	0.0234	96.98	97.66	2.45	2.5
Spectfheart	0.0581	0.0205	0.1037	0.0697	89.63	93.03	24.9	10.8
ILPD	0.2023	0.2154	0.242	0.2519	75.8	74.81	3.85	2.7
Wisconsin	0.0222	0.0185	0.0457	0.0442	95.43	95.58	10.85	3.7
Pima	0.2237	0.1732	0.2421	0.234	75.79	76.6	4.25	3.95
Thyroid	0.012	0.01	0.0157	0.0158	98.43	98.42	4.2	4.6
Hepatitis	0.0026	0.0369	0.0688	0.0726	93.12	92.74	6.45	3.6
Derm	0.0024	0.0018	0.0075	0.0102	99.25	98.98	16.4	11.5
HCC	0.0624	0.0916	0.1728	0.1631	82.72	83.69	15.95	5.2
Colon	0	0.0001	0.0199	0.0209	98.01	97.91	60.19	53.15
TOX_171	0.0006	0.0008	0.0144	0.0129	98.56	98.71	801.6	723.9
Leukemia	0.002	0.0012	0.018	0.0146	98.2	98.54	290.6	302.9

4.1.3. Time-Varying STFs and VTFs

Eight time-varying STFs and VTFs ($TF(z, \tau)$) are integrated with the Enhanced EO algorithm to convert the solutions into binary form. Table 2 presents the time-varying STFs and VTFs. Fig. 1 and Fig. 2 show the behaviour of the time-varying STFs and VTFs, respectively.

4.1.4. Binary Enhanced EO (BEE) approach for attribute selection

A set of random initial solutions are created by using the BEE approach. In the solution, each vector represents the indices of the corresponding features. BEE has been used for attribute selection to reduce the high dimensionality in biomedical datasets. Either bit value 0 or 1 is used, where the bit value 0 represents the non-selected points and the bit value 1 represents the selected attributes in the dataset.

For example, if the solution Q = {1,0,0,1,1,0,0,1}, it means that four attributes (1st, 4th, 5th, and 8th) are selected. Fig. 3 shows a feature subset sample.

The fitness function that reduces the number of attributes selected and enhances the classifier performance is used and defined as follows:

$$\downarrow \text{Fitness} = \alpha C_E + (1 - \alpha) \frac{|S|}{|L|} \quad (31)$$

where C_E is the classification error computed by the BMFK algorithm, $|S|$ is the length of the attributes selected, $|L|$ is the number of attributes and α is the parameter which controls the weight between classification error and selected attributes.

4.2. The hybrid RFE-BEE feature selection approach

The procedure of the RFE-BEE approach is discussed below:

- Normalization
- Filter methods
- Wrapper method
- Subset Evaluation

4.2.1. Normalization

Normalization is an essential step in the earliest stage of data analysis. There are many reasons for normalization. Some of them are differences in labelling, systematic biases in the measured expression levels

Table 8

Accuracies obtained by integrating four time-varying S-shaped functions with EEO approach.

Dataset	TV_{S1}	TV_{S2}	TV_{S3}	TV_{S4}
HS	86.33	85.48	87.49	86.08
BT	75.12	74.69	74.66	75.15
Mammo	83.86	84	82.93	82.93
Coimbra	86.91	87.75	86.42	87.02
Appen	94.46	92.4	91.21	92.15
E.coli	84.26	85.06	85.02	84.6
Hep-C	93.22	94.46	94.42	94.05
Cleve	88.37	87.02	87.09	88.02
Lymph	95.29	95.79	95.79	96.12
Haber	75.61	75.77	75.22	75.4
PO	74.66	75	77.6	78.06
NT	97.29	97.87	98.11	97.43
Spectf	91.1	91.82	90.6	91.47
ILPD	76.95	76.27	76.16	75.47
Wisconsin	96.42	96.94	96.13	96.2
Pima	75.25	75.58	75.43	75.88
Thyroid	97.4	98.43	98.4	97.6
Hepatitis	94.27	94.48	96.18	94.99
Derm	99.17	99.27	99.15	99.39
HCC	83.56	84.79	84.14	84.44
Colon	97.62	98.51	96.02	96.56
TOX_171	98.26	98.5	97.26	98.01
Leukemia	98.62	98.5	97.5	96.3
Friedman mean rank	2.74	1.85	2.96	2.46
Rank	3	1	4	2

Table 9

Accuracies obtained by integrating four time-varying V-shaped functions with EEO approach.

Dataset	TV_{V1}	TV_{V2}	TV_{V3}	TV_{V4}
HS	82.42	85.09	84.6	82.04
BT	74.09	73.2	73.91	72.94
Mammo	83.37	82.98	83.21	82.21
Coimbra	84.72	88.79	85.98	87.49
Appen	91.16	92.84	92.34	92.61
E.coli	83.63	84.16	85.39	84.23
Hep-C	94.14	93.99	94.3	94.16
Cleve	86.35	87.19	85.1	86.06
Lymph	94.26	91.99	93.04	93.54
Haber	76.29	76.31	76.91	76.16
PO	77.89	76.22	77.81	75.13
NT	97.03	97.88	97.31	97.43
Spectf	88.39	89.21	88.52	88.87
ILPD	75.54	76.14	75.71	75.32
Wisconsin	95.87	95.77	96.15	96.14
Pima	75.89	74.81	75.2	75.63
Thyroid	97.59	98.31	98.6	97.52
Hepatitis	93	93.1	93.69	92.63
Derm	98.82	98.92	99.08	98.95
HCC	85.48	84.83	85.09	84.06
Colon	98.56	97.63	98.6	98.5
TOX_171	97.01	97.62	98.01	97.96
Leukemia	98.71	97.65	97.62	98.06
Friedman Mean rank	2.61	2.48	2.04	2.87
Rank	3	2	1	4

and differences in detection efficiency between the fluorescent dyes used (Dabba et al., 2021). Normalization is the process of adjusting the data values measured on different scales to a common scale by removing the units of measurement or reducing the training error. This makes the data comparison easy. Some standard normalization techniques include transforming data using a z-score or t-score, clipping, log scaling and feature scaling (scaling to a range). Feature scaling is used to re-scale all the data values in the range [a, b]. In this work, we use the Min-Max method to normalize all the data values between [-1,1] using the equation below:

Table 10

Accuracies obtained with and without hybridization of filter and wrapper-based approaches.

Dataset	RF + FE + BEEO- BMFK (RFE- BEE)	RFE- BEE- KNN	BEEO- BMFK	BMFK	RF- BMFK	RF + FE- BMFK
HS	86.87	88.32	85.48	74.7	76.32	79.2
BT	76.77	69.83	74.69	69.49	67.93	70.29
Mammo	83.3	84.53	84	80.9	81.44	83.23
Coimbra	86.88	89.2	87.75	75.45	76.02	81.84
Appen	93.33	92.69	92.4	87.54	89.41	89.24
E. coli	84.7	86.44	85.06	79.89	80.6	84.36
Hep-C	94.55	93.63	94.46	92.61	91.72	92.77
Cleve	88.27	88.34	87.02	75.17	77.49	80.01
Lymph	95.48	93.41	95.79	85.51	84.61	89.22
Haber	76.91	76.83	75.77	75.64	71.88	76.75
PO	76.21	75.66	75	70.8	72.66	72.44
NT	97.07	97.01	97.87	96.04	96.61	97.03
Spectf	92.41	90.9	91.82	81.51	81.86	85.66
ILPD	75.8	76.76	76.27	70.84	71.65	73.17
Wisconsin	97.12	96.38	96.94	93.45	93.43	94.27
Pima	75.67	77.64	75.58	72.51	71	74.88
Thyroid	98.6	98.63	98.43	96.69	96.69	97.03
Hepatitis	95.35	93.8	94.48	84.21	86.97	87.56
Derm	99.37	99.45	99.27	94.94	95.15	97.23
HCC	86.85	83.76	84.79	69.96	69.99	74.96
Colon	98.6	95.72	98.51	87.96	88.37	88.55
TOX_171	98.72	87.42	98.5	87.13	87.13	88.1
Leukemia	98.4	96.35	98.5	82.89	86.07	84.3
Average	89.45	88.38	89.06	81.99	82.39	84.44
Friedman mean rank	1.75	2.05	2.35	5.65	5.25	3.95
Rank	1	2	3	6	5	4

Table 11

Comparison of best fitness values using the RFE-BEE approach and other heuristic algorithms.

Dataset	RFE-BEE- BMFK	GA- BMFK	PSO- BMFK	GWO- BMFK
HS	0.0596	0.0764	0.0947	0.0779
BT	0.1459	0.1437	0.1908	0.1908
Mammo	0.1054	0.1173	0.1332	0.1093
Coimbra	0.0464	0.0486	0.0453	0.0056
Appen	0.0029	0.0043	0.0029	0.0043
E. coli	0.0987	0.0958	0.1091	0.112
Hep-C	0.0386	0.0376	0.0448	0.0367
Cleve	0.0526	0.0706	0.0541	0.1178
Lymph	0.0017	0.0033	0.0028	0.0028
Haber	0.1852	0.1819	0.169	0.1819
PO	0.1163	0.115	0.1687	0.1687
NT	0.004	0.004	0.004	0.004
Spectf	0.0421	0.0223	0.0237	0.0216
ILPD	0.2164	0.2239	0.2154	0.2098
Wisconsin	0.0134	0.0098	0.002	0.0013
Pima	0.2121	0.2121	0.2004	0.1979
Thyroid	0.012	0.012	0.0139	0.0106
Hepatitis	0.0026	0.0369	0.0385	0.0369
Derm	0.0024	0.0021	0.0021	0.0024
HCC	0.0618	0.091	0.1222	0.0616
Colon	0	0.0035	0.0041	0.0062
TOX_171	0.0003	0.0049	0.0047	0.0013
Leukemia	0.0001	0.0043	0.0046	0.0014
Friedman mean rank	2.19	2.63	2.87	2.3
Rank	1	3	4	2

$$Z_{new} = (b - a) \frac{Z - Z_{min}}{Z_{max} - Z_{min}} + a \quad (32)$$

where Z_{new} is the normalized value, Z is the original value, Z_{max} is the

Table 12

Comparison of mean fitness values using the RFE-BEE approach and other heuristic algorithms.

Dataset	RFE-BEE-BMFK	GA- BMFK	PSO- BMFK	GWO- BMFK
HS	0.1313	0.1469	0.1565	0.163
BT	0.2323	0.2628	0.291	0.2557
Mammo	0.167	0.1661	0.1647	0.1769
Coimbra	0.1312	0.1274	0.1483	0.1714
Appen	0.0667	0.0811	0.0859	0.0834
E. coli	0.153	0.1453	0.1531	0.1613
Hep-C	0.0545	0.0572	0.0618	0.0564
Cleve	0.1173	0.1567	0.1549	0.15
Lymph	0.0452	0.0772	0.0693	0.0565
Haber	0.2309	0.2422	0.2424	0.2281
PO	0.2379	0.2581	0.2374	0.2561
NT	0.0293	0.02	0.0248	0.0293
Spectf	0.0759	0.0732	0.0723	0.07
ILPD	0.242	0.2538	0.2545	0.2506
Wisconsin	0.0288	0.043	0.0503	0.0374
Pima	0.2433	0.2467	0.2538	0.2447
Thyroid	0.0154	0.0159	0.0195	0.0169
Hepatitis	0.0465	0.084	0.0824	0.0696
Derm	0.0063	0.0108	0.0111	0.0128
HCC	0.1315	0.1946	0.2083	0.1647
Colon	0.0042	0.0448	0.0746	0.0299
TOX_171	0.0096	0.0713	0.0513	0.0367
Leukemia	0.0075	0.0751	0.0719	0.0335
Friedman mean	1.5	2.78	3.17	2.54
rank				
Rank	1	3	4	2

Table 13

Comparison of accuracies using the RFE-BEE method and other heuristic algorithms.

Dataset	RFE-BEE-BMFK	GA- BMFK	PSO- BMFK	GWO- BMFK
HS	86.87	85.31	84.35	83.7
BT	76.77	73.72	70.9	74.43
Mammo	83.3	83.39	83.53	82.31
Coimbra	86.88	87.28	85.17	82.86
Appen	93.33	91.89	91.41	91.66
E. coli	84.7	85.47	84.69	83.87
Hep-C	94.55	94.28	93.82	94.36
Cleve	88.27	84.33	84.51	85
Lymph	95.48	92.28	93.07	94.35
Haber	76.91	75.78	75.76	77.19
PO	76.21	74.19	76.26	74.39
NT	97.07	98	97.52	97.07
Spectf	92.41	92.68	92.77	93
ILPD	75.8	74.62	74.55	74.94
Wisconsin	97.12	95.7	94.97	96.26
Pima	75.67	75.33	74.62	75.53
Thyroid	98.46	98.41	98.05	98.31
Hepatitis	95.35	91.6	91.76	93.04
Derm	99.37	98.92	98.89	98.72
HCC	86.85	80.54	79.17	83.53
Colon	99.58	95.52	92.54	97.01
TOX_171	99.04	92.87	94.87	96.33
Leukemia	99.25	92.49	92.81	96.65
Friedman mean	1.5	2.78	3.17	2.54
rank				
Rank	1	3	4	2

maximum original value and, Z_{min} is the minimum original value. This method can provide a stable convergence of weights and biases (Jain et al., 2015).

