



Multi-objective Optimization Based Feature Selection Using Correlation

Rajib Das¹, Rahul Nath¹, Amit K. Shukla²(✉), and Pranab K. Muhuri¹

¹ Department of Computer Science, South Asian University, New Delhi, India
rajiblxp@gmail.com, rahul.nath@outlook.com,
pranabmuhuri@cs.sau.ac.in

² Faculty of Information Technology, University of Jyväskylä, Box 35 (Agora),
40014 Jyväskylä, Finland
amit.k.shukla@jyu.fi

Abstract. The optimal feature selection (FS) problem is widely targeted in the field of machine learning (ML). There are several ways to select the best features when the dataset dimension is small. However, when the dataset and number of features tend to increase, the solution becomes unrealistic as we need to evaluate every subset performance with the model. Various existing heuristics are partially useful as they portray premature convergence and exponential or high computational complexity. To solve this issue, evolutionary approaches-based FS has been extensively used in obtaining the optimal subset of features while maintaining the accuracy of the model. This paper proposes an efficient evolutionary-based multi-objective feature selection approach with a correlation coefficient filter method called Multi-Objective Optimization based Feature Selection (MOOFS). We introduce a two-stage process to select the best optimal features. In the first stage, a subset of features is randomly selected, and then a novel mutual correlation coefficient technique is used to get the important and relevant subset of features. The proposed MOOFS is experimented on several datasets and compared with the classical approach to demonstrate its efficiency.

Keywords: Feature selection · Correlation coefficient · Multi-objective optimization · NSGA-II

1 Introduction

In the present technology-oriented world, huge number of data are generated intentionally or unintentionally which is stored with some characteristics, mostly referred as features. All the features in the dataset may not be of importance due to dependency implying features dependence on each other and redundancy indicating unwanted and less important features. The process of selecting most optimal subset of feature is commonly termed as feature selection (FS), which is often considered one of the most critical and challenging tasks in machine learning. The reduced feature set improve the computational complexity and further helps in additional analysis. FS reduces feature space ($m \times n$) from very large space to a smaller ($n \times d$) space where $d < m$. Mathematically,

a feature selection problem can be defined as follows: Suppose a dataset D contains n number of features, and the objective is to select the optimal subset of features from D which are relevant. D is represented as follows:

$$D = \{f_1, f_2, f_3, \dots, f_n\} \quad (1)$$

where $f_1, f_2, f_3, \dots, f_n$ represents the features of any dataset. In FS, we extract a subset $S = \{f_1, f_2, f_3, \dots, f_d\}$, where, $d < n$.

In general, FS methods are classified into three categories namely filter, wrapper and embedded (hybrid) approaches. Filter approaches use general characteristics of the data to select features and are independent of learning algorithms. Wrapper methods always include a learning algorithm and according to its performance (increase or decrease), features are selected. The wrapper methods have high computational cost but provide more accurate results. On the other hand, filter approaches have low computational cost with less reliability. Lastly, the embedded methods include of filter and wrapper approaches both where FS is a part of the training process that is held with a learning algorithm.

According to set theory, if any dataset has n number of features, then 2^n number of subsets is possible, and our task is to pick best subset by which machine learning model can give best accuracy. To select best subset, we need to evaluate every subset performance with the model, which is unrealistic, when n increase to a huge number. Various existing heuristics are partially useful as they portray premature convergence, exponential or high computational complexity. To solve this issue, evolutionary or multi-objective optimization (MOO) based FS approaches have been extensively used in obtaining the optimal subset of features while preserving the accuracy of the model [1, 5]. These approaches tend to be efficient, effective, and reliable methods. In practice, the FS based on evolutionary, or MOO approaches falls under the wrapper method.

