Journal of Engg. Research Vol. 11 No(2B) pp.71-83 DOI:10.36909/jer.16863

A Novel Feature Selection Algorithm Using Multi-Objective Improved Honey Badger Algorithm and Strength Pareto Evolutionary Algorithm-II

Anusha Papasani* and Nagaraju Devarakonda**

*,** *School of Computer Science and Engineering VIT-AP University, Vijayawada, India*

* *Corresponding Author: anoosha.papasani@gmail.com*

Submitted : 18-01-2022

Revised : 29-03-2022

Accepted : 03-04-2022

ABSTRACT

Feature selection is an important task in classification that removes redundant or irrelevant features from the dataset. Many researchers favor a multi-objective feature selection approach. However, these approaches fail to maintain high classification accuracy while removing redundancy in the features. In this work, a wrapper-based feature selection technique is proposed using a hybrid of the Multi-Objective Honey Badger Algorithm (MO-HBA) and the Strength Pareto Evolutionary Algorithm-II, called MOHBSP2, to balance classification accuracy and redundancy removal. Classification accuracy improvement and the removal of redundant features are considered the multi-objective optimization functions of the proposed multi-objective feature selection technique. The Levy flight algorithm is used to initialize the population and enhance the exploration and exploitation of MO-HBA. The regularized Extreme Learning Machine (ELM) is used to classify the selected features. To evaluate the performance of the proposed feature selection technique, 18 benchmark datasets are used and results are compared with the four well-known multi-objective feature selection techniques in terms of accuracy, hamming loss, ranking loss, mean value, standard deviation, feature length, and training time. The proposed approach achieved a maximum accuracy of 99% with the maximum value of selected features as 80. The minimum value of hamming loss, ranking loss, mean value, and standard deviation value achieved by the proposed approach are 0.0092, 0.0003, 0.018, and 0.001, respectively. The experimental results show that the proposed approach can improve classification accuracy and remove redundancy in large datasets.

Keywords: Strength Pareto Evolutionary Algorithm-II; Multi-objective Optimization; Wrapper Feature Selection; Levy Flight; Honey Badger Algorithm.

INTRODUCTION

The main challenge in data mining and machine learning is classification. Even with mostly relevant features, collecting and creating a dataset is a difficult task and the collected features are redundant and irrelevant to the classifiers (Hu et al., 2020). To obtain a more compact dataset, it is necessary to select the maximum number of relevant features, and feature selection is an effective dimensionality reduction technique to achieve this (Xue et al., 2021). The feature selection technique enables classifiers to obtain better interpretability, improves the classifier's generalization ability, and reduces over-fitting (Dash et al., 1997). Filters and wrappers are the two main categories of feature selection algorithms. The key features are selected by the filter methods using the intrinsic data characteristics. Based on roughest theory, distance, and information theory, the important features are selected by the filter methods (Labani et al., 2020). In the wrapper method, learning algorithms are used to evaluate the significance of selected features. Based on the various search strategies, the important features are selected by the wrapper methods (Xue et al., 2013). In classification performance, the wrapper method is better than the filter methods (Vijayanand et al., 2020) but wrapper methods were slower on large datasets than filter methods (Bermejo et al., 2014). For minimizing the size of feature subsets and improving classification accuracy, the wrapper feature selection can be termed a multi-objective optimization model from an optimization standpoint

(Li, et al., 2020). Due to their powerful searching ability, a lot of attention has been focused on metaheuristic optimization techniques (Nayyar et al., 2018; Zhang et al., 2018; Le et al., 2018). They have been broadly used in several real-world applications like path planning of unmanned aerial vehicles, flow-shop scheduling problems, and wireless sensor networks. In wrapper algorithms, the metaheuristic optimization algorithm plays an important role. Particle swarm optimization (PSO), bat algorithm (BA), whale optimization algorithm (WOA), genetic algorithm (GA), and grey wolf optimization (GWO) algorithm are the most widely used optimization algorithms; PSO (Banka et al., 2015; Xue et al., 2014) and GA (Eroglu et al. 2017; Hamdani et al., 2011) are widely used in many studies. Based on the number of objective functions, these evolutionary algorithms are further divided into multi-objective methods and single-objective methods (Han et al., 2015). When compared to single-objective methods, the multi-objective methods have more advantages (Fu et al., 2014). Though the generally used single-objective optimization algorithms such as PSO and GA achieve better global search performance results, their exploitation ability of identified regions is weak. Honey badger algorithm (HBA) (Hashim et al., 2022) is a recently proposed optimization algorithm with better exploration and exploitation abilities. The multi-objective methods such as Strength Pareto Evolutionary Algorithm-II (SPEA-II) (Zitzler et al., 2001), non-dominated sorting genetic algorithm II (NSGA-II) (Deb et al., 2002), and Pareto envelope-based selection algorithm (PESA) (Knowles et al., 1999) have shown better performance in handling many objective functions. Among them, SPEA-II is the popular and significant evolutionary algorithm that is the improved version of SPEA. It is the best multi-objective framework when compared with NSGA-II and PESA, particularly in high-dimensional spaces. Its feature selection is aimed at minimizing the error rate while removing redundant features. Thus, multi-objective optimization techniques are considered more in the feature selection approaches. The existing multi-objective optimization algorithms, such as SPEA-II, NSGA-II and PESA, produce non-dominated solutions and show less accuracy with a reduction in the feature size (Dong et al., 2020). Combining many optimization algorithms can improve classification accuracy while reducing the feature size. In this work, the combination of HBA with SPEA-II, named MOHBSP2, is proposed as the multi-objective optimization algorithm to select the features. The Levy flight algorithm (Liu et al., 2020) is used to generate the initial population of MOHBSP2. This method can extend the search space, improve the performance of the existing SPEA-II approach, and ensure search speed. The performance of the classifier will be affected due to the imbalanced dataset. To balance the dataset, the most-used effective sampling method is Synthetic Minority Over-sampling Technique (SMOTE) (Huang et al., 2020). In the proposed work, the preprocessing step uses SMOTE to balance the dataset. To calculate the accuracy of the selected features, a classifier is used by the proposed wrapper technique. There are many efficient and accurate classification algorithms such as Naive Bayes, Decision Tree, SVM, Artificial Neural Network, and Convolutional Neural Network used by researchers to classify the selected features. Among them, Extreme Learning Machine (ELM) works well and quickly. In this study, Regularized Extreme Learning Machine (RELM) (Wang et al., 2020) is employed to evaluate the performance of the proposed approach. Compared to conventional classifiers, RELM gives satisfactory results in multi-label classification of large datasets. The results of the proposed method are compared to popular feature selection algorithms such as the multi-objective binary genetic algorithm integrating an adaptive operator selection mechanism (MOBGA-AOS) (Xue et al., 2021), multi-objective PSO (MO-PSO) (Paul et al., 2021), the multi-objective binary cuckoo optimization algorithm, and the non-dominated sorting genetic algorithms NSGA III (BCNSG3) (Usman et al., 2020) and modified whale optimization algorithm (MWOA) (Vijayanand et al., 2020).

