

FEATURE SELECTION USING MULTI-OBJECTIVE MICRO-CHC GENETIC ALGORITHM: A HYBRID APPROACH

Seema Rathee

Guru Jambheshwar University of Science & Technology, Hisar,
Haryana, India
Seema27rathee@gmail.com

Saroj Ratnoo

Guru Jambheshwar University of Science & Technology, Hisar,
Haryana, India
Ratnoo.saroj@gmail.com

Abstract - The dimensionality reduction problem can be handled by employing feature selection techniques. Feature selection (FS) is a key preprocessing technique which selects more informative and useful features from data based on some criterion. Researchers have suggested many feature selection algorithms that differ in their selection criteria. Furthermore, feature selection is intrinsically a multi-objective problem with several conflicting objectives like size, redundancy and error rate. Thus, Multi-Objective Genetic Algorithms (MOGAs) are an obvious choice for solving the feature selection problem. Moreover, Micro-GA is used with a very small size population in cooperated with different type of elitisms. This work presents a Multi-objective Micro-CHC based Algorithm (MO-Micro-CHC) for feature selection. The proposed hybrid algorithm exploits the peculiarities of CHC (Cross generational elitist selection, Heterogeneous recombination, and Cataclysmic mutation), NSGA-II (Non- Dominated Sorting Genetic Algorithm-II) the most popular MOGA and Micro-GA to the advantage of arriving at better Pareto optimal solutions. The algorithm has been validated and compared to several other similar approaches on many datasets available from UCI data repository. The comparison endorses the superiority of the suggested approach.

Keywords: MOGA; Micro-GA; MO-CHC; Micro-CHC; Feature Selection.

1. Introduction

As the heap of data grow exponentially in sheer amount and diversity, the complications in processing data to seek information from it scale-up accordingly. The curse of dimensionality is a challenge to design effective data mining algorithms. Hence, feature selection has become an essential step for preprocessing of data. Feature selection reduces the features by eliminating extraneous and irrelevant features and, thereby, increases the efficiency as well as efficacy of data mining algorithms. Hence, in data mining literature, although, there already exist many feature selection techniques based on filter and wrapper approaches, yet, more and more feature selection techniques are essential to have successful applications of data mining algorithms in the fields like image processing, bio-informatics and genomics.

Evolutionary algorithms (EAs)-evolution-based generational stochastic methods renowned for their comprehensive search capability- have been widely applied in the realm of feature selection. A genetic algorithm (GA) is an evolutionary algorithm that arrives at optimal solution with the help of a randomly initialized population of candidate solutions, an objective function to evaluate the worth of a candidate solution, a fitness proportionate selection strategy and genetic operators for manipulating the solutions to further create diverse better fit solutions in the search space over successive generations. Many variations of genetic algorithms have also been applied for reducing feature and instance space [Derrac et al., (2010); Garcia et al., (2012); Jović et al., (2015); Olvera-López et al., (2010)]. In [Cano et al., (2003)], Cano has been concluded that, CHC (Cross generational elitist selection, Heterogeneous recombination and Cataclysmic mutation), a variant of GA, performs excellent for Instance Selection. CHC is a non-traditional GA which simultaneously performs an elitist selection, an effective recombination and a resume process to maintain balance between selection bias and diversity. In [García-Pedrajas et al., (2009)], authors have devised a new instance selection method with the help of CHC.

Out of the two major limitations of premature convergence and long execution time of genetic algorithms, CHC addresses the first one through its highly disruptive crossover operator and restart process on convergence. A GA can run thousands of generations before arriving at an optimal solution for complex optimization problem in the domain of data mining. This can make applications of GA infeasible in real-time optimization data mining applications. A variant of GA, known as Micro GA, suggested by [Krishnakumar, (1990)], has been reported to optimize complex combinatorial problems within reasonable time limit. A Micro-GA works with very small size populations. Micro-GA too has a reinitialization process on the loss of diversity in the population.