4.2.2. Filter methods

Different filter methods use different evaluation measures to assign ranks to a feature. Each method has its own merits and demerits. If there is a misclassification due to some members of the ensemble, there is a

Table 14

Comparison of number of features selected using the RFE-BEE approach and other heuristic algorithms.

Dataset	RFE-BEE-BMFK	GA- BMFK	PSO- BMFK	GWO- BMFK
HS	5	5.05	4.45	4.5
BT	3.25	3.15	3.1	3.15
Mammo	1.8	2.25	1.7	1.7
Coimbra	3.8	4.25	3.7	3.15
Appen	2.15	2.3	2.35	2.25
E. coli	5.1	5	4.8	5.35
Hep-C	5.55	5.4	5.5	4.95
Cleve	5.6	5.25	5.05	4.1
Lymph	7.55	6.8	8.05	6.5
Haber	1.05	1.1	1.4	1.25
PO	3.7	1.9	2.9	2.45
NT	2.99	2.5	2.6	2.55
Spectf	16.45	17.8	18.35	11.95
ILPD	2.95	3.75	3.6	2.65
Wisconsin	6.35	4.05	6.25	4.35
Pima	3.25	4.25	4.3	3.45
Thyroid	4.65	5.9	6.85	4.7
Hepatitis	4.3	5.15	5.45	4.6
Derm	12.1	11.2	12.45	11.25
HCC	12.15	5.5	13.8	5.75
Colon	32.45	713.05	885.35	195.5
TOX_171	471.5	2365.8	2723.8	1004.8
Leukemia	299.5	3104.8	3346.7	1173.3
Friedman mean	2.48	2.54	3.11	1.87
rank				
Rank	2	3	4	1
Average	39.7	273.31	307.5	107.14

chance that the other members correctly classify an example. This improves the final accuracy. Hence, instead of using a single filter method for ranking a feature, we design an ensemble of two methods of different categories - distance and information based to combine the merits of individual algorithms. As we use filter methods of different categories, there is less chance of losing some important information. This improves the performance of the combination (Bolón-Canedo et al., 2015).

An ensemble of filter techniques is designed regarding the union of top-n attributes of ReliefF and Fuzzy Entropy methods. This ensemble acts as a pre-selector that selects the top-2n features from the dataset, reducing the search space. An ensemble is generally used based on the assumption that a combination of FS methods produces better results than an individual method. To evaluate a learning system, accuracy is usually used as a measure in classification systems. However, in the case of ensembles, diversity and stability are other factors that have relevance in this process. On the one hand, the members of the ensemble must produce diverse results. On the other hand, the ensemble needs to be robust. Therefore, we conduct a diversity and robustness study on the ensemble (Bolón-Canedo & Alonso-Betanzos, 2019). The description of the datasets used is given in Table 5 and Table 22.

A diversity study is conducted to ensure significant diversity in the acquired ensemble. A diversity study is conducted between ReliefF and FE methods using 11 datasets. The final ranks obtained from the two methods are compared using Spearman's rank correlation coefficient in Table 3. The coefficient value of 1 corresponds to a perfect positive correlation, and -1 corresponds to a perfect negative correlation between ranks. The coefficient value of 0 signifies no correlation. The values in Table 3 are far from one, which signifies enough diversity.

Even when there is a small change in the data, the FS methods must return similar results. This property is known as stability. Five high-dimensional small sample size datasets have been used to analyse the stability of FS techniques as they pose a significant challenge in feature selection. To assess the effect of instance perturbation on the FS outputs, we used a 10-fold cross-validation procedure and compared the feature rankings. The Spearman rank correlation is used to compare the feature rankings (Saeys et al., 2008). The results of the robustness of the two

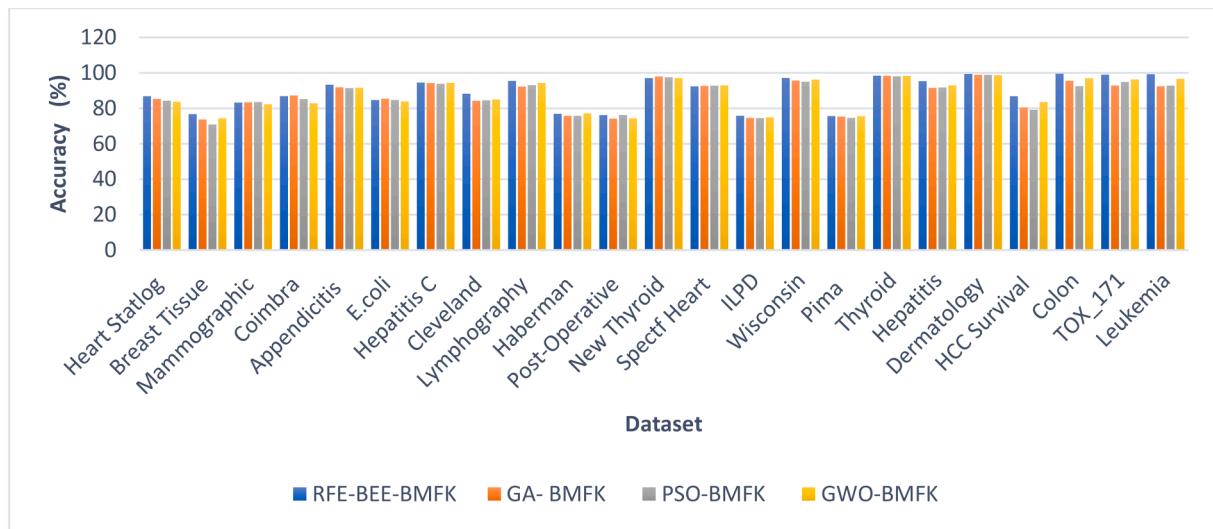


Fig. 5. Comparison of accuracies obtained using the RFE-BEE approach and other heuristic algorithms.

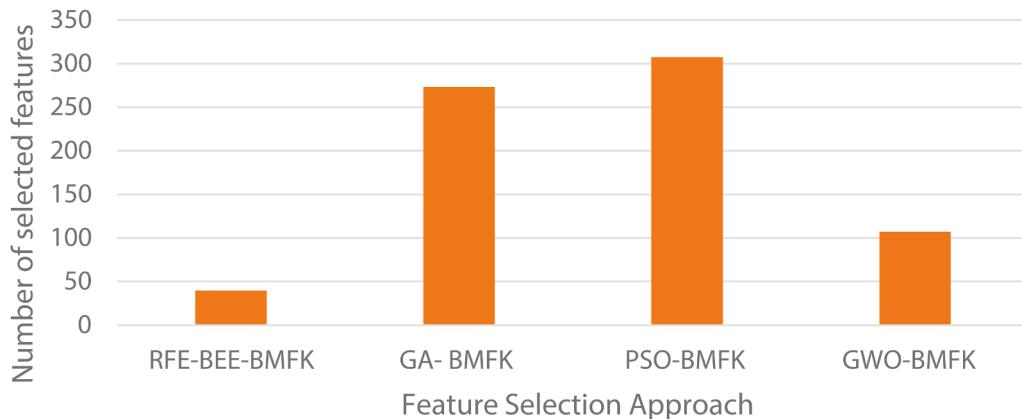


Fig. 6. Average number of features selected using the RFE-BEE approach and other heuristic algorithms.

filter methods (ReliefF and Fuzzy Entropy) and the ensemble are given in Table 4. From the results in Table 4, it can be observed that ensemble FS provides more stable results compared to the individual FS algorithms – ReliefF and Fuzzy Entropy. The robustness of the FS algorithm depends on the dataset.

4.2.3. Wrapper method

The unique features from the top $2n$ features (best features) are selected and given as input to the Binary Enhanced EO method. This method acts as a wrapper method. It iteratively searches the entire feature space and generates an optimal feature subset. Algorithm 3 describes the BEE optimization method in detail.

Algorithm 3: BEE optimization method

```

Initialize the population of the particles
Initialize the opposite population of the particles using Eq. (29)
Calculate the objective function of the initial and opposite population
Acquire the best particles from initial and opposite population
Define the constants  $b_1 = 2$ ,  $b_2 = 1$  and  $G_P = 0.5$ 
Set  $\tau_{mn}$  and  $\tau_{mx}$  values
Update  $\tau$  using Eq. (23)
for Itr = 1: Itrmax
    for i = 1: n
        Calculate the objective function of the particles
        Acquire the four best-so-far candidates,  $Z_{em(1)}, Z_{em(2)}, Z_{em(3)}, Z_{em(4)}$ 
    end for

```

(continued)

Algorithm 3: BEE optimization method

```

        Compute the average of four candidates,  $Z_{em(avg)}$  using Eq. (6)
        Construct the equilibrium pool,  $Z_{em,pool}$  using five candidates as in Eq. (5)
        Achieve memory saving (if Itr > 1)
        Assign time, t using Eq. (8)
        for i = 1: n
            Choose one candidate from  $Z_{em,pool}$  randomly
            Compute exponential term F using Eq. (7)
            Compute R using Eq. (9)
            Update the concentrations, Z using Eq. (25)
            Apply the Cauchy random mutation using Eq. (28)
            Calculate the objective function of particles obtained using Eq. (31)
            Update the best particles by selecting n best concentrations
        end for
        Calculate  $TF(Z)$  using equations from Table 2
        Update  $Z_{t+1}$  using Eq. (19) or Eq. (21)
    end for
    return the best solution

```

4.2.4. Subset evaluation

After the maximum number of iterations is achieved, the selected features from the BEE method are given as input to the BMFK classifier for evaluation. The BMFK classifier has been used because it is concerned with class distributions and mitigates the problem of class imbalance. Fig. 4 describes the proposed RFE-BEE feature selection approach.

(continued on next column)

Table 15

Comparison of best fitness values using the RFE-BEE approach and other existing methods.

Dataset	RFE-BEE-BMFK	BGWOPSO-KNN	TVBSSA-KNN	BGWO1-KNN	BGWO2-KNN
HS	0.0596	0.0947	0.0596	0.0955	0.0573
BT	0.1459	0.1908	0.1919	0.2402	0.1919
Mammo	0.1054	0.1133	0.1113	0.1312	0.1173
Coimbra	0.0464	0.0475	0.0486	0.0056	0.0453
Appen	0.0029	0.0029	0.0014	0.0014	0.0486
E. coli	0.0987	0.0796	0.1077	0.0662	0.0677
Hep-C	0.0386	0.0376	0.0367	0.0448	0.0528
Cleve	0.0526	0.0384	0.0691	0.1193	0.0533
Lymph	0.0017	0.0358	0.0375	0.0408	0.0364
Haber	0.1852	0.169	0.1981	0.1656	0.1852
PO	0.1163	0.115	0.1687	0.1725	0.1687
NT	0.004	0.004	0.004	0.006	0.004
Spectf	0.0421	0.003	0.0633	0.0617	0.0207
ILPD	0.2164	0.1672	0.2098	0.1938	0.2098
Wisconsin	0.0134	0.0192	0.0124	0.0134	0.0273
Pima	0.2121	0.2004	0.1939	0.2068	0.1862
Thyroid	0.012	0.0122	0.0148	0.0176	0.0095
Hepatitis	0.0021	0.0369	0.0037	0.0058	0.0026
Derm	0.0021	0.0029	0.0035	0.0047	0.0024
HCC	0.0618	0.0618	0.0957	0.1547	0.0606
Colon	0	0.0001	0.0048	0.0049	0.0008
TOX_171	0.0003	0.0319	0.0924	0.0934	0.0015
Leukemia	0.0001	0.0009	0.0049	0.0049	0.0011
Friedman	2.35	2.59	3.41	3.89	2.76
mean rank					
Rank	1	2	4	5	3

5. Experimental study

Extensive experiments are performed to prove the ability of the RFE-BEE algorithm using 27 datasets and compared with various algorithms. Eight variants of Enhanced Equilibrium Optimizer are developed and tested to show the efficiency of the proposed BEE algorithm. The description of the medical datasets used is given in Section 5.1. Section 5.2 discusses the four phases of the experimental study and the metrics used for evaluation in each phase. The parameter settings used by different algorithms are also given in the same section. The experimental results and the performance of the RFE-BEE against other algorithms are given in Sections 5.3 to 5.6.

5.1. Description of the medical datasets

The details of the biomedical datasets used for the experimental study are given in Table 5. Twenty-three datasets of various dimensions from the UCI and Keel repositories have been employed to test the efficiency of the proposed RFE-BEE approach.

5.2. Metrics and parameters

We performed four phases of experimental study and used the following performance measures to evaluate the efficiency of the RFE-BEE approach. In the first phase, we compare the proposed approach with and without hybridization in terms of accuracy. In the second phase, the RFE-BEE approach is compared with three other *meta-heuristic* approaches – GA, PSO and GWO using Fuzzy KNN based on Bonferroni mean (BMFK) as the learning model. In the third phase of experiments, this method is compared with the popular existing methods present in the literature using KNN as the learning model. To test the performance of these experiments, we used the best fitness value, mean fitness value, classification accuracy and the average number of features selected by different approaches as performance measures. In the fourth phase of the experimental study, we used four microarray datasets of high dimensions and a small number of instances where we used Accuracy, Precision, Sensitivity, F-measure, and the

Table 16

Comparison of mean fitness values using the RFE-BEE approach and other existing methods.