Mathematically, a multi-objective optimization problem (MOOP) can be formulated as follows:

$$\begin{aligned} \text{minimize } F(x) &= [f_1(x), (f_2(x)\dots, f_m(x))]^T \\ \text{s.t. } x &\in D, \end{aligned} \quad (2)$$

where D is the decision space and $x \in D$ is a decision variable. $F(x)$ consists of m objective functions $f_i : D \rightarrow O, i = 1, \dots, m$, where O is the objective space. The function of two objectives often trade-offs with each other, as, improvement in one objective may lead to degradation of another. A decision maker (DM) whose having expert domain knowledge, implicitly choose a solution which can optimize all objectives simultaneously. The best trade-off solutions are called the Pareto optimal solutions. In this paper, a feature selection problem is formulated under a multi-objective optimization approach where subset of features is selected by simultaneously optimizing more than one objective. Since there are multi-objectives, we propose to use the evolutionary algorithm (EA) of Non dominated sorted genetic algorithm-II (NSGA-II). Although, there are many EA's available in the literature and the selection of "best" algorithm certainly depends upon the characteristics of each problem (No free lunch), NSGA-II is used in this paper due to its scale-up capability [17]. The two contradictory objectives, considered in this paper are: i) the number of selected features, and ii) Accuracy. The main goal of this paper

is to propose an efficient evolutionary based multi-objective feature selection approach. As obtaining the maximum accuracy is the prime objective with minimum number of features, we select optimal feature in two stages. In the first stage, subset of features is randomly selected by the initialization step of NSGA-II. In the second stage, mutual correlation coefficient technique is used to get important, and relevant subset of features.

This approach is the hybrid filter-wrapper evolutionary approach where mutual correlation coefficient technique is incorporated inside the considered EA. Using mutual correlation coefficient technique, we calculate pair wise correlation for all features and then calculate average correlation for all features from pair wise correlation matrix. We use a threshold value and for reducing irrelevant and redundant feature we introduce percentile concept which make a novel correlation coefficient technique. In this concept, we only reduce percentage of features whose average correlation coefficient values are above the threshold value. Collectively, the major contribution of this paper is summarized as follows:

- 1) A novel correlation coefficient filter method is proposed and incorporated with NSGA-II, to obtain optimal subset of features.
- 2) This paper introduces novel way to reduce only less correlated features based on some threshold value in the filter approach. A novel randomized parameter, in the form of a percentile value, is introduced which decides the number of features to be reduced based on the size of the dataset.
- 3) The proposed approach enables to easily manage highly correlated dataset.

The remainder of this paper is organized as follows: In Sect. 2, literature review introduces the related existing works in this field. Section 3 details the explanation of the proposed approach of feature selection technique with the novel correlation coefficient technique. In Sect. 4, we will present experimental results of different dataset and compare the proposed approach with traditional NSGA-II. In Sect. 5, we will draw a summary of this article and outline the future research directions.

2 Literature Review

MOO approaches have achieved wide attention to solve the FS problems in many applications such as biomedical problems [2], text mining [3], image analysis [4], etc., where more than thousands of features are present. Various feature selection techniques with evolutionary algorithms (EAs) have been proposed in the literature. A survey of evolutionary feature selection techniques can be found in Xue et al. [5]. Binary genetic algorithms (Gas) are popularly used in EA when applied to feature selection. They use N dimensional binary vector for the number of features in the dataset. Here, “1” and “0”, shows whether the resultant feature is selected or not [6]. For more than 100 features, exhaustive search techniques of feature selection such as SFS (Sequential Forward Selection), SBS (Sequential Backward Selection), SFFS (Sequential Floating Forward Selection), SFBS (Sequential Floating Backward Selection) algorithms are become computationally infeasible in feature selection. To solve issue of selecting subset of features in large dataset, evolutionary computing (EC) algorithms have drawn attention of the