Further, feature selection is a familiar multi-objective optimization problem and hence it invites multi-objective treatment. A multi-objective algorithm needs to simultaneously optimize numerous fitness functions, often conflicting to each other, during the search. This leads to discovering a variety of non-dominated Pareto optimal solutions instead of the single best solution. The set of non-dominated solutions is also called the Pareto set [Deb et al., (2002); Grandchamp et al., (2015)]. Multi-objective optimization approaches use non-dominated sorting for selecting solutions for Pareto optimal front. Several multi-objective GAs have been applied to tackle the problem of feature selection [Kashyap et al., (2016); Khan and Baig, (2015); Paul and Das, (2015)]. Recently, a multi-objective CHC has been proposed by [Rathee et al., (2017), (2019), (2020)] for instance, feature as well as simultaneous selection of instance and feature selection. The authors have reported better reduction rates with competing accuracy attained over many datasets.

The objective of this research is to design a hybrid multi-objective feature selection method to harness the combined advantages of Micro-GA and CHC for feature subset selection. Multi-objective Micro CHC (MO-Micro-CHC) algorithm, proposed in this paper, is supposed to locate superior optimal solutions with a faster convergence. The two objectives considered are as usual accuracy and reduction rate. Here, feature selection is performed by using wrapper approach. The feature subsets are assessed for the task of classification and we have used KNN (1-NN) classifier.

The paper is structured as follows: Section 2 gives the essential details related to feature selection, Multi-objective optimization, CHC and Micro-GA to set the context for readers. Section 3 describes the projected hybrid approach (MO-Micro-CHC) for feature selection problem. Section 4 explains all the experiments conducted to validate the proposed approach. Finally, the last segment gives the conclusions along with future perspective of the work.

2. Background Details and Related Work

2.1. Multi-objective Optimization

Multi-objective optimization requires the optimization of multiple conflicting objective functions. In several real-life problems, there are multiple objectives to be optimized concurrently. In such situations, there is no single best solution; however, there exist a lot of solutions which are superior to each other on one or another objective function. If all the objectives are for maximization, a practical solution p is known to dominate another possible solution q ($p < q$), if and only if, $f_i(p) \geq f_i(q)$ for $i = 1, 2, 3 \dots n$ (where n depicts the total number of objective functions) and $f_j(p) > f_j(q)$ for at least one objective function j . A solution is called a Pareto optimal solution if it is not dominated by another solution in the solution space. Pareto optimal solutions will be used for this problem to find best trade-off among the objective functions. It produces multiple solutions and the set of all these multiple solutions is called Pareto optimal set. These multiple solutions facilitate a user to select a solution according to his or her choice in their application domain.

2.2. Micro-GA and CHC

The term Micro-GA (Micro-Genetic Algorithm) refers to a GA that evolves with small population size and a reinitialization procedure. Population sizes as small as 3 or 5 have shown to be sufficient to converge to an optimal solution according to [Coello and Pulido, (2001); Krishnakumar, (1990)]. In [Krishnakumar, (1990)], Krishnakumar compared simple and Micro GA in the domain of function optimization and concluded that micro-GA was quicker and gave superior results. Since Micro-GA was not capable of maintaining diversity due to its extremely small population size, it was augmented with a re-initialization process whenever convergence occurred. Although the population can be reinitialized every time when diversity drops, Micro-GA preserves best solution through elitism. Several other authors have applied micro-GAs to several other domains [Coello and Pulido, (2001); Mistry et al., (2017); Senecal, (2000); Pulido and Coello, (2003)].

2.3. About CHC

CHC, an unusual genetic search algorithm, has four basic components- an elitist selection, a highly disruptive crossover, an incest prevention mechanism and a population reinitialization process. The algorithm combines together the chromosomes of the parent and child populations. The top best chromosomes are selected permitting to the size of population for creating next population. Now, CHC applies HUX recombination operator with an incest prevention mechanism to maintain diversity and avoid early convergence. The HUX operator swaps half of the dissimilar bits subject to a minimum threshold on hamming distance between the participating parents. The parents are only allowed to mate if the hamming distance between them goes beyond

a threshold (d). Initially, d is set to half of the chromosome length (L), i.e., $L/2$. The threshold d is decremented by one when no offspring is obtained in any of the generation. When d becomes 0, it signifies a complete convergence. At this stage, CHC infuses diversity into the population through restart process which is known as cataclysmic mutation. A new population is produced by mutating 35 percent of the bits of the current population. At the same CHC maintains the global best individual into every successive generation through elitism.

CHC and its multi-objective variations have been tested by various authors in the field of instance selection. Application of Multi-objective Micro-CHC is a new experiment in the field of feature subset selection. Mo-Micro-CHC, suggested in this paper uses a small size population and elitism from Micro-GA and highly disruptive crossover genetic operator (HUX) from CHC and a restart process that uses elitism and memory archive. It also implements the non-dominated sorting process of NSGA-II.