The main contributions of the proposed MOHBSP2-based feature selection technique can be summarized as follows. The HBA is used for the multi-objective feature selection algorithm. The SPEA-II algorithm is combined with the MO-HBA to create MOHBSP2 to improve the performance of the existing HBA. To improve the exploration and exploitation ability of the existing SPEA-II, the Levy flight algorithm is used to initialize the population. The effectiveness of the proposed feature selection approach is evaluated by using the selected features in classification. The selected features are used to improve the classification performance of existing RELM. The performance of the proposed MOHBSP2-based feature selection algorithm is measured in terms of accuracy, number of selected features, computation time, hamming loss, ranking loss mean value, and standard deviation, to show the effectiveness of the proposed feature selection algorithm. The performance of the proposed approach is compared with four other well-known multi-objective evolutionary algorithms, MWOA, BCNSG3, MOBGA-AOS, and MO-PSO, on 18 datasets.

RELATED WORKS

Most feature selection methods are proposed as single-objective problems in the fitness function. There are few multi-objective feature selection methods. Zhang et al. (2020) introduced a binary differential evolution-based self-learning method. They used three operators, purifying search one-bit operator, probability difference binary mutation, and non-dominated sorting operator, based on crowding distance. This approach was evaluated using 20 standard datasets. This approach showed better performance and consumed less running time. With two objective functions, a wrapper-based multi-objective feature selection technique based on NSGA-II was proposed by Kozodoi et al. (2019). In the objective functions, feature subset reduction and expected profit maximizing were considered. They utilized 10 retail credit scoring datasets. This suggested approach showed better performance. Using breeding operators and NSGA-II, a wrapper-based multi-objective feature selection approach was proposed by Gonzalez et al. (2019). This method maintained stability with the feature ranking process. They applied four classifiers to evaluate the effectiveness of the proposed approach. Sharma and Rani (2019) introduced a feature selection approach that combined Salp Swarm Algorithm and multi-objective spotted hyena optimizer. There are two stages in this method. The irrelevant features are eliminated using a filter approach in the first stage. Then the most optimal features are explored using their hybrid method in the second stage. Kiziloz et al. (2018) proposed teaching a learning-based multi-objective feature selection approach. In this approach, three multi-objective TLBO algorithms, such as non-dominated selection, scalar transformation, and minimum distance are proposed. They utilized three classifiers, SVM, ELM, and LR to evaluate the performance of the proposed approach. Based on the frequency in the set of documents, the features are ranked using a PSO-based multi-objective approach proposed by Amoozegar and Minaei-Bidgoli (2018). Then, the particles are guided and a set of archives are enhanced using these levels. This proposed approach is compared with multi-objective GA and three variants of PSO. Hancer et al. (2018) introduced a hybrid of non-dominated sorting genetic operators and multi-objective artificial bee colony for feature selection. This approach used the k-nearest neighbor (k-NN) classifier to evaluate the selected feature subset. Dashtban et al. (2018) proposed a multi-objective approach for the classification of microarray results and gene selection. They used SVM, NBY, k-NN, and DT as the classifiers to evaluate the effectiveness of the proposed approach. For gene selection, Lai (2018) introduced a hybrid filter-wrapper-based multi-objective optimization approach. They selected the finest genes using an aggregate filter technique. In a multi-objective formula, Nayak et al. (2020) proposed an elitism-based differential evolution approach for filter-based feature selection. This approach removes the undesired and redundant features of the processed dataset by considering the nonlinear and linear dependencies. Usman et al. (2020) proposed a filter-based feature selection algorithm using the multi-objective binary cuckoo optimization algorithm and NSGA-III. They proposed four multi-objective filter-based feature selection techniques by utilizing entropy-based gain ratio and mutual information. Though their method derives the best feature subsets, a performance comparison with other evolutionary algorithms is not provided to show their effectiveness. Xue et al. (2021) introduced a multi-objective feature selection technique for classification using a binary genetic algorithm and an adaptive crossover operator. In their approach, different search characteristics are used with five crossover operators. According to the performance of the evolution process, a probability is assigned to each of the crossover operators. A Relative Discriminative Criterion (MORDC) algorithm with a multi-objective function is suggested by Labani et al. (2020) for text feature selection. They took the relevance of the text features to the target class as the first objective and selected the evaluation of the connection between the features as the second objective. They computed the redundancy and relevancy of the features using Pearson correlation and RDC measures. Gao et al. (2021) introduced a multi-objective optimization algorithm to select the features with the hybrid cat swarm optimization algorithm (HCSO). In this approach, they combined the inherent, competitive, and guided characteristics with the original CSO. They evolved global worst and best solutions during the HCSO execution. Though this approach gives better results, the comparison results with other existing optimization algorithms are not provided. Rostami et al. (2020) proposed an improved multi-objective PSO-based feature selection approach for medical datasets. They used a mutation operator to enhance the quality of a generated feature subset and the diversity of BPSO. Moreover, the convergence of the PSO algorithm improved by introducing a new node centrality-based approach to optimize the initial population of PSO. Abdollahzadeh and Gharehchopogh (2021) proposed Hybrid of Harris Hawks Optimization and Fruitfly Optimization Algorithm. In this multi-objective approach, the error rate was considered as one objective function and the number of features was considered as another objective function. The performance of these existing approaches showed less accuracy with the reduction in the feature size. Though the existing multi-objective optimization algorithms, and specifically SPEA-II, showed better performance than other approaches, SPEA-II produces non-dominated solutions and shows less accuracy with the reduction in the feature size. From this literature review, it can be known that the combination of optimization approaches produces better results in feature selection. A hybrid of various optimization algorithms is proposed in this paper to improve the performance of existing approaches such as SPEA-II and HBA.