researchers. It has non-exhaustive search procedure which is computationally expensive but not computationally infeasible. To solve FS problems effectively, metaheuristics algorithms are coupled with wrapper methods that search the lower dimensional dataset space by iteratively calling the learning algorithm [7]. Mostly used metaheuristics algorithms focused on single objective of FS problems are GA [7, 17], Genetic Programming (GP) [8], Particle Swarm Optimization (PSO) [9], Differential Evolution (DE) [4], Ant Colony Optimization (ACO) [10]. Multi-objective EAs like NSGA-II [11, 20], multi-objective evolutionary algorithm with domain decomposition (MOEA/D) [12], Multi-objective particle swarm optimization (MOPSO) [13] are mostly used for the multi-objective FS problems. Hu et al. proposed fuzzy multi-objective FS method with particle swarm optimization, called PSOMOFS, where a fuzzy dominance relationship is developed to compare the goodness of candidate particles and global leader of particles are determined by fuzzy crowding distance measure [14]. Xue et al. proposed PSO based multi-objective FS algorithms where feature selection problem addressed by non-dominated sorting and applying crowding distance, mutation, and dominance into PSO [15]. Chen et.al. Proposed an efficient ACO for image feature selection [16]. In ACO, graph is made to solve feature selection problem in which each feature is considered a node of the graph. Any feature i.e., node is selected if an ant has visited the node. Hancer et al. [22] developed first multi-objective artificial bee colony (MOABC) framework for feature selection in classification, where a new fuzzy mutual information-based criterion is proposed to evaluate the relevance of feature subsets. Khushaba et al. [18] proposed a novel feature selection algorithm by combining DE with ACO where DE was used to search for the optimal feature subset based on the solutions obtained by ACO.

3 Multi-objective Optimization Based Feature Selection (MOOFS)

Finding the optimal feature subset and maintaining the classification accuracy of low and high dimensional dataset are the goal of our proposed approach. For this purpose, we propose a novel Multi-Objective Optimization Based Feature Selection (MOOFS). The candidate solution of MOOFS is encoded as a vector of n bits where each bit can take value of 1 or 0 and number of selected features is decided according to the value of 1 in the vector. Each vector is an individual in the population. For example, binary vector solutions X of a dataset with n features represent as:

$$X = (x_1, x_2, x_3, x_4 \dots, x_n), x_j \in \{0, 1\} \quad (3)$$

Then, the selected feature subset of X is:

$$X = (x_1, x_2, x_3, x_4 \dots, x_d) \quad (4)$$

where $d < n$ and $x_j = 1$ represents that the corresponding j th feature is selected.

In our proposed MOOFS approach, we use binary NSGA-II algorithm with mutual correlation coefficient technique where for all feature pair, we calculate average absolute mutual correlation of a feature over k ($= n - 1$) features, where n is the total number of features and k has $(n - 1)$ features. Since NSGA-II algorithm randomly select initial population and offspring are generated using crossover and mutation, there is high

possibility to select unwanted and noisy feature in each generation. To select relevant features efficiently, after generating the population of feature set by NSGA-II, we use mutual correlation coefficient technique (MCCT) before evaluating the selected features. MCCT reduce the uncorrelated and less important features and gives the best subset of features. These subsets of features evaluate by classification model to get the accuracy. To clearly understand the proposed MOOFS, we introduce a novel algorithm (Algorithm 1) and then explain the step-by-step procedure.

Algorithm 1: Multi-Objective Optimization Based Feature Selection

Input: Population Size, Maximum Iteration, Dataset (D), Classification Algorithm,

Output: A set of non-dominated solutions.

1. Generate Randomly initialize binary solutions with Population Size
 2. For it 1 = to Maximum Iteration
 - a. Find out Subset-Features-Population according to value 1 in each individual of Population
 - b. Reduced Subset-Features-Population: = Call Percentile_Correlation_Coefficient (Subset of Features, D)
 - c. Evaluate the whole Subset-Features-Population with KNN Classifier.
 - d. Apply Non-Dominate sorting and Calculate Crowding distance.
 - e. Do Selection, Crossover, and Mutation
 - f. Generate New Population.
 3. Return: The non-dominated pareto optimal front (POF).
-

In Step 1, a population with M individuals is randomly initialized in the binary space $\{0, 1\}$. The variable size of a population is determined according to the feature size of a dataset. The core process of (MOOFS) algorithm starts from Step 2, which is iterated until the stopping criteria is met or a maximum number of iterations is reached. In Step 2(a), the subset of features is identified with a position of 1 in the chromosome. Then in Step 2(b), individuals are randomly generated, noisy and unwanted feature may be included which can degrade the accuracy of the classifier. After that in Step 2(c), k-nearest neighbors (KNN) classifier is used to evaluate the reduced subset of features with l-fold cross validation ($l = 10$) technique. KNN is one of the widely used classifier in evolutionary feature selection algorithms. The number of selected subset of features and accuracy with KNN classifier are two objectives of our MOOFS algorithms. The optimal features are selected by simultaneously optimizing the objectives which are number of selected features and accuracy. The objective functions are described as follows:

- i) The number of selected features: The objective of this function is minimizing number of features:

$$\min F_1(X) = |X| \quad (5)$$

where $|X|$ denotes the cardinality of selected subset of features.