2.4. Problem Formulation and Related Work

The main idea behind feature selection is to identify informative and relevant features from datasets. Feature selection reduces the computational overhead by decreasing dimensionality of data, improves the predictive performance and clarity of the classification models [Tan et al., (2008)]. According to search strategies, feature selection algorithms are categorized into three different models- filter, wrapper and hybrid models [Hira and Gillies, (2015); Tang et al., (2014); Jović et al., (2015); Salappa et al., (2007)]. The hybrid model takes advantage of filter and wrapper models by combining their evaluation criteria at different stages. In addition to this, FS is a branch of the Dimension Reduction problem [Hilario and Kalousis, (2008)] and Feature Extraction [Guyon et al., (2006)]. FS is an important task in many fields such as bioinformatics [Saeys et al., (2007)], text characterization [Forman, (2003)], data mining [Pappa et al., (2002)] and color image processing [Porebski et al., (2010)], etc.

A proficient global search technique is advantageous to deal with the problem of feature selection. Evolutionary Algorithms (EAs) are well-acknowledged for their global search ability. As FS consists of multiple objectives, therefore, the feature selection problem can be better solved using multi-objective meta-heuristics [Ahuja and Ratnoo, (2015); Anusha and Sathiaselvan, (2015); Khan and Baig, (2015); Mukhopadhyay et al., (2014); Pappa et al., (2002); Rathee and Ratnoo, (2020); Saroj and Jyoti, (2014)].

As discussed earlier, CHC (a variant of GA) and multi-objective CHC are considered most effective algorithm for instance, feature as well as dual selection [Cano et al., (2003), (2005); Derrac et al., (2009), (2012); Rathee et al., (2017), (2019), (2020); Ratnoo et al., (2018)]. Since GAs and MOGAs take long time to converge to optimal solutions, Micro-GA is the solution to optimize complex combinatorial problems like feature selection within reasonable time limits. A Micro-GA works with a very small size population, a reinitialization process and memory archives for implementing elitism. The detailed description of Micro-GA is given in [Coello and Pulido, (2001)]. Micro-GAs have been seldom applied in the area of feature selection [Mistry et al., (2017)]. The hybrid approach that simultaneously uses multi-objective micro-CHC for feature selection is not yet tried.

3. The Proposed Hybrid Approach for Feature Selection

The subject of feature subset selection is associated with large search space and it is multi-objective in nature. This section details the design of a hybrid multi-objective genetic algorithm for the problem of feature subset selection. The proposed algorithm combines CHC, micro GA and NSGA-II. The suggested method for feature selection has a small population inspired by Micro-GA, a highly disruptive crossover operator (HUX) adopted from CHC and imbibes elitism and restart process common to both algorithms (Micro GA and CHC). Elitism is applied through an external archive. In addition, non-dominated sorting is picked up from NSGA-II. The suggested algorithm is expected to maintain enough diversity and as well as converge early to better optimal non-dominated solutions.

3.1. Initializing Population and Memory Archives

In the population, each and every chromosome represents a set of features. A binary encoding is used to form a chromosome. The presence of a bit matches to the presence of the corresponding feature while the absence of a bit marks the exclusion of the respective feature in the chromosome. The proposed hybrid algorithm (MO-Micro-CHC) generates the initial population ($P_{t=0}$) randomly. A population memory is created which is divided into two separate parts- replaceable and non-replaceable portions in the ratio 2:3. The randomly initialized population is copied in proportionate manner in the replaceable and non-replaceable parts of the population memory. Additional external archive is initialized empty in the beginning.

3.2. Evaluating Population

MO-Micro-CHC exploits a fitness function with two objectives, i.e., maximize classification accuracy as well as minimize the number of features. The fitness function is shown below (see Eq. (1)).

$$Fitness = \left(f1 = accuracy; f2 = \left(1 - \frac{|Reduced\ feature\ set|}{|Total\ Features|} \right) * 100 \right) \quad (1)$$

The population is evaluated using the above fitness function and non-dominated fronts are computed by using the crowding distance-based sorting of NSGA-II [Deb et al., (2002)]. After evaluating the initial population, the first non-dominating front is copied into external archive.