PRELIMINARIES

Honey Badger Algorithm

HBA is inspired by the Honey Badger's intelligent foraging behavior (Hashim et al., 2022). Algorithm 1 explains the pseudo-code for HBA. Similar to the exploitation and exploration phases in HBA, the Honey Badger's search behavior comprises honey-finding and digging techniques. Representation of the population (P) of the candidate solution is given in Equation (1).

$$P = \begin{bmatrix} a_{11} & a_{12} & a_{13} \dots & a_{1D} \\ a_{21} & a_{22} & a_{23} \dots & a_{2D} \\ & \dots\dots\dots\dots\dots\dots. & & \\ a_{n1} & a_{n2} & a_{n3} \dots & a_{nD} \end{bmatrix} \tag{1}$$

The ith position of HBA is represented by using Equation (2).

$$a_i = [a_i^1, a_i^2, \dots, a_i^D] \tag{2}$$

Algorithm 1. Pseudocode for HBA

Initialize parameters t_{max}, N, β, C
Generate population P_{new} with random position between 0 and 1.
Evaluate fitness of each position x_i using objective function f_i
The best position x_{prey} and the fitness are assigned to f_{prey}
while t≤ t_{max} **do**
    Update the decreasing factor ∝
    for i=1 to N **do**
        Calculate the intensity I_i
        if r<0.5 **then**
            Update the position x_{new}
        else
            Update the position x_{new}
        end if
        Evaluate new position and assign to f_{new}
        if $f_{new} \leq f_i$ **then**
                Set x_i=x_{new} and $f_i = f_{new}$
        end if
        if $f_{new} \leq f_{prey}$ **then**
                Set x_{prey}=x_{new} and $f_{prey} = f_{new}$
        end if
    end for
end while Stop criteria satisfied
Return x_{prey}

Strength Pareto Evolutionary Algorithm-II (SPEA-II)

SPEA-II is one of the evolutionary multiple-objective algorithms (Zitzler et al., 2001). For the multiple-objective optimization problems, the original genetic algorithm is extended as SPEA-II. The objective of this algorithm is to preserve and identify a set of Pareto optimal solutions. The Pareto optimal set is all the Pareto optimal solutions. In the objective space, the best non-dominated solutions made the Pareto optimal set. For each solution, the two main parameters are considered. Algorithm 2 details the pseudocode for SPEA-II.

Algorithm 2. Pseudocode for SPEA-II

Population (P) initialization
Creation of empty external set E
    for i=1 **to** no of Generation
        Calculation of fitness for each feature in A and P
        Add Non-dominated features from A and P
        if capacity of A exceeds **then**
            By Truncation Operator Remove features from A
        if capacity of A not exceeds **then**
            To fill E using dominated features in P
        Binary tournament selection
        Mutation and crossover
    end for

PROPOSED METHOD

The proposed approach selects the features using the MOHBSP2 approach and the selected features are classified using regulated ELM. The existing multi-objective approaches select the non-dominated parameters and the accuracy decreases with the reduction in the feature size. Improving the accuracy of the classification process during the reduction in feature size is a complex task. The combination of more optimization algorithms can solve this issue. In this paper, the performance of the existing SPEA-II approach is improved by combining the HBA and Levy flight (Balakrishnan et al., 2021; Ewees et al., 2022) approaches. Thus, the proposed approach can produce better accuracy despite the reduction in the feature size.