- ii) Accuracy: Maximizing the accuracy of the classification model represent the higher performance of classification. In this paper, accuracy is calculated by the KNN

classifier with k fold cross-validation method ($k = 10$) and objective function is defined as follows:

$$\max F_2(X) = \left(\frac{1}{k} \sum_{i=1}^k \frac{N_{Cor}}{N_{All}} \right) \times 100 \quad (6)$$

where N_{Cor} denotes the correctly classified test samples, and N_{All} is the total number of test samples.

In Step 2(d), Non-dominated sorting used on the population and the crowding distance is calculated, which are key process of NSGA-II algorithm. In Step 2(e, f), selection, crossover and mutation is performed. The output of crossover and mutation is used to generate new population for the next generation. Repeat the step 2 until the maximum number of iterations/generations is reached. At the end in Step 3, the non-dominated Pareto optimal front (POF) is returned where POF includes the number of selected features and accuracy.

To improve the performance of the classifier, important, linearly independent features need to be selected by removing noisy and unwanted feature. This task is done using Algorithm 2 (Percentile_correlation_coefficient), which returns the reduced subset of features after removing constant, and highly correlated features. In this algorithm, basic correlation coefficient method is used with some conditional step. The correlation coefficient technique measures the linear dependency or uncorrelation between two features. Two features are uncorrelated when their correlation coefficient is 0 and linearly dependent when their correlation coefficient is +1 (positively correlated) or -1 (negatively correlated). Generally, the approach of algorithm 2 may be refereed as a filter approach.

Algorithm 2: Percentile_Correlation_Coefficient

Input: Subset of Features, Dataset D

Output: Reduced Features Subset

1. Initialize percentile value p , number of features n , threshold value
 2. Calculate w according percentile value with formula $w = \text{round}(n * p)$
 3. For $t=1$ to w
 - a. Set $k=n-1$
 - b. Calculate Correlation Coefficient of Data set D according to Eq. (5)
 - c. Compute average absolute mutual correlation of all features according to Eq. (4)
 - d. Set $s =$ largest average mutual correlation
 - e. If $s >$ threshold remove having largest average mutual correlation feature from Dataset D
 - f. Set $n=n-1$
 4. Return Reduced Subset of Feature
-

Algorithm 2 shows the pseudocode of the percentile correlation coefficient algorithm. In step 1 of this algorithm 2, we initialize the percentile value p ($p = 0.3$), number of features (n) according to the input parameter-subset of features, and threshold value (threshold = 0.9). Then, we calculate variable, w , according to the value of p and n . The

value of w decides how many times step 3 will iterate to reduce the feature following the condition of threshold value. In step 3(a), we define a variable $k = n-1$ and in 3(b), the correlation coefficient matrix of Dataset (D) is computed, according to Eq. (7):

$$r_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \tag{7}$$

Then, step 3(c) computes the average absolute mutual correlation of one feature to $(n-1)$ feature according to Eq. (8), which is given as follows:

$$r_{i,k(=n-1)} = \frac{1}{k} \sum_{j=1, j \neq i}^k |r_{x_i, x_j}| \tag{8}$$

where i is the i^{th} feature, k denotes the all $n-1$ features except i^{th} feature and j denotes 1 to k features except i^{th} feature.