3.3. Evolving Population

The proposed MO-Micro-CHC applies Half Uniform Crossover (HUX) operator to create child population (C_t). The HUX operator exchanges half of the bits which differ in the parent chromosomes. The bits to be exchanged are determined arbitrarily. The mutation is not applied throughout the recombination period of Micro-CHC genetic algorithm. Next, the populations (P_{t-1}) and (C_t) are merged and the best fit individuals equivalent to population size are selected to form the next generation population (P_{t+1}). The new population is evaluated, and non-dominated sorting is applied. The replaceable population memory and external archive are updated with the better fit non-dominated solutions. This cycle continues until convergence. The convergence is determined based on divergence rate threshold, a parameter to decide convergence. On convergence, a restart process is initiated. During the restart process the population is initialized from replaceable and non-replaceable portions of the population memory in the ratio 2:3. It is to be noted that non-replaceable memory never changes and contains random solutions from the initial population. This way the reinitialized population is a good mix of leading and random solutions. The above process repeats until stopping criteria is met. In the end, the best non-dominated Pareto solution from the external archive is taken as the optimal solution.

3.4. Updating memory archives and implementing elitism

Elitism is applied through population memory and external memory archive. The population replaceable memory and external archive memory get updated on every cycle of the evolutionary process to have better fit optimal solutions. If the non-dominated solutions from subsequent populations during the evolution process are better fit than the existing solution in the replaceable population memory and external memory, these memory archives are accordingly updated with better fit non-dominated solutions. A portion of better fit non-dominated solutions from replaceable population memory participate to form the new population at every reinitialization of population on convergence.

3.5. Stopping Criteria

The MO-Micro-CHC stops if there is no improvement in the solution stored in external archive for last 10 generations.

4. Experimental Details and Analysis

This section explains the detail of the experimental structure. All the experiments are conducted in MATLAB environment. Section 4.1 presents the description of the datasets. Section 4.2 enlists the parameters employed in each method. Section 4.3 compares the proposed method with some other benchmark approaches.

4.1. Datasets Description

We have tested the proposed algorithm, MO-Micro-CHC, for feature subset selection on 15 datasets taken from UCI Machine Learning Repository. These datasets are summarized in Table 1.

Table 1. Description of datasets

Sr. No.	Datasets	#Features	#Instances	#Classes
1.	Spambase	57	4701	2
2.	Waveform	40	5000	3
3.	Dermatology	34	366	6
4.	Ionosphere	34	351	2
5.	WDBC	30	569	2
6.	German	24	1000	2
7.	Vehicle	18	846	4
8.	Zoo	17	101	7
9.	Australian	14	690	2
10.	Wine	13	178	3
11.	Breast Cancer	10	683	2
12.	WBC	9	699	2
13.	Glass	9	214	6
14.	Heart	6	270	2
15.	Iris	4	150	3

4.2. Parameter Setting

For success of evolutionary algorithms, setting appropriate input parameter values is a task of utmost importance. The various parameters set for MO-Micro-CHC are tuned only experimentally in a few initial runs given in Table 2.

Table 2. The Parameter Settings for MO-Micro-CHC

Sr. No.	Name of the parameter	Value
1.	Size of Population	8
2.	Size of Population Memory	30
3.	Length of Chromosome	As per the length of the features
4.	Number of Generations	10
5.	Incest threshold	One fourth of #n

4.3. Comparison with other Baseline/Benchmark Techniques

The performance of the proposed algorithm is compared with regard to predictive accuracy and reduction rate with the following feature selection algorithms.

1. Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) -a wrapper based feature selection algorithm presented by [Kudo and Sklansky, (2000)].
2. Feature Selection based on NSGA-II, a well-known multi-objective evolutionary algorithm devised by [Deb et al., (2002)].
3. Feature selection based on Decomposition (MOEA/D)- The algorithm suggested by [Paul and Das, (2015)] decomposes an multi-objective (MO) problem into single objective problem. This approach is known as Tchebycheff approach.
4. MOGA-FS - Feature Selection methods based on several criteria like Inter-class Distance (IE), Attribute Class Correlation (AC), Inconsistent Example Pair (IP), Representation Entropy (RE) and Laplacian Score (LS) by [Spolaor et al., (2010)]. The authors have applied NSGA-II to find optimal feature subsets for ten combinations of five distinct criteria, taken two at a time.
5. MO-CHC- Feature Selection method based on Multi-objective CHC evolutionary algorithm. This algorithm is multi-objective CHC and does not use a micro size population and memory archive as done in the case of MO-Micro-CHC.