Dataset

Eighteen datasets from the UCI Machine Learning Repository (Asuncion and Newman, 2007) are used to evaluate the performance of the proposed feature selection approach. The datasets used to evaluate the proposed approach are WDBC, Zoo, Lymphography, Ionosphere, Credit, Heart, Dermatology, Sonar, Spect, Parkinson, Indian Pima, Scene, Kc1, Audiology, Tic-tac-toe, Waveform, Glass, and Wine and were collected from various domains such as medical and computer.

MOHBSP2-based Feature Selection

According to the basic HBA, a modified multi-objective method called MOHBSP2 is proposed in this paper. The exploitation ability can be improved by performing the local search with Levy Flight.

Initialization

In this phase, the parameters like maximum iteration count, archive size ($\overline{N}$), and feature size are initialized. At t = 0, the initial population(P_0) is generated. The size of the population and the respective positions are initialized based on Equation (3),

$$x_i = lb_i + r_1 \times (ub_i - lb_i) \tag{3}$$

where r_1 represents a random number between 1 and 0. The ith position of HBA, referring to a feature in a population P_0, is represented by x_i. The lower and upper bounds of the search domain are represented by lb_i and ub_i. Then an empty archive $\overline{(P_t)}$ is initialized by the Levy flight algorithm using Equation (4),

$$\overline{P_t} = x_i + \propto \oplus \, Levy(\lambda) \tag{4}$$

where a parameter of random step size is represented by $\propto$, and the distribution parameter of Levy flight is represented by λ. The entrywise multiplication is represented by $\oplus$. The *i*th solution is represented by x_i.

Fitness Assignment

The fitness values are calculated for both the population P_t and $\overline{P_t}$. A strength value *S*(*i*) is allocated by each individual *i* in the population P_t and archive $\overline{P_t}$. It denotes the dominant number of features as expressed in Equation (5),

$$S(i) = |\{j|j \in P_t + \overline{P_t} \wedge i \ \phi \ j)| \tag{5}$$

where the cardinality of a set is represented by $|\cdot|$. The Pareto dominance relation is represented by ϕ. The raw fitness *Rw*(*i*) of a feature is calculated based on the *S* values as given in Equation (6).

$$Rw(i) = \sum_{j \in P_t + \overline{P_t}, j \phi i} S(j) \tag{6}$$

Corresponding to *i*, the density D(*i*) is expressed as given in Equation (7),

$$D(i) = \frac{1}{\sigma_i^k + 2} \tag{7}$$

where the kth element for each individual *i* is represented by σ_i^k. For the raw fitness value *Rw*(*i*) the addition of D(*i*) produces its fitness *Ft*(*i*) as given in Equation (8).

$$Ft(i) = Rw(i) + D(i) \tag{8}$$

Fitness Function

Finding a set of optimal features with a small solution size and high classification accuracy is the aim of the multi-objective feature selection technique. Instead of maximizing the classification accuracy, minimizing the classification errors is taken as the first fitness function. The second fitness function considers the size of the solution. To evaluate the solutions, k-NN is used as the classifier. The k-NN uses n-fold cross-validation. The first fitness function can be calculated using Equation (9),

$$\min(f_1) = \left(\frac{1}{n}\sum_{l=1}^{n}\frac{N_{Error}}{N_{All}}\right) \times 100\% \tag{9}$$

where the feature is represented by X, the number of all the instances is represented by N_{All}, and the number of wrongly predicted instances is represented by N_{Error}. The second fitness function can be calculated using Equation (10),

$$\min(f_2) = \sum_{i=1}^{D} x_i \tag{10}$$

where the *i*th value in the X feature is represented by x_i. The number of original features is represented by *D*.

Environmental Selection

All the non-dominated features from both the populations $\overline{P_t}$ and P_t that have fitness value lower than 1 are copied to $\overline{P_{t+1}}$. It can be expressed using Equation (11),

$$\overline{P_{t+1}} = \{i|i \in P_t + \overline{P_t} \wedge F(i) < 1 \tag{11}$$

The environmental selection is completed if the front of the non-dominated fits exactly into the dataset $|\overline{P_{t+1}}| = \overline{N}$. If the size of the dataset is too small ($\overline{P_{t+1}} | < \overline{N}$), then $\overline{P_{t+1}}$ fills with dominated features from $\overline{P_t}$ and P_t. The truncation operator is used to reduce $\overline{P_{t+1}}$ if the size of the dataset is too large ($\overline{P_{t+1}} | > \overline{N}$).