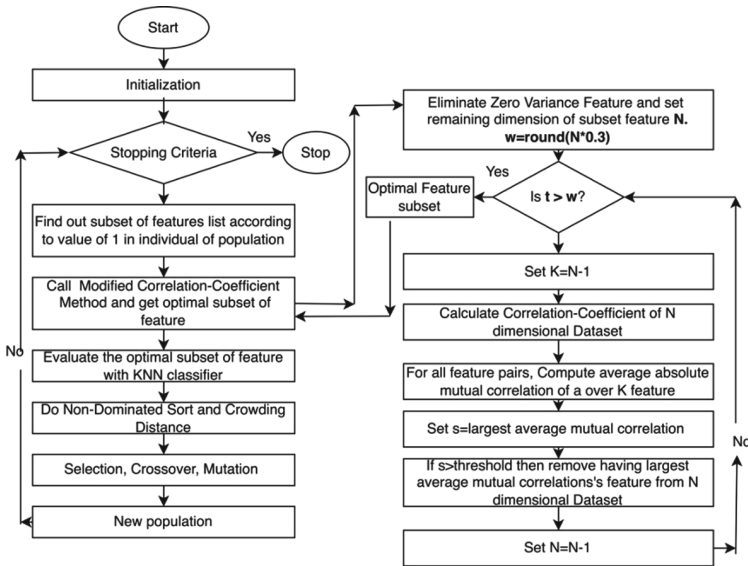


Fig. 1. Flowchart for multi-objective optimization based feature selection

Find out largest average mutual correlation value and assign to the variable ‘s’ (step 3(d)). Feature with the largest average mutual correlation will be removed in each iteration, if the value of s is greater than threshold value (step 3(e)). Repeat the step 3 until the t is equated to w . As step 3 will iterate up to the value of w which means we only reduce maximum w feature and minimum zero feature following the condition of step 3(e). Further, the process will be iterated up-to stopping criteria and return the best reduced subset of features (step 4). In this algorithm, we can manage highly correlated dataset, where all feature’s average absolute mutual correlation value is greater than the threshold which was impossible in only correlation coefficient technique. Figure 1 pictorially depicts the flowchart of proposed algorithm 1 and 2.

4 Experimental Results

The proposed MOOFS uses NSGA-II for the feature selection. To validate the efficiency of the proposed approach, we have used several datasets with various dimensions and compared MOOFS with traditional NSGA-II approach. This section is divided into two subsections. In first we have discussed about the datasets used and parameters settings of the MOO approaches. In the later sub-section, we present the experimentation results.

i. *Datasets and parameter settings*

The details of the datasets used in this paper in mentioned in Table 1 which are taken from UCI dataset [19]. The number of features varies from 33 to 241, while the number of instances goes from 351 to 4464. The maximum iteration in the NSGA-II is 500 while the number of populations is 50. In the percentile_correlation_coefficient method, the parameter threshold is experimented for 0.8 and 0.9 [21, 23].

Table 1. Datasets and their dimensions

Dataset	#Instances	#Features	Classes
Ionosphere	351	33	2
Connectionist-Bench (Sonar)	208	60	2
Hill-Valley	606	100	2
Musk1(Clean1)	476	166	2
Tuandromd	4464	241	2

ii. *Experimental result and discussion.*

Our proposed evolutionary Feature selection algorithm (MOOFS) is compared with standard NSGA-II algorithm to validate its efficiency. The first set of our experimental results are Tables 2 and 3, consisting of best accuracy value corresponding to the number of features. These tables also show the comparison among NSGA-II and our proposed MOOFS. Table 2 shows the accuracy and number of feature values when the threshold values is 0.8, whereas Table 3 shows the values for 0.9 threshold.

Table 2. Best accuracy of proposed MOOFS (with 0.8 threshold) and NSGA-II

Dataset	MOOFS		NSGA-II	
	Best accuracy	#Feature	Best accuracy	# Feature
Ionosphere	94.8571	7	94.2857	9
Sonar	76.5	7	76	7
HillValley	64.9697	10	65.8333	10
Musk1	86.6479	26	84.5202	28
Taandromd	96.9389	39	96.8280	46

From the Table 3, it's evident that best values are obtained when the threshold value is chosen as 0.9. For ionosphere dataset, MOOFS results in a higher accuracy of 94.865 with only six number of features, however, NSGA-II returned nine optimal features with lesser accuracy of 94.286. For Sonar and HillValley datasets, best accuracy is again achieved by MOOFS which is 76.286 and 65.97 with seven and ten features, respectively. The NSGA-II achieved less accuracy but with same number of features.