4.4. Comparison Results and Analysis

Tenfold cross validation sampling is adopted across all the feature selection algorithms. The performance evaluation is carried out by measuring accuracy and reduction rate by using KNN (1-NN) as the classification algorithm. The predictive accuracy and reduction rate reported are the average obtained over tenfold sampling. Table 3 depicts the comparison between classification accuracies of 1-NN and the proposed MO-Micro-CHC algorithm. The results show that the suggested algorithm achieves higher accuracy in most of the domains as compared to 1-NN. Concerning reduction rate, there is more than 80% reduction of features except in case of *iris* dataset.

Table 3. Comparison Results of MO-Micro-CHC with 1-NN Classification Accuracy

Datasets	1-NN Accuracy	MO-Micro-CHC Accuracy	MO-Micro-CHC Reduction Rate (RR)
Spambase	81.89	99.30	87.19
Waveform	89.90	99.11	87.50
Dermatology	89.79	99.44	94.12
Ionosphere	90.00	99.42	87.87
WDBC	97.76	98.21	83.33
German	61.00	100	95.00
Vehicle	65.17	80.90	88.89
Zoo	100	100	93.75
Australian	97.82	100	92.85
Wine	100	96.59	92.30
Breast Cancer	97.46	99.70	88.89
WBC	96.32	98.43	88.89
Glass	95.34	99.28	88.88
Heart	94.44	99.55	92.30
Iris	98.33	100	50.00

The proposed algorithm is further compared with five algorithms SFS, SBS, NSGA-II and MOEA/D in Table 4. We have taken results for SFS, SBS, NSGA-II approaches from [Paul and Das, (2015)] as shown in Table 4. The results demonstrate that MO-Micro-CHC attains higher classification accuracy in all the fourteen domains. The reduction rates achieved are also higher in all the domains. This confirms that MO-Micro-CHC converges to better non-dominated solutions with reference to accuracy and reduction rate.

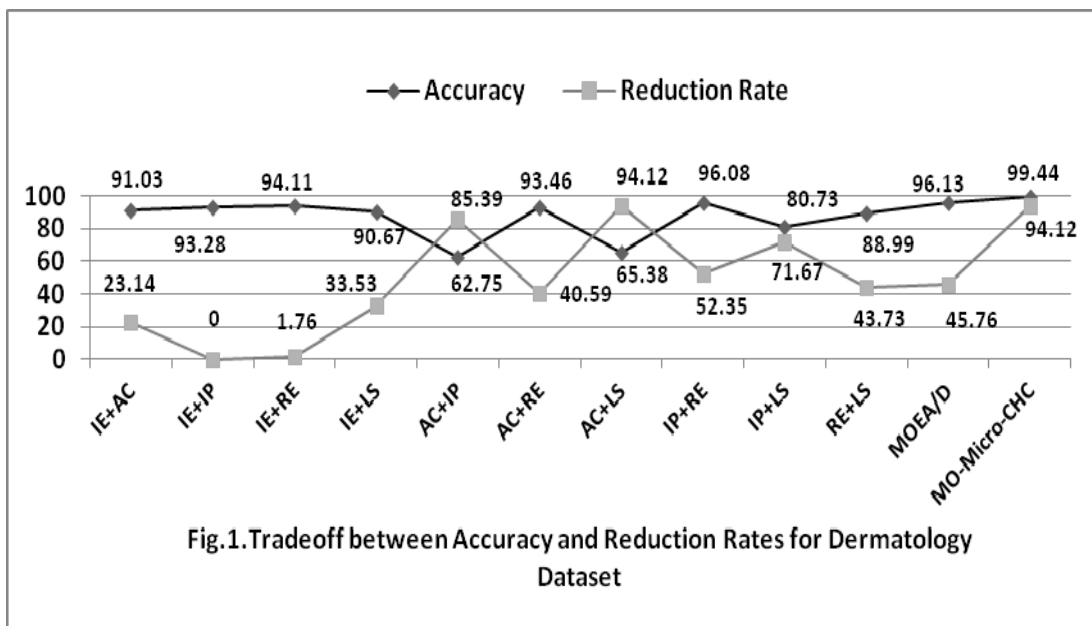
Table 4. Comparison of MO-Micro-CHC with SFS, SBS, DEMOFS, NSGA-II and MOEA/D