Dataset	RFE-BEE-BMFk	BGWOPSO-KNN	TVBSSA-KNN	BGWO1-KNN	BGWO2-KNN
HS	0.1313	0.1504	0.129	0.1723	0.1368
BT	0.2323	0.3021	0.3045	0.3365	0.2993
Mammo	0.167	0.1632	0.1565	0.1649	0.1608
Coimbra	0.1312	0.1355	0.1273	0.1458	0.1372
Appen	0.0667	0.0661	0.0639	0.1042	0.0895
E.coli	0.153	0.1373	0.1382	0.1384	0.1324
Hep-C	0.0545	0.0666	0.0653	0.0639	0.0696
Cleve	0.1173	0.1461	0.1299	0.1606	0.1332
Lymph	0.0452	0.0883	0.0936	0.1114	0.0847
Haber	0.2309	0.2319	0.2387	0.2538	0.2532
PO	0.2379	0.2719	0.2404	0.2815	0.2832
NT	0.0293	0.039	0.035	0.0401	0.0403
Spectf	0.0759	0.0578	0.1116	0.1024	0.0623
ILPD	0.242	0.2421	0.2377	0.2535	0.2472
Wisconsin	0.0288	0.0461	0.0414	0.0531	0.0513
Pima	0.2433	0.231	0.2252	0.2436	0.2255
Thyroid	0.0154	0.0154	0.0201	0.0278	0.015
Hepatitis	0.0465	0.0864	0.0713	0.0909	0.0774
Derm	0.0063	0.0127	0.0078	0.0128	0.0077
HCC	0.1315	0.1562	0.1778	0.236	0.1568
Colon	0.0042	0.0222	0.1046	0.1043	0.0216
TOX_171	0.0096	0.1093	0.181	0.1739	0.0644
Leukemia	0.0075	0.0226	0.0865	0.0688	0.0225
Friedman	1.85	2.93	2.78	4.52	2.91
mean rank					
Rank	1	4	2	5	3

Table 17

Comparison of accuracies using the RFE-BEE approach and other existing methods.

Dataset	RFE-BEE-BMFk	BGWOPSO-KNN	TVBSSA-KNN	BGWO1-KNN	BGWO2-KNN
HS	86.87	84.96	87.1	82.77	86.32
BT	76.77	69.79	69.55	66.35	70.07
Mammo	83.3	83.68	84.35	83.51	83.92
Coimbra	86.88	86.45	87.27	85.42	86.28
Appen	93.33	93.39	93.61	89.58	91.05
E.coli	84.7	86.27	86.18	86.16	86.76
Hep-C	94.55	93.34	93.47	93.61	93.04
Cleve	88.27	85.39	87.01	83.94	86.68
Lymph	95.48	91.17	90.64	88.86	91.53
Haber	76.91	76.81	76.13	74.62	74.68
PO	76.21	72.81	75.96	71.85	71.68
NT	97.07	96.1	96.5	95.99	95.97
Spectf	92.41	94.22	88.84	89.76	93.77
ILPD	75.8	75.79	76.23	74.65	75.28
Wisconsin	97.12	95.39	95.86	94.69	94.87
Pima	75.67	76.9	77.48	75.64	77.45
Thyroid	98.46	98.46	97.99	97.22	98.5
Hepatitis	95.35	91.36	92.87	90.91	92.26
Derm	99.37	98.73	99.22	98.72	99.23
HCC	86.85	84.38	82.22	76.4	84.32
Colon	99.58	97.78	89.54	89.57	97.84
TOX_171	99.04	89.07	81.9	82.61	93.56
Leukemia	99.25	97.74	91.35	93.12	97.75
Friedman	1.85	2.93	2.78	4.52	2.91
mean rank					
Rank	1	4	2	5	3

number of features selected as the performance measures. This method is also applied on a COVID-19 case study to predict the affected patient's death and recovery conditions. These results are given in Section 6.

Each experiment is run 20 independent times on each of the datasets. The maximum number of iterations is 30. Each dataset is randomly divided into two groups: 80 % of the samples for training and 20 % for testing. The common parameter values of all the algorithms are as

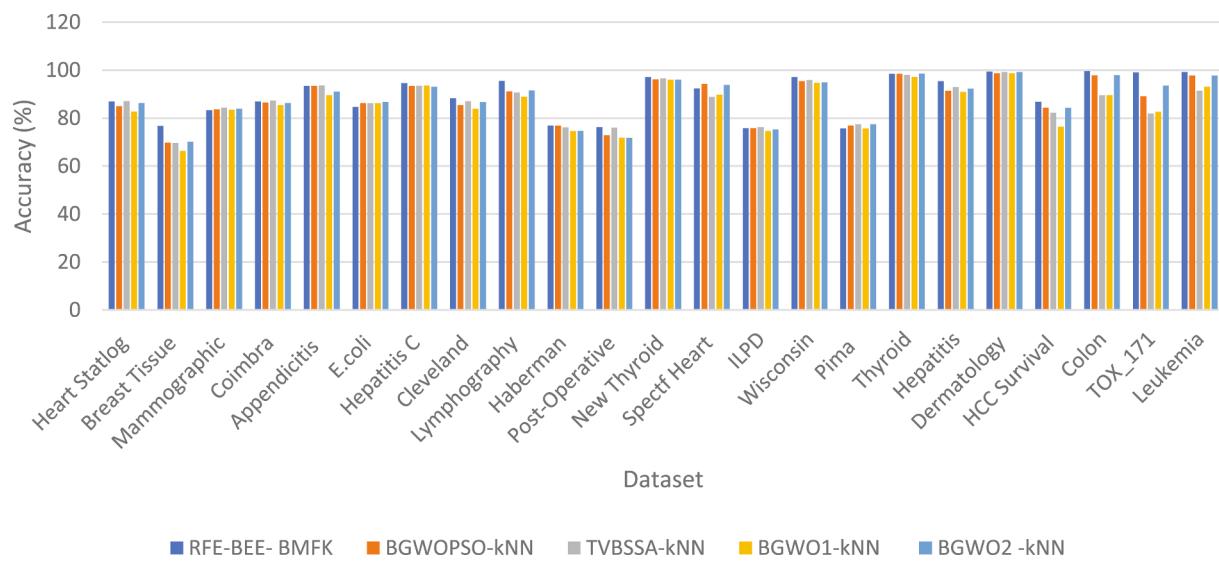


Fig. 7. Comparison of accuracies using the RFE-BEE approach and other existing methods.

Table 18
Friedman's and Holm's p values.

i	Methods	Friedman mean rank	Holm p-value	$\alpha/(k-i)$
1	BGWO1	4.52	$<10^{-5}$	0.0125
2	BGWOPSO	2.93	0.002	0.0166
3	BGWO2	2.91	0.022	0.025
4	TVBSSA	2.78	0.045	0.05

Friedman p-value: <0.00001 .

$\alpha = 0.05$, k = degrees of freedom.

follows: the population size (N) is ten, and the value of k in KNN and BMFK is 5. The value of α in the fitness function is set to 0.99. The number of dimensions (D) equals the number of features in the given dataset. The parameter settings for all the approaches are given in Table 6.

5.3. Results of the proposed BEE and RFE-BEE approach with and without hybridization

RFE-BEE is a hybrid approach consisting of filter and wrapper feature selection algorithms. Initially, we compare the Enhanced EO approach (wrapper-based method) with the EO approach, and the results are shown in Table 7. Out of the 23 datasets, the EEO obtained lower best fitness values for 13 datasets, optimal mean fitness values for 14 datasets and higher accuracies for 14 datasets. The number of features selected is higher for the EEO approach than for the EO approach.

Eight versions of the proposed EEO are developed, each integrating one variant of time-varying S and V-shaped TFs (STFs and VTFs). The performance of different versions of EEO for solving feature selection problems is evaluated using classification accuracies. The accuracies obtained by integrating four-time varying STFs are given in Table 8. In this work, we used the non-parametric Friedman test to test the significance of the proposed approach with other methods. We observe that by integrating EEO with the TV_{S2} transfer function (EEO-S2), better results are obtained for 11 datasets, followed by EEO-S4 for five datasets out of 23.

Table 9 gives the accuracies obtained by integrating the EEO approach with four TVV transfer functions. By integrating TV_{V3} (EEO-V3), we observe that better results are obtained for nine datasets out of 23. EEO-V2 is placed second, achieving better accuracies for seven datasets. When comparing the accuracies of EEO-S2 with EEO-V3, EEO-S2 outperforms EEO-V3 for nearly 74 % of the datasets. Hence, we used

the TV_{S2} function for converting EEO to binary form (BEE).

The accuracies obtained with and without hybridization of ReliefF, Fuzzy Entropy methods (filter-based approaches) and Binary EE approach (wrapper-based approach) are given in Table 10. We also compared the proposed RFE-BEE approach using BMFK and KNN as learning models. The RFE-BEE-BMFK approach delivered better results among all the six methods. Improved classification accuracy of 89.45 % is obtained by integrating the filter and wrapper-based feature selection approaches compared to the accuracies of 89.06 % and 84.44 % using the wrapper and filter-based FS approaches alone. These results show the hybrid method's superiority compared to the individual approaches.

The RFE-BEE-BMFK algorithm obtained the best rank, delivering higher accuracies for 11 datasets, followed by the RFE-BEE-KNN for nine datasets. This shows the efficiency of the BMFK classifier compared to the KNN as a learning algorithm. Also, an ensemble RF + FE-BMFK obtained an accuracy of 84.44 %, better than the single RF method with an accuracy of 82.39 %. This shows the ability of an ensemble to deliver higher accuracies.

5.4. Comparison of the RFE-BEE approach with other FS approaches using heuristic algorithms

In the second phase of the experimental study, we compare the RFE-BEE method with three other popular FS approaches using heuristic algorithms – GA (Chyzyk et al., 2014), PSO (Unler et al., 2011) and GWO (Emary et al., 2016) in terms of fitness values, accuracies and number of features selected. BMFK is used as the learning model. Table 11 reports the performance of the RFE-BEE approach in terms of best fitness values. It is found that the RFE-BEE approach delivered lower fitness values for ten datasets, followed by the GWO approach for nine datasets out of 23.

Regarding the mean fitness values, the results are shown in Table 12. The results show that the RFE-BEE approach gives optimal mean fitness values for 16 out of 23 datasets, nearly 70 % of the total datasets. GWO stood second, achieving lower mean fitness values for two datasets – Haberman and Spectf Heart.

Table 13 reports the classification accuracies obtained by different approaches for 23 datasets. According to these results, it is observed that the RFE-BEE ranked first, achieving higher classification accuracies for 16 datasets, followed by the GWO approach in second place. When considering the high dimensional datasets like Colon, TOX_171 and Leukemia, we obtained 99.58 %, 99.04 % and 99.25 % (nearly 100 %) accuracies, respectively, using a minimal number of features. The

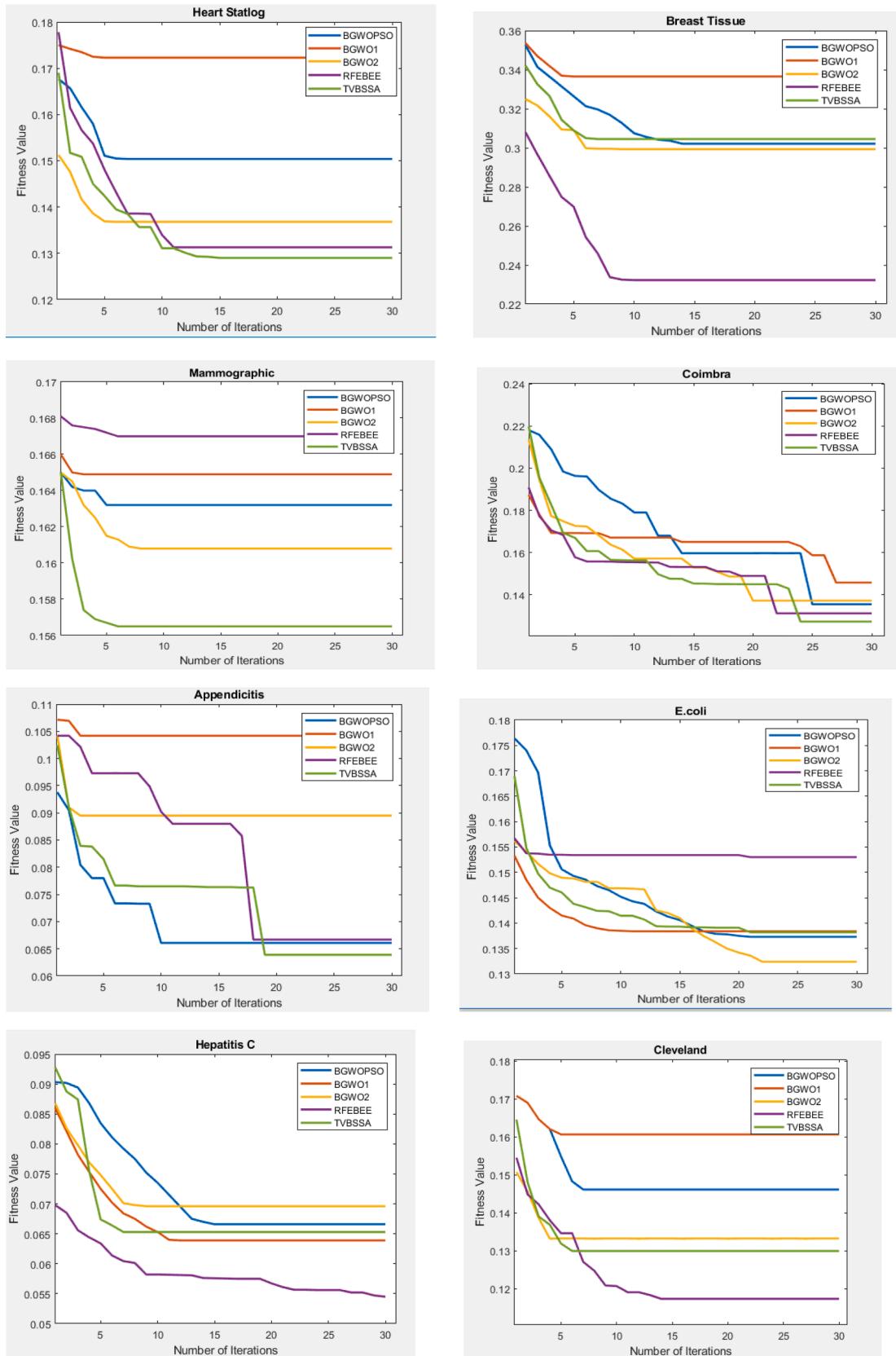


Fig. 8. Convergence behaviour of the RFE-BEE approach and other existing methods on 23 datasets.