Table 3. Best accuracy of proposed MOOFS (with 0.9 threshold) and NSGA-II

Dataset	MOOFS		NSGA-II	
	Best accuracy	#Feature	Best accuracy	# Feature
Ionosphere	94.8650	6	94.2857	9
Sonar	76.2857	7	76	7
HillValley	65.9696	10	65.8333	10
Musk1	85.0622	23	84.5202	28
Taundromd	97.1384	39	96.8280	46

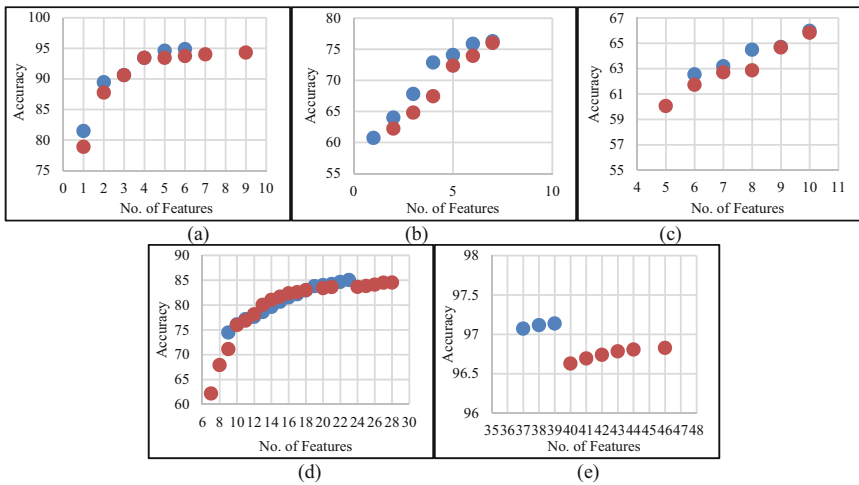


Fig. 2. POF of accuracy vs. no. of features of datasets: (a) Ionosphere, (b) Sonar (c) Hill Valley, (d) Musk1, (e) Taundromd (Blue Color: Proposed MOOFS; Red Color: NSGA-II) (Color figure online)

The significant amount of improvement is achieved for the Musk1 and Taundromd datasets, where the accuracy is significantly higher with a smaller number of features as compared to the traditional NSGA-II approach. The improvement in accuracy is achieved with around 18% (Musk1) and 15% (Taundromd) reduction in number of features for the datasets. The marginal improvement in accuracy for HillValley dataset is due to the highly correlation among the features. It is clearly observed that our proposed MOOFS algorithm shows better results compare to standard NSGA-II. Figure 2 shows the POFs of all the datasets. In Fig. 3(a-e), accuracy convergence graph of the five datasets is shown where best accuracy of every iteration is considered.

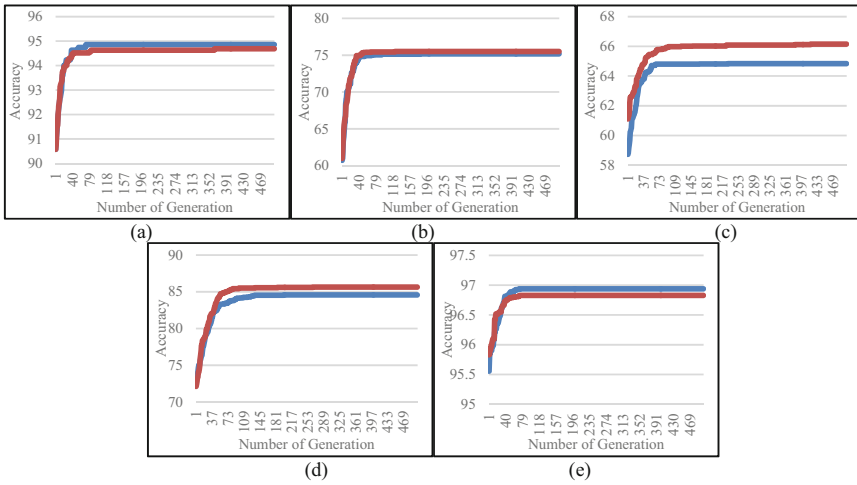


Fig. 3. Accuracy convergence graph of Features of datasets: (a) Ionosphere, (b) Sonar (c) Hill Valley, (d) Musk1, (e) Taundromd (Blue Color: Proposed MOOFS; Red Color: NSGA-II) (Color figure online)