Algorithm Datasets	SFS		SBS		NSGA-II		MOEA/D		MO-Micro-CHC	
	Acc	RR	Acc	RR	Acc	RR	Acc	RR	Acc	RR
Spambase	87.40	37.89	87.01	34.56	88.25	3.50	88.48	54.38	99.30	87.19
Waveform	77.82	54.00	78.46	54.25	80.14	1.50	83.65	60.00	99.11	87.5
Ionosphere	88.70	78.82	85.92	73.23	87.06	2.94	88.31	66.17	99.42	87.87
WDBC	90.12	53.66	89.77	40.66	96.34	0.00	94.06	55.00	98.21	83.33
German	68.20	49.16	65.80	55.00	71.02	4.16	71.30	56.25	100	95.00
Vehicle	68.56	40.00	67.34	40.55	68.06	3.88	65.26	49.44	80.90	88.89
Zoo	94.89	47.05	98.00	23.52	94.00	18.23	95.42	35.29	100	93.75
Australian	83.02	73.57	82.83	78.57	84.45	5.00	84.64	66.42	100	92.85
Wine	91.44	53.84	91.44	42.30	95.90	13.84	96.05	46.90	96.59	92.30
Breast Cancer	95.14	39.00	94.85	39.00	96.05	2.00	96.53	57.00	99.70	88.89
WBC	95.99	28.88	95.13	18.88	95.86	4.44	96.05	53.33	98.43	88.89
Glass	63.10	35.55	63.61	28.57	66.77	17.77	67.76	51.11	99.28	88.88
Heart	65.12	68.33	62.59	61.66	78.89	0.00	80.00	11.66	99.55	92.30
Iris	93.33	17.5	93.33	35.00	96.03	50.00	97.27	50.00	100	50.00

[Spolaor et al., (2010)] have done extensive work in the domain of feature selection. They have applied five distinct measures– Attribute Class Correlation (AC), Laplacian Score (LS), Inconsistent Example Pairs (IP), Inter-Class Distance (IE) and Representation Entropy (RE)- for feature selection. They have made ten combination of these five objective measures also optimized by using NSGA-II [Deb et al., (2002)] to obtain the best possible subset of features. Again, the results shown in *Table 5* for all of the above said approaches are taken from [Paul and Das, (2015)]. The classification accuracy and reduction rates generated by MOGA-FS and MOEA/D are compared with the proposed approach, i.e., MO-Micro-CHC by applying rank-based method. The ranking is given from 1 to 12 for each dataset with rank 1 allotted to each of the method with maximum accuracy and increasing ranks for decreasing accuracies. Subsequent to that, an average rank is calculated for different datasets. Hence, smallest rank method may be considered as the best among the others. Whenever the results are same, then, the ranks are divided between them. The same ranking system is followed for Reduction Rates as well. *Table 5* presents the results of all MOGA-FS methods with different combinations of the criteria taken two at a time, MOEA/D and the proposed MO-Micro-CHC along with their ranks for three datasets (Dermatology, Vehicle and Wine) found common in these research works. The results show that MO-Micro-CHC has the top rank of 1 among the further methods of MOGA-FS and MOEA/D in case of classification accuracy. For reduction rates, MO-Micro-CHC and AC+LS are the two top ranking algorithms, however, AC+LS performs worse regarding accuracy. *Table 5* again establishes that the proposed algorithm settles for a better trade-off in between accuracy and reduction rate which is graphically depicted in Fig. 1, 2 & 3.