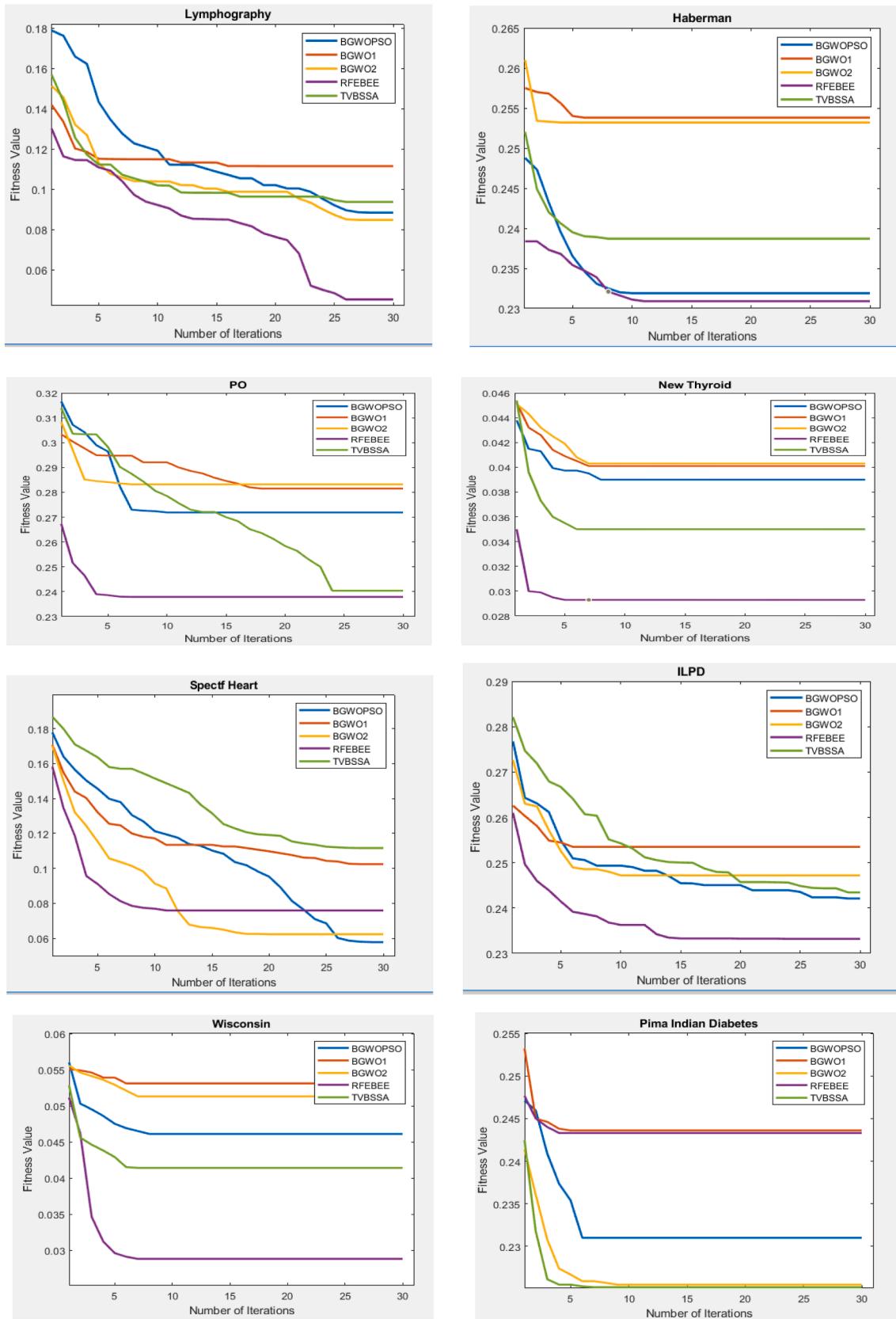


Fig. 8. (continued).

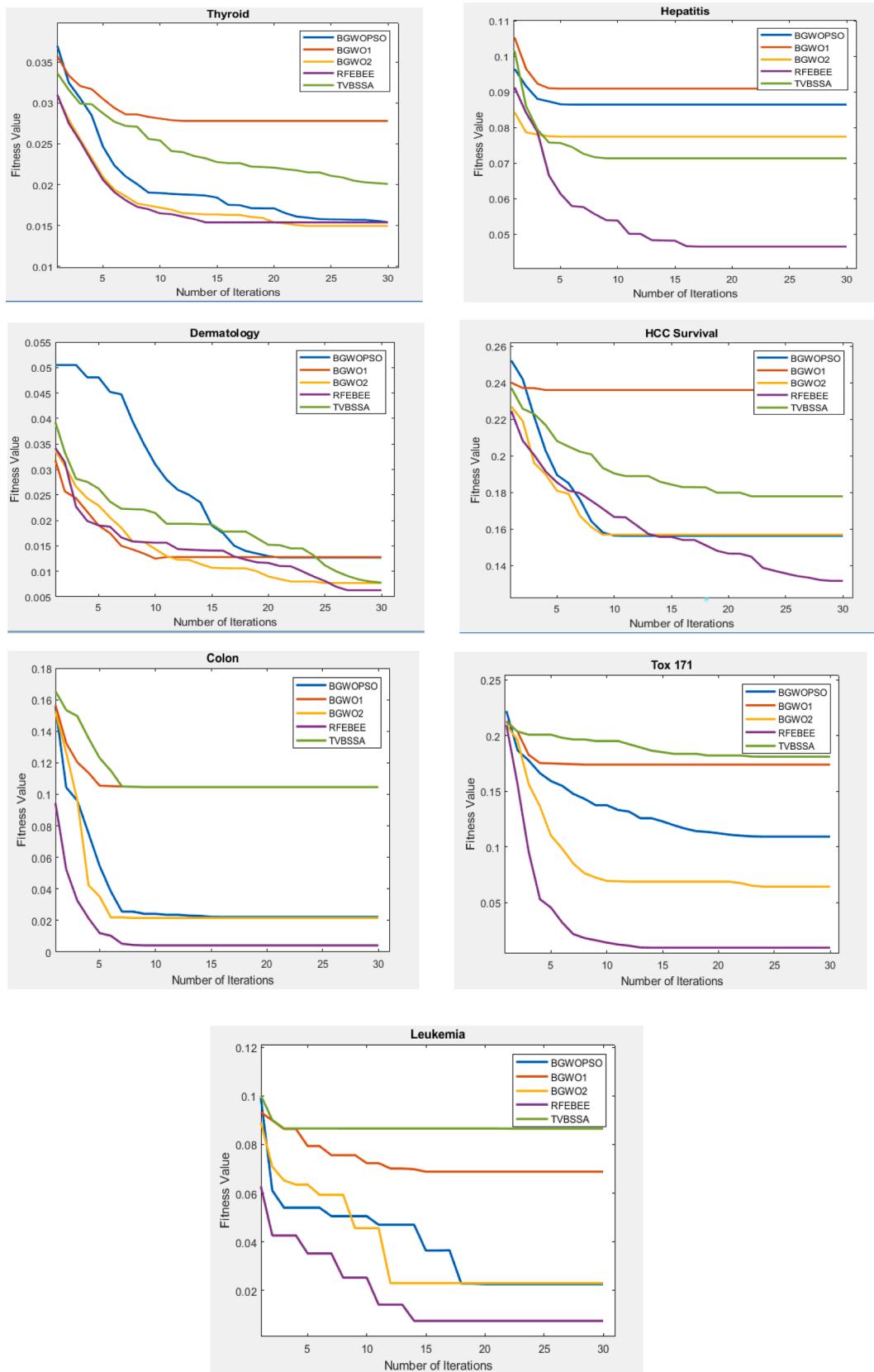
**Fig. 8. (continued).**

Table 19

Comparison of average number of features selected using the RFE-BEE approach and other existing methods.

Dataset	RFE-BEE-BMFK	BGWO PSO- KNN	TVBSSA-KNN	BGWO1-KNN	BGWO2-KNN
HS	5	3.7	5.65	7.05	3.9
BT	3.25	2.45	2.55	3.7	2.05
Mammo	1.8	2.3	2.75	2.85	2.1
Coimbra	3.8	3.8	4.2	5.1	3.85
Appen	2.15	1.7	1.85	2	1.6
E.coli	5.1	5.1	4.7	5.35	4.75
Hep-C	5.55	4.2	4.95	6.5	4.9
Cleve	5.6	5.5	5.8	7.15	5.75
Lymph	7.55	5.4	8.75	9.95	4.95
Haber	1.05	1.65	1.75	1.65	1.7
PO	3.7	1.95	3.1	3	2.15
NT	2.99	2.25	2.55	2.8	2.3
Spectf	16.45	11.9	26.6	27.5	10.95
ILPD	2.95	3.5	4.3	5.1	3.1
Wisconsin	6.35	5.45	13.7	13.4	2.8
Pima	3.25	3.9	4.4	4.65	3.6
Thyroid	4.65	4.85	8.95	10.7	4.8
Hepatitis	4.3	3	7.75	8.1	2.65
Derm	12.1	13.15	17.3	20.3	10.1
HCC	12.15	8.25	26.05	24.5	3.85
Colon	32.45	308.35	1113.2	1059.1	199.45
TOX_171	471.5	1757.7	3624.7	3738.1	1025.5
Leukemia	299.5	972.9	3683.1	3611.4	879.55
Friedman mean rank	2.52	2.2	3.91	4.54	1.83
Rank	3	2	4	5	1

number of features selected is presented in [Table 14](#).

As per the results, the RFE-BEE approach delivers the highest accuracy of 99.58 %, using only 32 features for the Colon dataset. Considering the TOX_171 dataset, RFE-BEE achieved the highest accuracy of 99.04 %, using just 471 features. Similarly, for the Leukemia dataset, RFE-BEE achieved the highest accuracy of 99.25 %, using only 300 features. This shows the ability of the RFE-BEE algorithm to classify high-dimensional datasets. [Fig. 5](#) illustrates the classification accuracies obtained using RFE-BEE and other heuristic algorithms for all 23 datasets.

[Table 14](#) shows that the RFE-BEE approach exhibited superior performance regarding the number of features selected compared with GA and PSO. GWO attained the first rank using a fewer number of features for seven datasets. The RFE-BEE approach works very well, especially for high-dimensional datasets like Colon, TOX_171 and Leukemia, delivering excellent results using a minimal number of features. The

Table 20

Comparison of accuracies obtained by the RFE-BEE method with other improved versions of EO.

Dataset	RFE-BEE- BMFK	CEO	BEO	IBEO
HS	86.87	83.44	84.49	84.44
BT	76.77	68.12	69.34	69.98
Mammo	83.3	84.87	83.75	84.18
Coimbra	86.88	83.65	83.03	84.43
Appen	93.33	92.42	91.06	92.2
E.coli	84.7	87.35	86.27	85.78
Hep-C	94.55	93.07	93.11	93.34
Cleve	88.27	81.56	82.56	87.24
Lymph	95.48	88.77	90.51	91.85
Haber	76.91	75.28	75.43	74.65
PO	76.21	74.59	74.31	72.93
NT	97.07	96.2	95.61	95.66
Spectf	92.41	86.52	88.52	90.51
ILPD	75.8	75.29	75.29	74.83
Wisconsin	97.12	96.02	95.3	95.27
Pima	75.67	77.03	76.28	77.57
Thyroid	98.46	97.58	97.97	98.34
Hepatitis	95.35	93.07	91.31	90.29
Derm	99.37	97.31	98.65	98.95
HCC	86.85	77.63	80.35	79.37
Colon	99.58	78.59	81.05	85.15
TOX_171	99.04	78.81	81	90
Leukemia	99.25	94.81	97.25	95.35
Friedman mean rank	1.39	3.07	2.89	2.65
Rank	1	4	3	2

average number of features selected by each approach is shown in [Fig. 6](#).