Position Update using HBA

Multi-objective HBA is utilized to update the positions of the population $\overline{P_{t+1}}$. The positions of $\overline{P_{t+1}}$ are updated using Equation (12) if the value of r is less than 0.5; otherwise, Equation (13) can be used.

$$\overline{P_{t+1}} = x_{prey} + F \times \beta \times I \times x_{prey} + F \times r_3 \times \alpha \times d_i \times |\cos(2\pi r_4) \times [1 - \cos(2\pi r_5]| \quad (12)$$

$$\overline{P_{t+1}} = x_{prey} + F \times r_7 \times \propto \times d_i \quad (13)$$

where the global best position is represented by x_{prey}. The honey badger's ability to search for food is represented by β. The distance between prey and the i^{th} honey badger is represented by d_i and it can be calculated using Equation (14). The random numbers between 0 and 1 are represented by r_3, r_4, r_5, and r_7. The search direction is represented by F and it can be calculated using Equation (15). The decreasing factor α is updated using Equation (16),

$$d_i = x_{prey} - x_i \quad (14)$$

where the global best position is represented by x_{prey} and the i^{th} honey badger position is represented by x_i.

$$\text{F} = \begin{cases} 1 & if\ r_6 \leq 0.5 \\ -1 & else \end{cases} \quad (15)$$

The random number between 0 and 1 is represented by r_6,

$$\alpha = C \times \exp\left(\frac{-t}{t_{max}}\right) \quad (16)$$

where t_{max} represents the maximum count of iterations. The constant C has a default value of 2.

Feature selection using SPEA-II

If the maximum iteration is reached then the non-dominated features are represented by A with the set of decision vectors. To fill the mating pool, with the replacement of $\overline{P_{t+1}}$, the binary tournament selection is performed. For the mating pool, the mutation and crossover operators are applied and the resulting population is $\overline{P_{t+1}}$. Then the generation counter is incremented to $t = t + 1$. Then the fitness is calculated.

Computational Complexity

The computational complexity of the proposed approach is O(mn^2), where the population size is represented by n and the number of selected features is represented by m.

System Model

The proposed feature selection approach is established with the following steps. Step 1: The collected datasets were preprocessed using the scaling function. Step 2: The high dimensionality of the preprocessed dataset was reduced using the proposed MOHBSP2-based feature selection approach. Step 3: The selected features were classified using RELM. The proposed approach showed better accuracy with a reduced feature size and less computational time. At first, the parameters used in the proposed approach were initialized. The population was initialized with random numbers between 0 and 1. Then the archive set was generated by using the Levy flight algorithm. The fitness value was evaluated for each population. Based on the fitness value the positions were updated using Equations (12) and (13). Thus, the position update was performed with HBA. Then the selection, crossover, and mutation processes were performed based on SPEA-II. Based on these values the optimal features were selected from the dataset. The selected features were used in the RELM-based classification.

Algorithm 3. Pseudo code for proposed feature selection algorithm

Initialize parameters t_{max}, N, β, C, $\overline{N}$
Generate new population P_0 with random position between 0 and 1
Generate archive set $\overline{P_0}$ (t=0) based on Levy flight using Equation (4) **// Levy flight**
while t≤ t_{max} **do**
Update the decreasing factor $\propto$ using Equation (16)
for i=1 to Number of populations **do**
Calculate fitness of each feature in P_t and $\overline{P_t}$
$\overline{P_{t+1}}$ = Copy all non-dominated feature from $\overline{P_t}$ and $\overline{P_t}$ to $\overline{P_{t+1}}$
if r<0.5 **then**
Update the position $\overline{P_{t+1}}$ using Equation (12) **// HBA**
else
Update the position $\overline{P_{t+1}}$ using Equation (13)
end if
end for
Perform selection **//SPEA-II**
Apply crossover and mutation
end while stop criteria satisfied
Calculate and save feature subset's error rate (solutions)
Using non-dominated sorting ranking the population
Return the selected features
end

EXPERIMENTAL RESULTS

In Figure 1, the non-dominated solutions are plotted for the performance evaluation of the proposed approach on feature subset searching. The proposed approach result is from the Zoo dataset and compared to existing approaches. The size of the feature subset is represented by the horizontal axis on the graph and the classification error is represented by the vertical axis on the graph. The Pareto front results show that the proposed MOHBSP2 approach can achieve fewer classification errors while ensuring a smaller solution size for all 18 datasets.

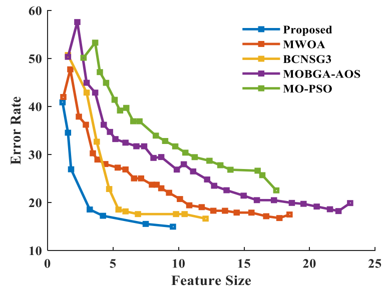
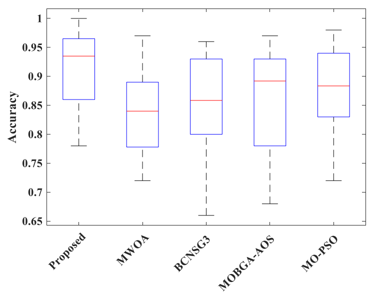
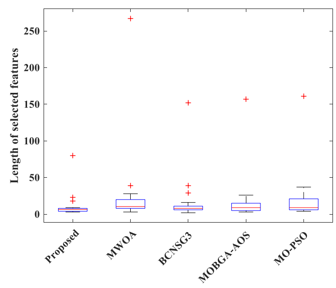