Table 5. Comparison of MOGA-FS, MOEA/D and MO-Micro-CHC on the bases of classification accuracy and feature reduction rate (RR)

Feature selection procedure	Dermatology		Vehicle		Wine		Average Rank of Acc (RR)
	Acc	RR	Acc	RR	Acc	RR	
IE+AC	91.03(7)	23.14 (10)	72.43 (5)	6.85 (10)	93.24 (3)	37.69 (10)	5.00 (10)
IE+IP	93.28 (6)	00.00 (12)	74.00 (2)	00.00 (12)	92.12 (4.5)	00.00 (11.5)	4.16 (11.83)
IE+RE	94.11 (4)	01.76 (11)	73.96 (3)	00.19 (11)	92.12 (4.5)	00.00 (11.5)	3.83 (11.16)
IE+LS	90.67 (8)	33.53 (9)	72.74 (4)	14.07 (9)	89.35 (8)	64.62 (8)	6.66 (8.66)
AC+IP	62.75 (11)	85.39 (3)	69.99 (6)	62.59 (7)	89.93 (7)	78.46 (4)	8.00 (4.66)
AC+RE	93.46 (5)	40.59 (8)	63.84 (10)	88.89 (2.5)	90.02 (6)	69.74 (6)	7.00 (5.5)
AC+LS	65.38 (12)	94.12 (1.5)	52.49 (11.5)	94.44 (1.5)	83.10 (11)	88.46 (2)	11.5 (1.66)
IP+RE	96.08 (3)	52.35 (5)	67.62 (8)	76.30 (5)	77.48 (12)	69.23 (5)	7.66 (5)
IP+LS	80.73 (10)	71.67 (4)	68.44 (7)	63.33 (6)	89.31 (9)	83.85 (3)	8.66 (4.33)
RE+LS	88.99 (9)	43.73 (7)	52.49 (11.5)	94.44 (1.5)	87.61 (10)	69.23 (7)	10.16 (5.16)
MOEA/D	96.13 (2)	45.76 (6)	65.26 (9)	49.44 (8)	96.05 (2)	46.90 (9)	4.33 (7.66)
MO-Micro-CHC	99.44 (1)	94.12 (1.5)	80.90 (1)	88.89 (2.5)	96.59 (1)	92.30 (1)	1(1.66)