5.5. Comparison of the RFE-BEE approach with other existing methods

In the third phase of the experimental study, we compare the proposed RFE-BEE approach with four other existing methods and three modified versions of the EO algorithm proposed by different authors. The four existing methods used for comparison are BGWOPSO ([Al-Tashi et al., 2019](#)), TVBSSA ([Faris et al., 2019](#)), BGWO1 and BGWO2 by ([Emary et al., 2016](#)). Here KNN is used as an induction algorithm for comparison.

[Tables 15 and 16](#) provide the best and mean fitness values using different methods. It is seen that the proposed RFE-BEE is placed first, providing optimal best fitness values for nine datasets. BGWOPSO attained the second rank delivering optimal best fitness values for five datasets out of 23.

Regarding the mean fitness values, RFE-BEE delivered lower mean fitness values for 14 datasets out of 23, >60 % of the datasets used. The TVBSSA method attained the second rank, delivering lower mean fitness

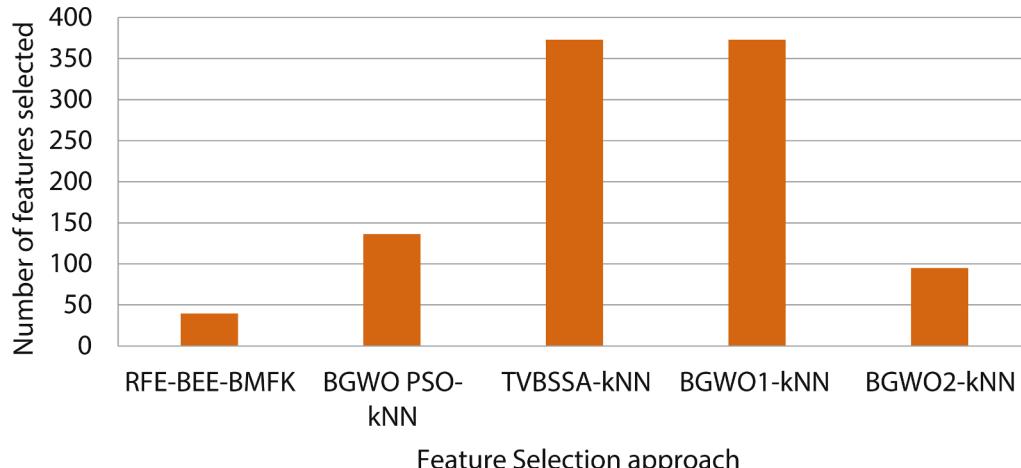
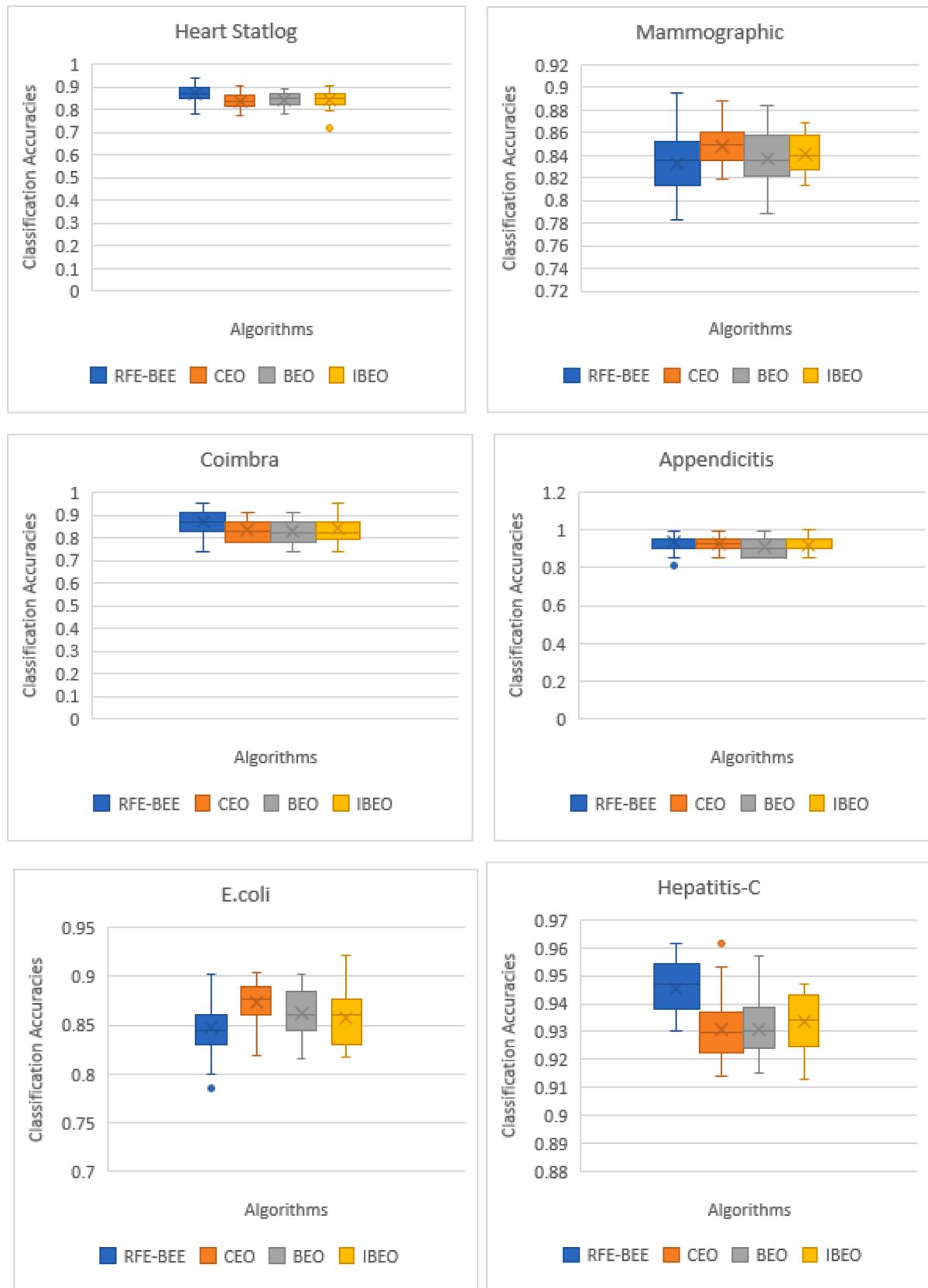


Fig. 9. Number of features selected using RFE-BEE and other existing methods.

**Fig. 10.** Box plots of the classification accuracies of different versions of EO.

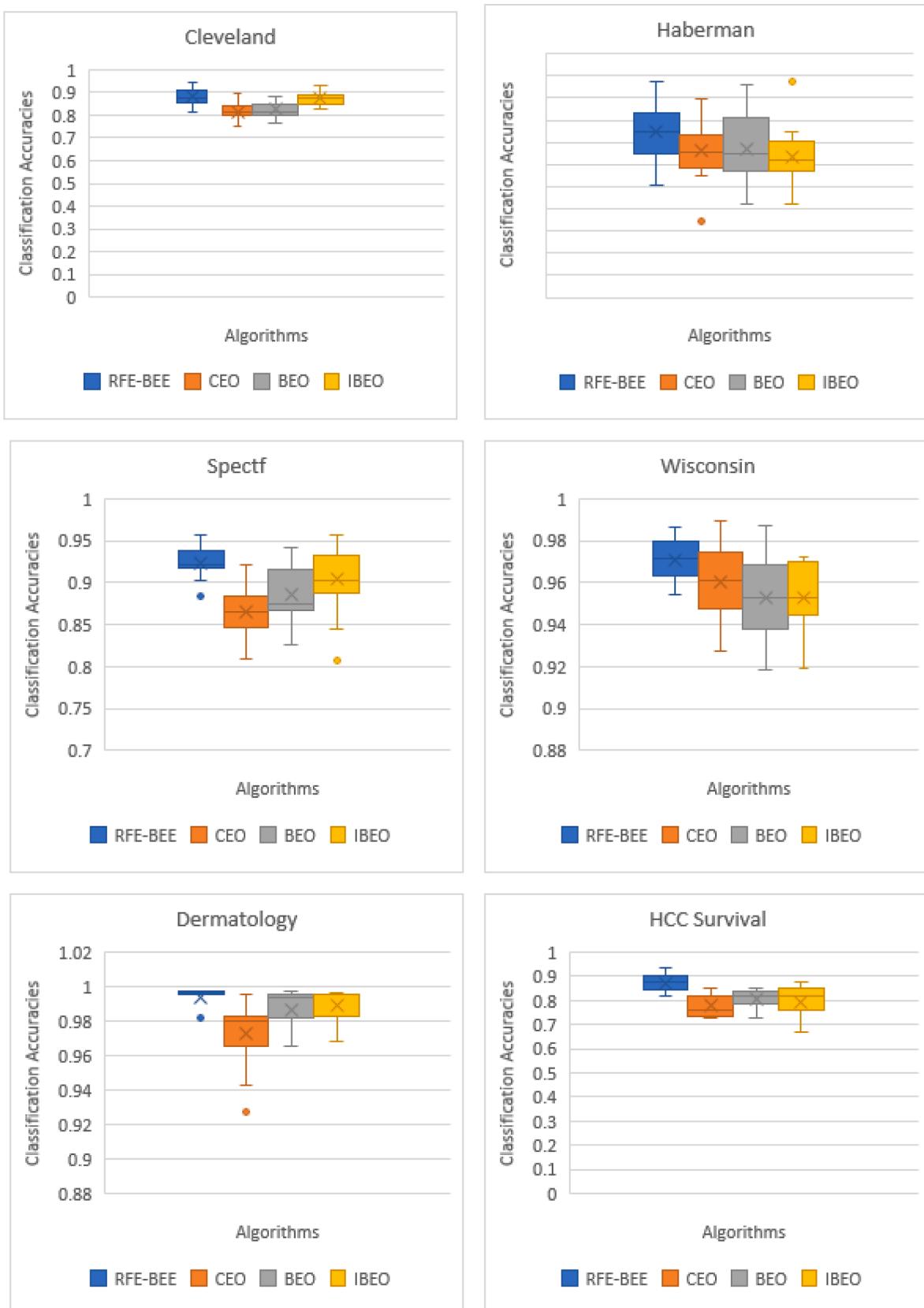


Fig. 10. (continued).

Table 21

Comparison of accuracies obtained by the RFE-BEE approach with seven other existing methods on Parkinson's dataset.

RFE-BEE	BGWOPSO	BGWO1	BGWO2	TVBSSA	CEO	IBEO	BEO
99.45	98.35	97.73	98.60	98.24	98.29	98.84	98.54

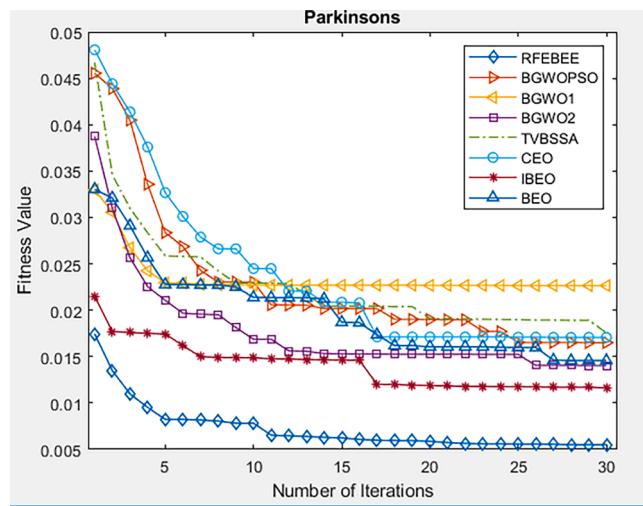


Fig. 11. Convergence curves of the RFE-BEE approach and other methods on Parkinson's dataset.

values for six datasets. BGWO1 is assigned the last rank, as it did not report a better outcome in any dataset than other algorithms.

Similarly, Table 17 shows that the RFE-BEE method exhibited a sturdy performance attaining higher accuracy values for 14 datasets with an average accuracy of 89.53 %. TVBSSA stood second, achieving higher accuracy values for six datasets. Of all the five methods, BGWO1 attained the last rank. Fig. 7 compares the accuracies using RFE-BEE and other existing methods for all 23 datasets.

We performed the non-parametric Friedman test for statistical analysis. The Friedman mean ranks are calculated for all five methods, which are given in Table 18. The Friedman p-value obtained is <0.00001 ($p < 0.00001$), which signifies a significant difference in the results obtained using the five methods.

Table 18 also presents the results of Holm's post hoc method, which

Table 22
Description of four microarray datasets used.

Dataset	Instances	Features	Classes
Colon	62	2000	2
DLBCL	77	5469	2
Lung Cancer	203	12,600	5
SRBCT	83	2308	4

Table 23
Comparison of the RFE-BEE approach with other existing methods for classification of microarray datasets in terms of accuracies.

Dataset	RFE-BEE-BMFK	BGWO PSO	TVBSSA	BGWO1	BGWO2
Colon	99.58	97.78	89.54	89.57	97.84
DLBCL	100	99.08	96.99	97.7	98.63
Lung Cancer	99.43	99.51	98.94	98.81	99.42
SRBCT	100	100	99.5	99.63	99.02
Friedman mean rank	1.375	1.875	4.5	4	3.25
Rank	1	2	5	4	3

compares the RFE-BEE approach with the other four algorithms. From the p-values in the table, we find a significant difference between the proposed RFE-BEE and other algorithms.