This convergence graph is made by five independent runs with 500 iterations of the MOOFS and standard NSGA-II. Except HillValley and Musk1 dataset, we got higher accuracy convergence graph and it is also shown that MOOFS showed stable result around 250 iterations. We have further compared the proposed MOOFS with other well-known classifier in machine learning using basic correlation coefficient with threshold 0.9. The results are compiled in Table 4. Hill Valley data set select only one feature because this dataset is highly correlated. It is observed that except Hill Valley dataset, remaining 5 datasets select more number of features which are 33 for Ionosphere, 57 for Sonar, 106 for Musk1, 154 for Tuandromd. Also, most of the classifier showed less accuracy than our proposed approach. This trade-off of selecting optimal feature with higher accuracy is efficiently achieved by our proposed MOOPS approach.

Table 4. Accuracy of different classifier

Algorithm	Ionosphere	Sonar	Hill Valley	Musk1	Taundromd
Gradient boosting	91.17	57.26	50.50	62.26	98.40
Bagging classifier	91.17	57.45	51.83	66.26	98.60
Extra tree classifier	94.59	54.88	52.82	64.33	98.60
Random forest	93.73	55.90	51.83	65.00	98.40
AdaBoost	92.30	66.45	50.34	67.24	98.60
DT	86.31	58.40	50.00	62.02	98.40
Logistics regression	85.50	51.95	47.51	63.53	98.80
KNN	84.07	42.76	47.04	63.41	97.20
Support vector machine	64.14	38.09	48.00	56.87	75.00
NB	64.41	39.47	48.00	54.73	98.40
NSGA-II with KNN	94.28(9)	76(7)	65.83(10)	84.52(28)	96.78 (43)
MOOFS	94.86(6)	76.28(7)	65.96(10)	85.06 (23)	97.13 (39)

5 Conclusion and Future Work

This paper introduces a novel feature selection approach formulated under a multi-objective optimization approach where subset of features is selected by simultaneously optimizing more than one objective. Due to the scale-up capability, NSGA-II is utilized where the two considered objectives are optimal number of features, and accuracy. Several datasets with various dimensions are considered for the experimentation. The proposed approach is named Multi-Objective Optimization based Feature Selection (MOOFS), an efficient evolutionary-based multi-objective feature selection approach with a correlation coefficient filter method. The experiments are done on different datasets. From the results, a significant improvement in the accuracy with a considerable reduction in the number of features can be seen. Overall, our approach can reduce irrelevant and redundant features, which helps reduce training time as well as improve the performance of ML algorithms. One of the limitations of the proposed approach is to define the threshold value for reducing the number of features, which can be improved in future work. Further, the capability of the proposed approach for the big datasets shall be experimented. This work can also be extended to self-adapting parameter values where the correlation of the data set, dynamic in nature, can be identified.

References

1. De La Iglesia, B.: Evolutionary computation for feature selection in classification problems. Wiley Interdisc. Rev. Data Min. Knowl. Disc. **3**(6), 381–407 (2013)
2. Ahmed, S., Zhang, M., Peng, L.: Enhanced feature selection for biomarker discovery in LC-MS data using GP. In: 2013 IEEE Congress on Evolutionary Computation, pp. 584–591. IEEE (2013)