Figure 1. Pareto front analysis

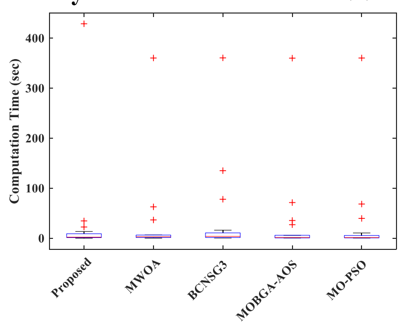
In Figure 2(a) the accuracy of the proposed approach is compared to existing approaches in a box plot analysis for all the datasets. For the proposed approach, the datasets Ionosphere, Parkinson, and Wine achieved 99% accuracy. The upper quartile in the box plot for the proposed approach is 99%. For the existing approaches MWOA, BCNSG3, MOBGA-AOS, and MO-PSO the highest values in the upper quartile are 97%, 96%, 97%, and 98%, respectively. Therefore, the proposed approach has the highest value in the upper quartile compared to the existing approaches in the box plot. The median line of the box for the proposed approach is 93%, higher than the existing approaches. There are no potential outliers in the box plot, as all the approaches have better-than-average performance. Similarly, the selected feature length is compared to existing approaches in the box plot analysis in Figure 2(b). The highest length of the selected features in the proposed approach is 80. For the existing approaches MWOA, BCNSG3, MOBGA-AOS, and MO-PSO the highest value of the selected features are 267, 152, 157, and 161, respectively. Due to the large size of the dataset, there are some potential outliers in the box plot in all the approaches. Thus, the proposed approach selects fewer features than the existing approaches. The computation time taken by the proposed approach is compared to existing approaches in the box plot analysis in Figure 2(c). The highest value in computation time taken by the proposed approach is 378.99 seconds. For the existing approaches MWOA, BCNSG3, MOBGA-AOS, and MO-PSO the highest value in computation time is 350.58 seconds, 328.82 seconds, 337.23 seconds, and 365.59 seconds, respectively. The computation time taken by the proposed approach is 13.31 seconds higher than the existing approaches. The difference in the computation time is in seconds. From these results it can be determined that the proposed feature selection approach can maintain higher accuracy with the reduction in feature size.



(a) Accuracy



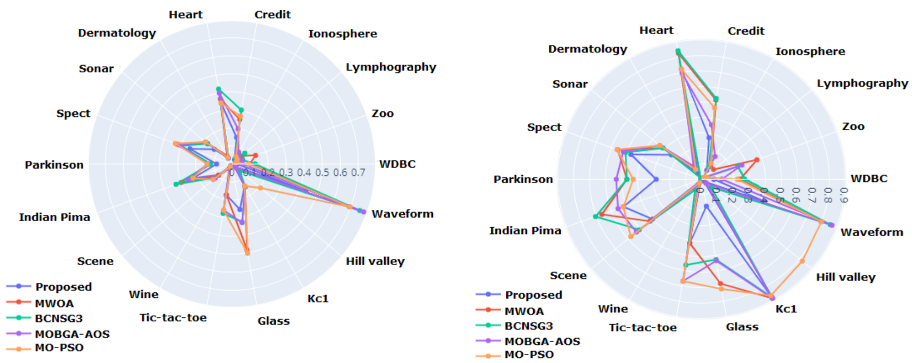
(b) Feature length



(c) Computation time

Figure 2. Box plot accuracy analysis

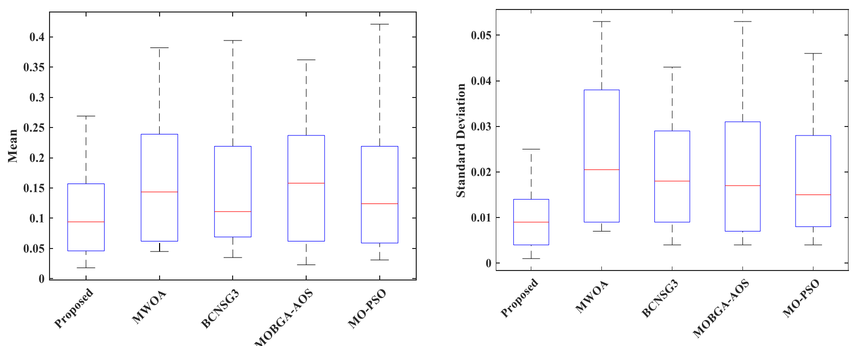
In Figure 3(a) and 3(b), the hamming loss and ranking loss results of the proposed feature selection approach are compared with four other well-known feature selection approaches, MWOA, BCNSG3, MOBGA-AOS, and MO-PSO, using a radar chart. The minimum value of hamming loss in the proposed approach is 0.0092. For the existing approaches MWOA, BCNSG3, MOBGA-AOS, and MO-PSO the lowest values in hamming loss in the box plot are 0.0294, 0.0198, 0.0268, and 0.0153, respectively. The minimum value in ranking loss for the proposed approach is 0.0003. For the existing approaches MWOA, BCNSG3, MOBGA-AOS, and MO-PSO the lowest ranking losses are 0.0029, 0.0020, 0.0027, and 0.0015, respectively. To improve the performance of the feature selection approach, the hamming loss and ranking loss values should be less. The radar charts show that the proposed feature selection approach has less hamming loss and ranking loss when compared to other techniques.