The convergence trends of the RFE-BEE approach and other existing methods based on the fitness values on 23 datasets are revealed in Fig. 8. It presents the mean convergence of 20 runs for 30 iterations. The best algorithm is the one that settles to a smaller fitness value in a lower number of iterations. Overall, the RFE-BEE approach shows a better convergence behaviour in many datasets. We can observe that RFE-BEE is exceptionally superior in Breast Tissue, Hepatitis-C, Cleveland, Lymphography, Haberman, PO, New Thyroid, Wisconsin, Hepatitis, Dermatology, HCC Survival, Colon, TOX_171 and Leukemia datasets. This shows the superiority of the RFE-BEE approach over other methods in convergence behaviour.

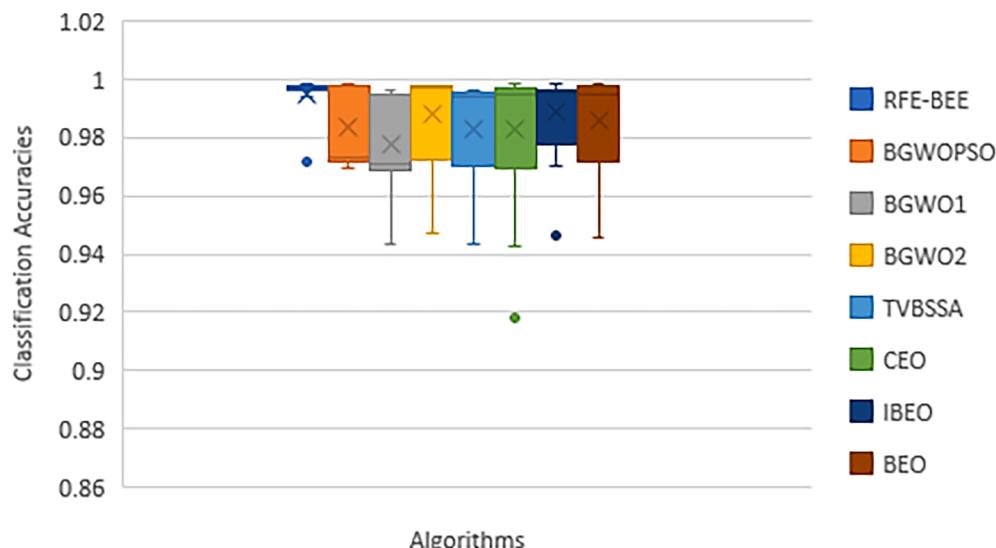


Fig. 12. Boxplots of the classification accuracies on Parkinson's dataset.

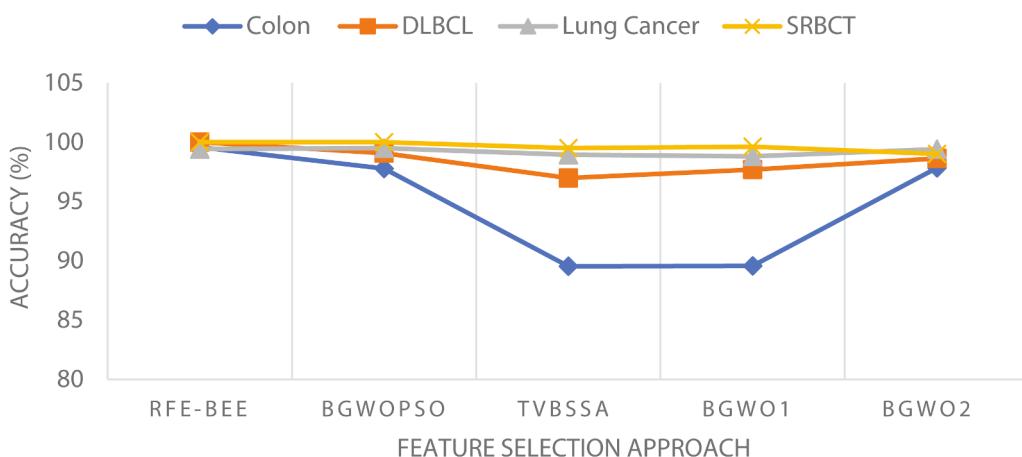


Fig. 13. Comparison of accuracies obtained using the RFE-BEE approach and other existing methods.

Table 24

Comparison of the RFE-BEE approach with other existing methods for classification of microarray datasets in terms of precision.

Dataset	RFE-BEE-BMFK	BGWOPSO	TVBSSA	BGWO1	BGWO2
Colon	99.38	97.5	86.56	86.88	97.19
DLBCL	100	99.13	97.27	97.86	98.67
Lung Cancer	95.62	97.43	91.42	89.69	99.42
SRBCT	100	100	99.08	99.29	98.08
Friedman mean rank	1.625	1.875	4.5	4	3
Rank	1	2	5	4	3

Table 25

Comparison of the RFE-BEE approach with other existing methods for classification of microarray datasets in terms of sensitivity.

Dataset	RFE-BEE-BMFK	BGWOPSO	TVBSSA	BGWO1	BGWO2
Colon	99.72	98.17	91.46	91.72	98.39
DLBCL	100	97.75	93.38	94.13	96.92
Lung Cancer	98.35	99.16	98.12	97.56	99.12
SRBCT	100	100	98.75	99.06	97.77
Friedman mean rank	1.625	1.875	4.5	4	3
Rank	1	2	5	4	3

The number of features selected by different algorithms is given in [Table 19](#). We observe that the BGWO2 approach uses a minimal number of features and is placed first. In the case of high-dimensional datasets like Colon, TOX_171 and Leukemia, RFE-BEE uses fewer features than any other algorithm. [Fig. 9](#) depicts the average number of features selected by each method.

5.5.1. Comparison of the RFE-BEE approach with other improved versions of EO

In this section, we compare the accuracies obtained by the RFE-BEE approach with other improved versions of EO – Chaotic Equilibrium Optimizer ([Sayed et al., 2022](#)), Binary Equilibrium Optimizer ([Gao et al., 2020](#)) and Improved Binary Equilibrium Optimizer ([Elmanakhly et al., 2021](#)). [Table 20](#) unveils the results of the accuracies compared with the other improved versions of the equilibrium algorithm for feature selection. RFE-BEE stood first of all the modified versions achieving higher classification accuracies for nearly 87 % of the datasets showing its superiority in solving FS problems. IBEO attained the second rank and the last rank by the CEO.

Table 26

Comparison of the RFE-BEE approach with other existing methods for classification of microarray datasets in terms of F-measure.

Dataset	RFE-BEE-BMFK	BGWOPSO	TVBSSA	BGWO1	BGWO2
Colon	99.59	97.83	88.94	89.23	97.78
DLBCL	100	98.44	95.29	95.96	97.79
Lung Cancer	96.97	98.23	94.65	93.46	99.27
SRBCT	100	100	98.91	99.17	97.92
Friedman mean rank	1.625	2.625	4.5	4.0	3.0
Rank	1	2	5	4	3

[Fig. 10](#) depicts the boxplot graphs of the RFE-BEE method and other modified EO versions for 12 datasets. These boxplots give a better understanding of the distribution of the accuracy outcomes. When examining the boxplots, we find that RFE-BEE stably identified the optimal solutions in Heart Statlog, Coimbra, Appendicitis, Hepatitis-C, Cleveland, Haberman, Spectf Heart, Wisconsin, Dermatology and HCC Survival datasets. In the Mammographic and E. coli datasets, the CEO has reached a stable solution. The boxplot graphs indicate that the exploration capability of RFE-BEE is better than the other versions of EO.

5.5.2. Comparison of the RFE-BEE approach with seven other existing methods on Parkinson's dataset

An additional binary dataset, Parkinson's ([Dua & Graff, 2019](#)), with 22 features and 195 samples, is used to compare RFE-BEE with other existing methods. [Table 21](#) gives the classification accuracy results using the RFE-BEE approach and seven other existing methods. The highest accuracy of 99.45 % is obtained using the RFE-BEE approach followed by IBEO. BGWO1 attained the lowest accuracy of 97.73 %. [Fig. 11](#) shows the convergence behaviour of these methods on the Parkinson's dataset. We observe that RFE-BEE shows a better convergence behaviour than the other methods. It stably settled to a lower fitness value in a small number of iterations. [Fig. 12](#) depicts the boxplots of the accuracy outcomes of the RFE-BEE and other methods on the Parkinson's dataset. It is seen that the RFE-BEE approach obtained the optimal solution compared with all other algorithms.

6. Microarray datasets and COVID-19 dataset

6.1. Application of the RFE-BEE approach on high-dimensional microarray datasets

Four gene expression datasets of high dimensions and a small number of instances are used to test the efficiency of the proposed approach

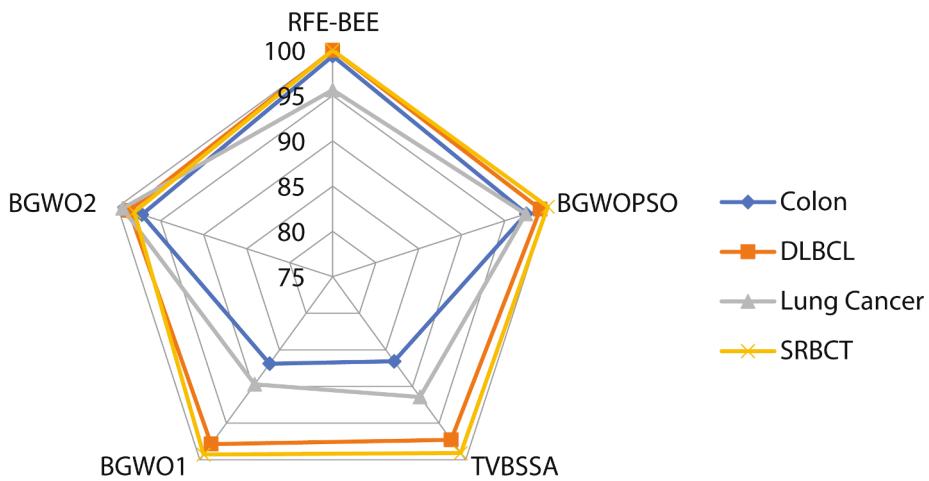


Fig. 14. Comparison of the RFE-BEE approach and other existing methods in terms of precision values.

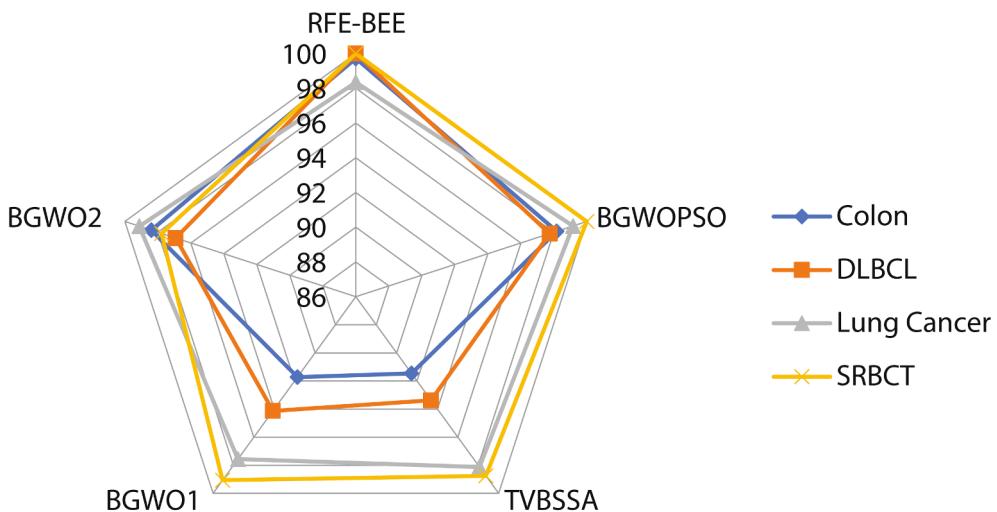


Fig. 15. Comparison of the RFE-BEE approach with other existing methods in terms of sensitivity values.

Table 27

Comparison of the RFE-BEE approach with other existing methods in terms of number of features selected.

Dataset	RFE-BEE-BMFK	BGWO PSO	TVBSSA	BGWO1	BGWO2
Colon	32.45	308.35	1113.2	1059.1	199.45
DLBCL	419.65	673.45	2927.9	2808.6	1445.1
Lung Cancer	2188.3	2024.5	7149.3	6905.1	3608.6
SRBCT	20.75	345.75	1171.3	1152.7	570.95
Friedman mean rank	1.25	2	5	4	3
Rank	1	2	5	4	3

and compared with four other approaches. It is challenging for feature selection approaches to deal with datasets containing many attributes and a small number of instances. Large no. of attributes increases the search space because the heuristic methods cannot explore all the regions, and a small number of instances will not be sufficient to train the learning method (Bolón-Canedo et al., 2015).

In this section, we used four popular datasets, two binary and two multiclass microarray datasets, to test the efficiency of the proposed

approach in terms of Accuracy, Precision, Sensitivity, F-measure and number of features selected. Each method is run for thirty iterations and 20 independent runs, and the results are observed. The description of the micro-array datasets (Dabba et al., 2021) is given in Table 22.