3. Aghdam, M.H., Ghasem-Aghaee, N., Basiri, M.E.: Text feature selection using ant colony optimization. *Expert Syst. Appl.* **36**(3), 6843–6853 (2009)
4. Ghosh, A., Datta, A., Ghosh, S.: Self-adaptive differential evolution for feature selection in hyperspectral image data. *Appl. Soft Comput.* **13**(4), 1969–1977 (2013)
5. Xue, B., Zhang, M., Browne, W.N., Yao, X.: A survey on evolutionary computation approaches to feature selection. *IEEE Trans. Evol. Comput.* **20**(4), 606–626 (2015)
6. Siedlecki, W., Sklansky, J.: A note on genetic algorithms for large-scale feature selection. In: *Handbook of Pattern Recognition and Computer Vision*, pp. 88–107 (1993)
7. Xue, Y., Tang, Y., Xu, X., Liang, J., Neri, F.: Multi-objective feature selection with missing data in classification. *IEEE Trans. Emerg. Topics Comput. Intell.* **6**(2), 355–364 (2021)
8. Muni, D.P., Pal, N.R., Das, J.: Genetic programming for simultaneous feature selection and classifier design. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **36**(1), 106–117 (2006)
9. Unler, A., Murat, A.: A discrete particle swarm optimization method for feature selection in binary classification problems. *Eur. J. Oper. Res.* **206**(3), 528–539 (2010)
10. Yan, Z., Yuan, C.: Ant colony optimization for feature selection in face recognition. In: Zhang, David, Jain, Anil K. (eds.) *ICBA 2004*. LNCS, vol. 3072, pp. 221–226. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-25948-0_31
11. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.A.M.T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**(2), 182–197 (2002)
12. Zhang, Q., Li, H.: MOEA/D: a multiobjective evolutionary algorithm based on decomposition. *IEEE Trans. Evol. Comput.* **11**(6), 712–731 (2007)
13. Zhang, X., Zheng, X., Cheng, R., Qiu, J., Jin, Y.: A competitive mechanism based multi-objective particle swarm optimizer with fast convergence. *Inf. Sci.* **427**, 63–76 (2018)
14. Hu, Y., Zhang, Y., Gong, D.: Multiobjective particle swarm optimization for feature selection with fuzzy cost. *IEEE Trans. Cybern.* **51**(2), 874–888 (2020)
15. Xue, B., Zhang, M., Browne, W.N.: Particle swarm optimization for feature selection in classification: a multi-objective approach. *IEEE Trans. Cybern.* **43**(6), 1656–1671 (2012)
16. Chen, B., Chen, L., Chen, Y.: Efficient ant colony optimization for image feature selection. *Sig. Process.* **93**(6), 1566–1576 (2013)
17. Muhuri, P.K., Nath, R., Shukla, A.K.: Energy efficient task scheduling for real-time embedded systems in a fuzzy uncertain environment. *IEEE Trans. Fuzzy Syst.* **29**(5), 1037–1051 (2020)
18. Khushaba, R.N., Al-Ani, A., AlSukker, A., Al-Jumaily, A.: A combined ant colony and differential evolution feature selection algorithm. In: Dorigo, M., Birattari, M., Blum, C., Clerc, M., Stützle, T., Winfield, A.F.T. (eds.) *ANTS 2008*. LNCS, vol. 5217, pp. 1–12. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87527-7_1
19. <https://archive.ics.uci.edu/ml/datasets.php?format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table>
20. Shukla, A.K., Nath, R., Muhuri, P.K., Lohani, Q.D.: Energy efficient multi-objective scheduling of tasks with interval type-2 fuzzy timing constraints in an industry 4.0 ecosystem. *Eng. Appl. Artif. Intell.* **87**, 103257 (2020)
21. Wang, F., Yang, Y., Lv, X., Xu, J., Li, L.: Feature selection using feature ranking, correlation analysis and chaotic binary particle swarm optimization. In: *2014 IEEE 5th International Conference on Software Engineering and Service Science*, pp. 305–309. IEEE (2014)
22. Hancer, E., Xue, B., Zhang, M., Karaboga, D., Akay, B.: A multi-objective artificial bee colony approach to feature selection using fuzzy mutual information. In: *2015 IEEE congress on evolutionary computation (CEC)*, pp. 2420–2427. IEEE (2015)
23. Xie, Z.-X., Qing-Hua, H., Da-Ren, Y.: Improved feature selection algorithm based on SVM and correlation. In: Wang, J., Yi, Z., Zurada, J.M., Bao-Liang, L., Yin, H. (eds.) *ISNN 2006*. LNCS, vol. 3971, pp. 1373–1380. Springer, Heidelberg (2006). https://doi.org/10.1007/11759966_204