(a) Hamming Loss **(b) Ranking Loss**

Figure 3. Radar chart for Hamming Loss and Ranking Loss

The performance of the proposed feature selection algorithm was evaluated using the IGD metric in terms of standard deviation and mean value for all 18 datasets. Figure 4 shows a box plot analysis of mean value and standard deviation. The results of the proposed MOHBSP2 method were compared to other multi-objective feature selection algorithms such as MOBGA-AOS, MO-PSO, BCNSG3, and MWOA. These results show that the proposed feature selection approach has less standard deviation and mean values for all 18 datasets when compared to the other algorithms. The performance of the proposed MOHBSP2 was compared to the other existing multi-objective algorithms such as MWOA, BCNSG3, MOBGA-AOS, and MO-PSO for all 18 datasets. The proposed approach has the highest accuracy value when compared to the other four algorithms for all 18 datasets. Moreover, the proposed approach selects fewer features than other algorithms. Thus, the proposed approach minimizes classification errors while minimizing the number of selected features. However, the training time taken by the proposed approach is higher than the other approaches. From these results, it can be concluded that the proposed multi-objective feature selection technique can provide better results than other conventional multi-objective feature selection techniques.



(a) Mean value **(b) Standard deviation value**

Figure 4. Box plot analysis for mean value and standard deviation value

CONCLUSION

A novel multi-objective wrapper feature selection technique is proposed in this paper to select fewer features with fewer classification errors. A hybrid of the Multi-Objective Honey Badger Algorithm and the Strength Pareto Evolutionary Algorithm-II combined with the Levy flight algorithm, called MOHBSP2, is the proposed feature selection algorithm. The effectiveness of the proposed approach was evaluated using 18 benchmark datasets and compared with four well-known multi-objective feature selection techniques. RELM was used to evaluate the selected features using a proposed feature selection approach. The classification results showed that the proposed feature selection technique outperformed the other feature selection approaches with fewer selected features and fewer classification errors. The proposed approach achieved 99% accuracy with a maximum length of selected features of 80. The minimum value of hamming loss, ranking loss, mean value, and standard deviation value achieved by the proposed approach were 0.0092, 0.0003, 0.018, and 0.001, respectively. Though the proposed approach shows better performance in terms of accuracy, ranking loss, and hamming loss, it is slightly slower than other approaches. Reducing the computation complexity of the proposed multi-objective feature selection algorithm will be investigated in future work.

REFERENCES

Abdollahzadeh, B., & Gharehchopogh, F. S. 2021. A multi-objective optimization algorithm for feature selection problems. Engineering with Computers 1-19.

Amoozegar, M., & Minaei-Bidgoli, B. 2018. Optimizing multi-objective PSO based feature selection method using a feature elitism mechanism. Expert Systems with Applications 113: 499-514.

Asuncion, A., & Newman, D. 2007. UCI machine learning repository.

Balakrishnan, K., Dhanalakshmi, R. & Khaire, U.M. 2021. Improved Salp swarm algorithm based on the Levy flight for feature selection. The Journal of Supercomputing 77(11): 12399-12419.

Banka, H., & Dara, S. 2015. A Hamming distance based binary particle swarm optimization (HDBPSO) algorithm for high dimensional feature selection, classification and validation. Pattern Recognition Letters 52: 94-100.

Bermejo, P., Gámez, J. A., & Puerta, J. M. 2014. Speeding up incremental wrapper feature subset selection with Naive Bayes classifier. Knowledge-Based Systems 55: 140-147.

Dash, M., & Liu, H. 1997. Feature selection for classification. Intelligent Data Analysis 1:131–156.

Dashtban, M., Balafar, M., & Suravajhala, P. 2018. Gene selection for tumor classification using a novel bio-inspired multi-objective approach. Genomics 110(1): 10-17.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T. 2002. A fast and elitist multi objective genetic algorithm: NSGA-II. IEEE transactions on evolutionary computation 6(2): 182-197.

Dong, H., Sun, J., Sun, X., & Ding, R. 2020. A many-objective feature selection for multi-label classification. Knowledge-Based Systems 208.

Eroglu, D. Y., & Kilic, K. 2017. A novel Hybrid Genetic Local Search Algorithm for feature selection and weighting with an application in strategic decision making in innovation management. Information Sciences, 405: 18-32.

Ewees, A. A., Mostafa, R. R., Ghoniem, R. M., & Gaheen, M. A. 2022. Improved seagull optimization algorithm using Lévy flight and mutation operator for feature selection. Neural Computing and Applications 1-36.

Fu, W., Xue, B., & Zhang, M. 2014. Multi-objective feature selection in classification: A differential evolution approach. In Asia-Pacific Conference on Simulated Evolution and Learning. Springer, Cham.

Gao, X. Z., Nalluri, M. S. R., Kannan, K., & Sinharoy, D. 2021. Multi-objective optimization of feature selection using hybrid cat swarm optimization. Science China Technological Sciences 64(3): 508-520.

Gonzalez, J., Ortega, J., Damas, M., Martín-Smith, P., & Gan, J. Q. 2019. A new multi-objective wrapper method for feature selection–Accuracy and stability analysis for BCI. Neurocomputing 333: 407-418.

Hamdani, T. M., Won, J. M., Alimi, A. M., & Karray, F.2011. Hierarchical genetic algorithm with new evaluation function and bi-coded representation for the selection of features considering their confidence rate. Applied Soft Computing 11(2): 2501-2509.