There are only two classes in the binary classification; one is positive (P) and the other negative (N). The four possible outcomes from the classification are true positive (TP), false positive (FP), true negative (TN) and false negative (FN). The metrics used are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (33)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (34)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (35)$$

$$\text{F1 - Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (36)$$

The accuracies obtained by the RFE-BEE approach and other existing methods are reported in Table 23. The RFE-BEE approach obtained the highest accuracies for three datasets out of four, i.e., for 75 % of the total

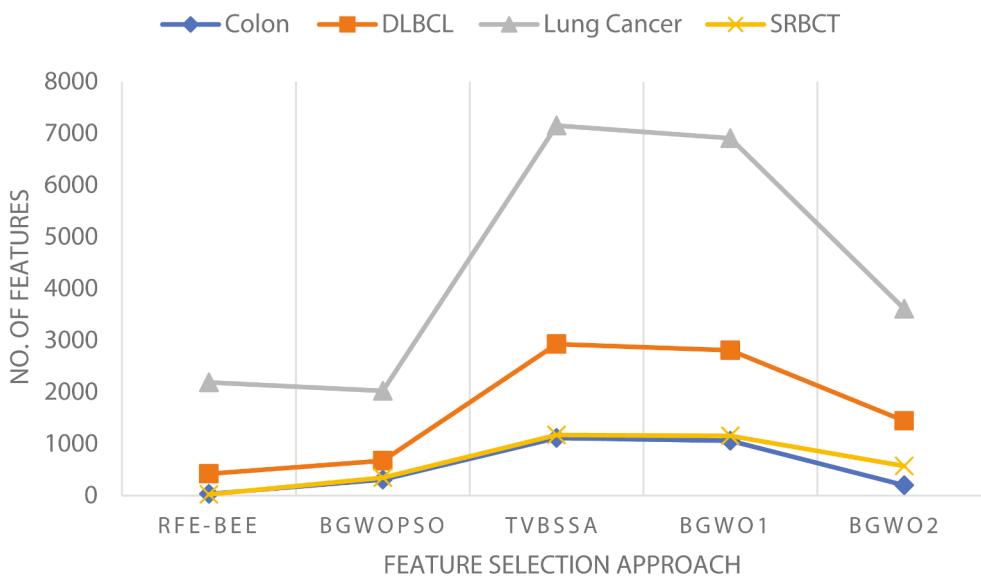


Fig. 16. Number of features selected using the RFE-BEE approach and other existing methods.

Table 28
Description of the COVID-19 dataset.

No.	Feature	Description
1	id	The ID of patients
2	location	The location where patient belongs to
3	country	The country where patient belongs to
4	gender	Gender of patients
5	age	The ages of patients
6	vis_wuhan	Whether the patients visited Wuhan, China
7	from_wuhan	Whether the patients from Wuhan, China
8	symptom1	Fever
9	symptom2	Pneumonia
10	symptom3	Cough
11	symptom4	Difficulty in breathing
12	symptom5	Fatigue
13	symptom6	Malaise
14	diff_sym_hos	The day's difference between the symptoms being noticed and visit the hospital
15	result	Class: Death or Recovery

datasets used, with an average accuracy of 99.75 %. This approach achieved 100 % accuracies for two datasets, DLBCL and SRBCT, proving its efficiency in classifying high-dimensional datasets. The accuracies obtained for four microarray datasets using RFE-BEE and other existing methods are shown in Fig. 13.

We also used three more metrics to prove the efficiency of the proposed RFE-BEE approach. The results of precision, sensitivity and F-measure outcomes are given in Tables 24–26. Higher values of these measures indicate better performance of the algorithm. The RFE-BEE approach attained the first rank obtaining higher precision values of 99.38 %, 100 % and 100 % for Colon, DLBCL and SRBCT datasets, respectively, with an average precision value of 98.75 %. BGWOPSO is ranked second, attaining a higher precision value for one dataset, lung cancer. Table 24 reports the precision values of all the FS approaches.

Regarding the sensitivity results, the RFE-BEE approach obtained the highest outcomes of 99.72 %, 100 % and 100 % for Colon, DLBCL and SRBCT datasets, respectively, with an average of 99.52 %, which are tabulated in Table 25. Fig. 14 and Fig. 15 depict the precision and sensitivity values using the RFE-BEE approach and other existing methods, respectively.

Similarly, the RFE-BEE approach attained higher F-measure values and is ranked first, proving its capability even in microarray datasets classification. These are tabulated in Table 26.

Table 27 shows that the proposed RFE-BEE approach uses a minimal number of features for very high dimensional datasets, using an average of 665.29 features for four datasets. This method selects fewer features for three datasets out of four than the other existing methods. Fig. 16 shows the number of features selected for each of the four microarray datasets.

6.2. Application of the RFE-BEE approach on a COVID-19 case study

COVID-19 is an infectious and contagious disease caused by SARS-CoV-2. The first case was known to be detected in Wuhan, China, and it spread rapidly worldwide, causing many deaths (Chowdhury et al., 2020). Designing effective techniques is required to combat the spread of this type of disease. Many techniques have been developed for the early diagnosis and detection of COVID-19. In this section, the proposed RFE-BEE approach is applied to a real-world COVID-19 dataset (Nadimi-Shahraki et al., 2022) to predict an affected patient's death and recovery conditions using the different factors given in the dataset. This dataset consists of 15 features. A detailed description of the COVID-19 dataset used is given in Table 28. Each experiment is run 20 times, and the results are compared to evaluate the performance of RFE-BEE with other algorithms.

Fig. 17 illustrates the comparison of RFE-BEE with other approaches on the COVID-19 dataset. Fig. 17a shows the boxplot analysis of classification accuracies of RFE-BEE and other algorithms. It is seen that the proposed RFE-BEE algorithm achieved the highest mean classification accuracy of 97.23 % compared to the other seven approaches using an average of 6.5 features. The best accuracy value obtained by RFE-BEE is 99.14 % using a subset of only four features (1–3–5–14), and the worst obtained is 95.04 %. Based on the results, age is the most selected feature, and symptom 6 is never selected by the RFE-BEE approach in 20 runs. The CEO approach obtained the worst mean accuracy of 95.56 % of all the algorithms. Fig. 17b shows the no. of features selected by different approaches, and Fig. 17c shows the convergence behaviour of RFE-BEE and other algorithms. It can be seen that RFE-BEE settled to a lower fitness value in less no. of iterations compared to the other algorithms.

By observing these results, it can be inferred that the proposed RFE-BEE method exhibits superior performance compared with GA, PSO, GWO and many other existing methods in terms of accuracies and fitness values. BMFK classifier is mainly concerned with class distributions and mitigates the class imbalance problem. Hence, the proposed approach

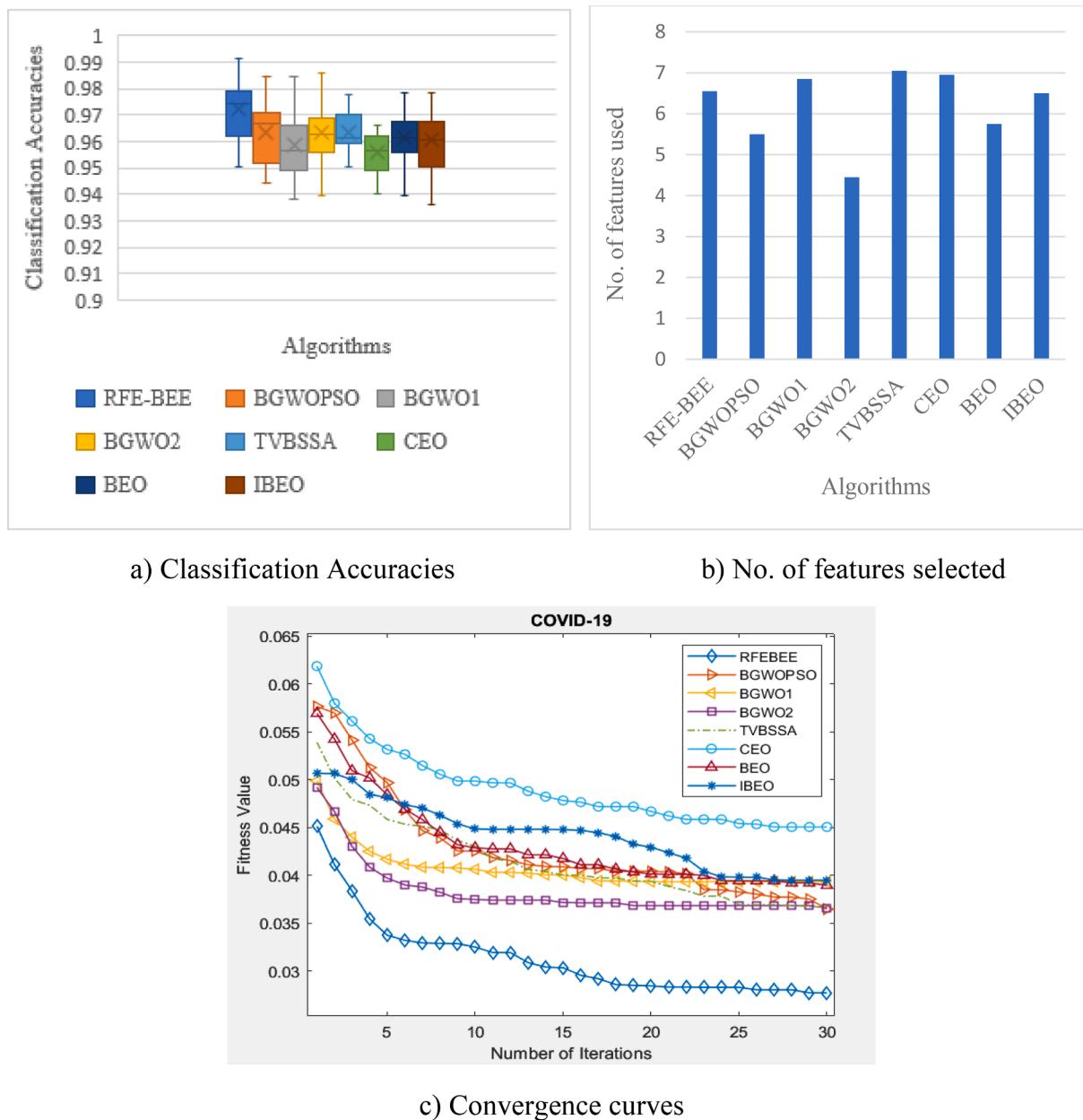


Fig. 17. Comparison of RFE-BEE with other approaches on COVID-19 dataset.

works well, especially for high dimensional datasets with a smaller number of instances delivering higher classification accuracies using fewer features.

As the BMFK classifier uses the mean values calculated from the classes, this process takes more time than other methods, especially for large datasets. Another limitation is the selection of threshold values (value of n to select the best features from the filter method). For smaller values of n , informative features are lost. A larger value of n makes the selection process hard. Sometimes, the union may lead to selecting a whole set of features. Hence, finding a threshold value is one of the crucial factors for the smooth performance of the algorithm.

7. Conclusion

This paper introduces a novel hybrid feature selection approach (RFE-BEE) that integrates filter and wrapper techniques for feature selection. Two diverse filter techniques, ReliefF and Fuzzy Entropy, are combined to remove irrelevant features and decrease the search space.

The Equilibrium Optimization method is combined with Opposition Based Learning, Cauchy Mutation, and a novel search strategy to obtain an Enhanced Equilibrium method (EEO). The Opposition Based Learning strategy improves the diversity at the initialization phase. Cauchy Mutation avoids the solutions of getting trapped in the local optima, and the novel search strategy improves the exploration ability of the algorithm. Eight time-varying S and V-shaped transfer functions are integrated with EEO to convert it into binary form. The union of top- n features from each filter technique is given as input to the Binary Enhanced Equilibrium Optimizer to extract the significant features. Fuzzy KNN based on Bonferroni mean is used as the induction algorithm.

The ability of the RFE-BEE approach is tested using 23 standard datasets and four microarray datasets. This method is also tested on a real-world dataset, COVID-19, with the highest accuracy of 97.23 %. The results show that the proposed approach outperformed other FS approaches in terms of fitness values, accuracy, precision, sensitivity, and F-measure. RFE-BEE exhibited a credible performance, especially for high-dimensional datasets utilizing a minimal number of features.