Han, M., & Ren, W. 2015. Global mutual information-based feature selection approach using single-objective and multi-objective optimization. Neurocomputing 168: 47-54.

Hancer, E., Xue, B., Zhang, M., Karaboga, D., & Akay, B. 2018. Pareto front feature selection based on artificial bee colony optimization. Information Sciences 422: 462-479.

Hashim, F. A., Houssein, E. H., Hussain, K., Mabrouk, M. S., & Al-Atabany, W. 2022. Honey Badger Algorithm: New metaheuristic algorithm for solving optimization problems. Mathematics and Computers in Simulation 192:84-110.

Hu, P., Pan, J. S., & Chu, S. C. 2020. Improved Binary Grey Wolf Optimizer and Its application for feature selection. Knowledge-Based Systems 195.

Huang, L., Zhang, H., Wu, C. Q., & Li, Z. 2020. An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset. Computer Networks 177.

Kiziloz, H. E., Deniz, A., Dokeroglu, T., & Cosar, A. 2018. Novel multiobjective TLBO algorithms for the feature subset selection problem. Neurocomputing 306: 94-107.

Knowles, J., & Corne, D. 1999. The Pareto archived evolution strategy: a new baseline algorithm for Pareto multiobjective optimisation. Proceedings of the 1999 Congress on Evolutionary Computation, 98–105.

Kozodoi, N., Lessmann, S., Papakonstantinou, K., Gatsoulis, Y., & Baesens, B. 2019. A multi-objective approach for profit-driven feature selection in credit scoring. Decision support systems 120: 106-117.

Labani, M., Moradi, P., & Jalili, M. 2020. A multi-objective genetic algorithm for text feature selection using the relative discriminative criterion. Expert Systems with Applications 149.

Lai, C. M. 2018. Multi-objective simplified swarm optimization with weighting scheme for gene selection. Applied Soft Computing 65: 58-68.

Le, D. N., Nayyar, A., & Nguyen, N. G. 2018. Advances in swarm intelligence for optimizing problems in computer science. CRC press.

Li, A. D., Xue, B., & Zhang, M. 2020. Multi-objective feature selection using hybridization of a genetic algorithm and direct multisearch for key quality characteristic selection. Information Sciences 523: 245-265.

Liu, Y., & Cao, B. 2020. A novel ant colony optimization algorithm with Levy flight. IEEE Access 8: 67205-67213.

Nayak, S. K., Rout, P. K., Jagadev, A. K., & Swarnkar, T. 2020. Elitism based multi-objective differential evolution for feature selection: A filter approach with an efficient redundancy measure. Journal of King Saud University-Computer and Information Sciences 32(2): 174-187.

Nayyar, A., & Nguyen, N. G. 2018. Introduction to swarm intelligence. Advances in swarm intelligence for optimizing problems in computer science 53-78.

Paul, D., Jain, A., Saha, S., & Mathew, J. 2021. Multi-objective PSO based online feature selection for multi-label classification. Knowledge-Based Systems 222.

Rostami, M., Forouzandeh, S., Berahmand, K., & Soltani, M. 2020. Integration of multi-objective PSO based feature selection and node centrality for medical datasets. Genomics 112(6): 4370-4384.

Sharma, A., & Rani, R. 2019. C-HMOSHSSA: Gene selection for cancer classification using multi-objective meta-heuristic and machine learning methods. Computer methods and programs in biomedicine 178: 219-235.

Usman, A. M., Yusof, U. K., & Naim, S. 2020. Filter-based multi-objective feature selection using NSGA III and cuckoo optimization algorithm. IEEE Access 8: 76333-76356.

Vijayanand, R. & Devaraj, D. 2020. A novel feature selection method using whale optimization algorithm and genetic operators for intrusion detection system in wireless mesh network. IEEE Access 8: 56847-56854.

Wang, D., Wang, P., Yuan, Y., Wang, P., & Shi, J. 2020. A fast conformal predictive system with regularized extreme learning machine. Neural Networks 126: 347-361.

Xue, B., Zhang, M., & Browne, W. N. 2013. Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach. IEEE Transactions on Cybernetics 43: 1656–1671.

Xue, B., Zhang, M., & Browne, W. N. 2014. Particle swarm optimization for feature selection in classification: Novel initialisation and updating mechanisms. Applied soft computing 18: 261-276.

Xue, Y., Zhu, H., Liang, J., & Słowik, A. 2021. Adaptive crossover operator based multi-objective binary genetic algorithm for feature selection in classification. Knowledge-Based Systems 227.

Zhang, X., Zhang, Q., Chen, M., Sun, Y., Qin, X., & Li, H. 2018. A two-stage feature selection and intelligent fault diagnosis method for rotating machinery using hybrid filter and wrapper method. Neurocomputing 275: 2426-2439.

Zhang, Y., Gong, D. W., Gao, X. Z., Tian, T., & Sun, X. Y. 2020. Binary differential evolution with self-learning for multi-objective feature selection. Information Sciences 507: 67-85.

Zitzler, E., Laumanns, M., & Thiele, L. 2001. SPEA2: Improving the strength pareto evolutionary algorithm, EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems. Athens. Greece.