Based on Friedman's test, RFE-BEE is ranked first compared with other existing methods in the literature. In future, this work can be extended by integrating a technique that automatically finds the threshold value (value of n) to select the informative features from the ensemble of filter methods.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- Abdel-Basset, M., Chang, V., & Mohamed, R. (2020). A novel equilibrium optimization algorithm for multi-thresholding image segmentation problems. *Neural Computing and Applications*, 33(17), 10685–10718. <https://doi.org/10.1007/s00521-020-04820-y>
- Alcalá-Fdez, J., Fernandez, A., Derrac, J., Garcia, S., Sanchez, L., & Herrera, F. (2011). Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Multiple Valued Logic Soft Computer*, 17(2–3), 255–287.
- Al-Tashi, Q., Abdul Kadir, S. J., Rais, H. M., Mirjalili, S., & Alhussian, H. (2019). Binary Optimization Using Hybrid Grey Wolf Optimization for Feature Selection. *IEEE Access*, 7, 39496–39508. <https://doi.org/10.1109/ACCESS.2019.2906757>
- Awadallah, M. A., Al-Betar, M. A., Braik, M. S., Hammouri, A. I., Doush, I. A., & Zitar, R. A. (2022). An enhanced binary Rat Swarm Optimizer based on local-best concepts of PSO and collaborative crossover operators for feature selection. *Computers in Biology and Medicine*, 147, Article 105675. <https://doi.org/10.1016/j.combiomed.2022.105675>
- Awadallah, M. A., Hammouri, A. I., Al-Betar, M. A., Braik, M. S., & Elaziz, M. A. (2022). Binary Horse herd optimization algorithm with crossover operators for feature selection. *Computers in Biology and Medicine*, 141, Article 105152. <https://doi.org/10.1016/j.combiomed.2021.105152>
- Bolón-Canedo, V., Sánchez-Marzoño, N., & Alonso-Betanzos, A. (2015). Feature selection for high-dimensional data. *Springer*. <https://doi.org/10.1007/978-3-319-21858-8>
- Bolón-Canedo, V., & Alonso-Betanzos, A. (2019). Ensembles for feature selection: A review and future trends. *Information Fusion*, 52, 1–12. <https://doi.org/10.1016/j.inffus.2018.11.008>
- Bonferroni, C. (1950). Sullemedie multiple di potenze. *Bulletino dell'Unione Matematica Italiana*, 5, 267–270.
- Chen, G., & Chen, J. (2015). A novel wrapper method for feature selection and its applications. *Neurocomputing*, 159, 219–226. <https://doi.org/10.1016/j.neucom.2015.01.070>
- Chowdhury, M. E., Rahman, T., Khandakar, T., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Islam, K.R., Khan, M.S., Iqbal, A., & Al Emadi, N. (2020) Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* 8, 132665–132676. Doi: 10.1109/ACCESS.2020.3010287.
- Chyzyk, D., Savio, A., & Graña, M. (2014). Evolutionary elm wrapper feature selection for Alzheimer's disease cad on anatomical brain MRI. *Neurocomputing*, 128, 73–80. <https://doi.org/10.1016/j.neucom.2013.01.065>
- Dabba, A., Tari, A., Meftali, S., & Mokhtari, R. (2021). Gene Selection and Classification of Microarray Data Method Based on Mutual Information and Moth Flame Algorithm. *Expert Systems with Applications*, 114012. <https://doi.org/10.1016/j.eswa.2020.114012>
- De Luca, A., & Termini, S. (1972). A definition of a non-probabilistic entropy in the setting of fuzzy sets theory. *Information and Control*, 20, 301–312. [https://doi.org/10.1016/S0019-9958\(72\)90199-4](https://doi.org/10.1016/S0019-9958(72)90199-4)
- Dessì, N., & Pes, B. (2015). Similarity of feature selection methods: An empirical study across data intensive classification tasks. *Expert Systems with Applications*, 42, 4632–4642. <https://doi.org/10.1016/j.eswa.2015.01.069>
- Dorigo, M., Birattari, M., & Stützle, T. (2006). Ant colony optimization. *IEEE Computational Intelligence Magazine*, 1, 28–39. <https://doi.org/10.1109/3477.484436>
- Drotar, P., Gazda, J., & Smekal, Z. (2015). An experimental comparison of feature selection methods on two-class biomedical datasets. *Computers in Biology and Medicine*, 66, 1–10. <https://doi.org/10.1016/j.combiomed.2015.08.010>
- Dua, D., & Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
- Eberhart, R., & Kennedy, J. (1995). A new optimizer using particle swarm theory. in: Proc. Sixth Int. Symp. Micro Mach. Hum. Sci. MHS, 95, pp. 39–43. 10.1109/MHS.1995.494215.
- Elmanakhly, D. A., Saleh, M. M., & Rashed, E. A. (2021). An Improved Equilibrium Optimizer Algorithm for Features Selection: Methods and Analysis. *IEEE Access*, 9, 120309–120327. <https://doi.org/10.1109/ACCESS.2021.3108097>
- Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, 172, 371–381. <https://doi.org/10.1016/j.neucom.2015.06.083>
- Faramarzi, A., Heidarnejad, M., Stephens, B., & Mirjalili, S. (2020). Equilibrium optimizer: A novel optimization algorithm. *Knowledge-Based Systems*, 191, 105–190. <https://doi.org/10.1016/j.knosys.2019.105190>
- Faris, H., Heidari, A. A., Al-Zoubi, A. M., Mafarja, M., Aljarah, I., Eshtay, M., et al. (2019). Time-varying hierarchical chains of salps with random weight networks for feature selection. *Expert Systems with Applications*, 140, Article 112898. <https://doi.org/10.1016/j.eswa.2019.112898>
- Fix, E., & Hodges, J. L. (1951). Discriminatory analysis-nonparametric discrimination: Consistency properties. *Technical Report, California Univ Berkeley*. <https://doi.org/10.2307/1403797>
- Gao, Y., Zhou, Y., & Luo, Q. (2020). An Efficient Binary Equilibrium Optimizer Algorithm for Feature Selection. *IEEE Access*, 8, 140936–140963. <https://doi.org/10.1109/ACCESS.2020.3013617>
- Ghosh, M., Adhikary, S., Ghosh, K. K., Sardar, A., Begum, S., & Sarkar, R. (2019). Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods. *Medical & Biological Engineering & Computing*, 57, 159–176. <https://doi.org/10.1007/s11517-018-1874-4>
- Got, A., Moussaoui, A., & Zouache, D. (2021). Hybrid filter-wrapper feature selection using whale optimization algorithm: A multi-objective approach. *Expert Systems with Applications*, 183, Article 115312. <https://doi.org/10.1016/j.eswa.2021.115312>
- Gupta, S., Deep, K., & Mirjalili, S. (2020). An efficient equilibrium optimizer with mutation strategy for numerical optimization. *Applied Soft Computing*, 96, Article 106542. <https://doi.org/10.1016/j.asoc.2020.106542>
- Hammami, M., Bechikh, S., Hung, C., & Said, L. B. (2019). A Multi-objective hybrid filter wrapper evolutionary approach for feature selection. *Memetic Computing*, 11, 193–208. <https://doi.org/10.1007/s12293-018-0269-2>
- Jain, A., Nandakumar, K., & Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern recognition*, 38, 2270–2285. <https://doi.org/10.1016/j.patcog.2005.01.012>
- Keller, J. M., Gray, M. R., & Givens, J. A. (1985). A fuzzy k-nearest neighbor algorithm. *IEEE Transaction Systems, Man and Cybernetics*, 15(4), 580–585. <https://doi.org/10.1109/TSMC.1985.6313426>
- Kennedy, J., & Eberhart, R. C. (1997). A discrete binary version of the particle swarm algorithm. *IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, 5, 4104–4108. 10.1109/ICSMC.1997.637339.
- Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In D. Sleeman, P. Edwards (Eds.), *Proceedings of the Ninth International Conference on Machine Learning (ICML-92)*, Morgan Kaufmann, pp. 249–256. 10.1016/B978-1-55860-247-2.50037-1.
- Kumbure, M. M., Luukka, P., & Collan, M. (2020). A new fuzzy k-nearest neighbor classifier based on the Bonferroni mean. *Pattern Recognition Letters*, 140, 172–178. <https://doi.org/10.1016/j.patrec.2020.10.005>
- Li, S., Chen, H., Wang, M., Heidari, A. A., & Mirjalili, S. (2020). Slime Mould Algorithm: A new method for Stochastic Optimization. *Future Generation Computer Systems*, 111, 300–323. <https://doi.org/10.1016/j.future.2020.03.055>
- Luukka, P. (2007). Similarity classifier using similarity measure derived from Yu's norms in classification of medical data sets. *Computers in Biology and Medicine*, 37, 1133–1140. <https://doi.org/10.1016/j.combiomed.2006.10.005>
- Mafarja, M., & Mirjalili, S. (2018). Whale optimization approaches for wrapper feature selection. *Applied Soft Computing*, 62, 441–453. <https://doi.org/10.1016/j.asoc.2017.11.006>
- Mafarja, M., Aljarah, I., Heidari, A. A., Faris, H., Viger, P. F., Li, X., et al. (2018). Binary Dragonfly Optimization for Feature Selection using Time-Varying Transfer functions. *Knowledge-Based Systems*, 165, 185–204. <https://doi.org/10.1016/j.knosys.2018.08.003>
- Meenachi, L., & Ramakrishnan, S. (2021). Metaheuristic Search Based Feature Selection Methods for Classification of Cancer. *Pattern Recognition*, 119, Article 108079. <https://doi.org/10.1016/j.patcog.2021.108079>
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software*, 69, 46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
- Nadimi-Shahroki, M. H., Zamani, H., & Mirjalili, S. (2022). Enhanced whale optimization algorithm for medical feature selection: A COVID-19 case study. *Computers in Biology and Medicine*, 148, Article 105858. <https://doi.org/10.1016/j.combiomed.2022.105858>
- Ouadfel, S., & Elaziz, M. A. (2022). Efficient high-dimension feature selection based on enhanced equilibrium optimizer. *Expert Systems with Applications*, 187, Article 115882. <https://doi.org/10.1016/j.eswa.2021.115882>
- Robnik-Sikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*, 53, 23–69. <https://doi.org/10.1023/A:1025667309714>
- Sapre, S., & Mini, S. (2019). Opposition-based moth flame optimization with Cauchy mutation and evolutionary boundary constraint handling for global optimization. *Soft Computing*, 23, 6023–6041. <https://doi.org/10.1007/s00500-018-3586-y>
- Sayed, G. I., Khoriba, G., & Haggag, M. H. (2022). 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, 3137–3162. <https://doi.org/10.1007/s12652-021-03151-7>
- Saeys, Y., Abeel, T., & Van der Peer, Y. (2008). Robust feature selection using ensemble feature selection techniques, in: W. Daelemans, et al. (Eds.), Proc. European Conference on Machine Learning (ECML PKDD), LNCS 5212, 313–325. Doi: 10.1007/978-3-540-87481-2_21.

- Spoilaor, N., Cherman, E. A., Monard, M. C., & Lee, H. D. (2013). A comparison of multi-label feature selection methods using the problem transformation approach. *Electronic Notes in Theoretical Computer Science*, 292, 135–151. <https://doi.org/10.1016/j.entcs.2013.02.010>
- Sun, L., Kong, X., Xu, J., Xue, Z., Zhai, R., & Zhang, S. (2019). A Hybrid Gene Selection Method Based on ReliefF and Ant Colony Optimization Algorithm for Tumor Classification. *Scientific Reports*, 9, 8978. <https://doi.org/10.1038/s41598-019-45223-x>
- Tizhoosh, H.R. (2005). Opposition-based learning: A new scheme for machine intelligence. Proceedings of International Conference on Computational Intelligence for Modelling Control and Automation – CIMCA, Vienna, Austria, 1, 695-701. 10.1109/CIMCA.2005.1631345.
- Too, J., & Mirjalili, S. (2020). General Learning Equilibrium Optimizer: A New Feature Selection Method for Biological Data Classification. *Applied Artificial Intelligence*, 35 (3), 247–263. <https://doi.org/10.1080/08839514.2020.1861407>
- Tsai, C. F., & Hsiao, Y. C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems*, 50, 258–269. <https://doi.org/10.1016/j.dss.2010.08.028>
- Unler, A., Murat, A., & Chinnam, R. B. (2011). mr2psos: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification. *Information Sciences*, 181(20), 4625–4641. <https://doi.org/10.1016/j.ins.2010.05.037>
- Varzaneh, A. Z., Hossein, S., Mood, S. E., & Javidi, M. M. (2022). A new hybrid feature selection based on Improved Equilibrium Optimization. *Chemometrics and Intelligent Laboratory Systems*, 228, Article 104618. <https://doi.org/10.1016/j.chemolab.2022.104618>
- Vommi, A. M., & Battula, T. K. (2023). A binary Bi-phase mutation-based hybrid Equilibrium Optimizer for feature selection in medical datasets classification. *Computers and Electrical Engineering*, 105, Article 108553. <https://doi.org/10.1016/j.compeleceng.2022.108553>
- Vommi, V. B., & Vemula, R. (2018). A very optimistic method of minimization (VOMMI) for unconstrained problems. *Information Sciences*, 454, 255–274. <https://doi.org/10.1016/j.ins.2018.04.046>
- Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67–82. <https://doi.org/10.1109/4235.585893>
- Yao, Y. Y., Wong, S. K., & Butz, C. J. (1999). On information-theoretic measures of attribute importance. *Pacific Asia Conference on Knowledge Discovery and Data Mining*, 133–137. https://doi.org/10.1007/3-540-48912-6_18
- Yu, L., & Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research* 5, 1205–1224. 24. <http://dl.acm.org/citation.cfm?Id=1005332.1044700>.
- Zhang, G., Hou, J., Wang, J., Yan, C., & Luo, J. (2020). Feature Selection for Microarray Data Classification Using Hybrid Information Gain and a Modified Binary Krill Herd Algorithm. *Interdisciplinary Sciences: Computational Life Sciences*, 12, 288–301. <https://doi.org/10.1007/s12539-020-00372-w>