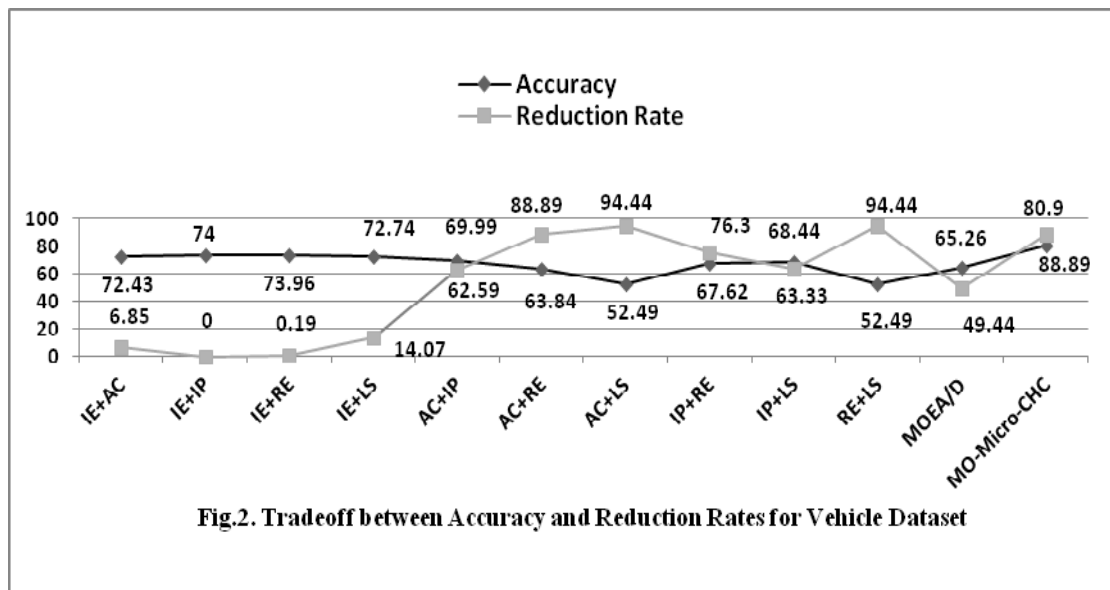


Fig.2. Tradeoff between Accuracy and Reduction Rates for Vehicle Dataset

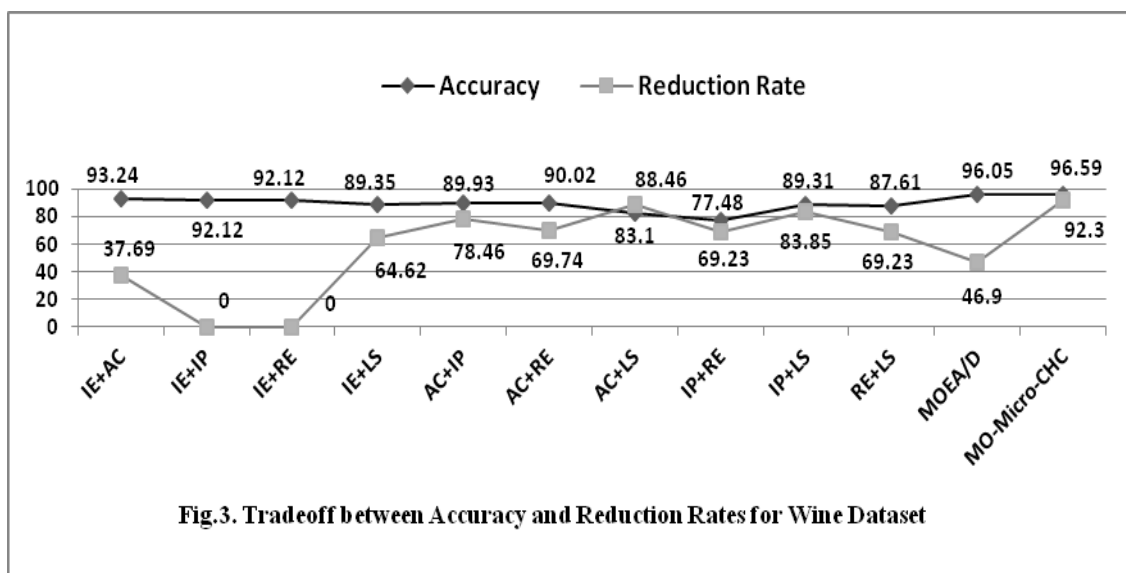


Fig.3. Tradeoff between Accuracy and Reduction Rates for Wine Dataset

The algorithm is further compared with MO-CHC for feature selection to study the behavior of the two algorithms concerning predictive accuracy, reduction rate and run-time. The comparison is shown in Table 6. We have used Wilcoxon signed rank test to see if there is any difference among the two algorithms regarding accuracy and reduction rate. The test results are negative with p values 0.255 and 0.635 for accuracy and reduction rates respectively. Hence, we can conclude that the two algorithms realize comparable accuracy and reduction rates. Furthermore, MO-Micro-CHC algorithm is far better in terms of execution time. It achieves a minimum and maximum speed gain of 2.17 and 11.0 respectively.

Table 6. Accuracy, Reduction Rate and Time comparison in between MOCHC and MO-Micro-CHC for Feature Selection

Datasets	Accuracy of MO-CHC	RR of MO-CHC	Time (in sec) of MO-CHC	Accuracy of MO-Micro-CHC	RR of MO-Micro-CHC	Time (in sec) of MO-Micro-CHC (gain)
Spambase	99.26	98.43	3387	99.30	87.19	1264 (2.67)
Waveform	98.78	97.56	3614	99.11	87.5	1306 (2.76)
Dermatology	99.32	94.28	305	99.44	94.12	140(2.17)
Ionosphere	97.56	96.96	242	99.42	87.87	34(7.11)
WDBC	99.79	88.88	251	98.21	83.33	52 (4.82)
German	100	95.00	368	100	95.00	72 (5.11)
Vehicle	78.52	94.73	228	80.90	88.89	70 (3.25)
Zoo	100	93.75	259	100	93.75	40 (6.47)
Australian	100	92.85	242	100	92.85	22 (11.0)
Wine	100	92.30	188	96.59	92.30	25 (8.54)
Breast Cancer	98.67	72.34	239	99.70	88.89	31 (7.70)
WBC	97.31	88.88	165	98.43	88.89	32 (5.15)
Glass	96.19	66.66	216	99.28	88.88	28 (7.71)
Heart	100	92.30	162	99.55	92.30	21 (7.71)
Iris	98.66	50.00	117	100	50.00	34 (3.44)

5. Conclusion

In this paper, we have devised a novel algorithm, Multi-objective Micro-CHC, for feature selection. The proposed Micro-CHC is compared with four similar algorithms based on accuracy and reduction. Since the Multi-objective micro-CHC attains high predictive performance and reduction rates in all the domains considered in this study, we can conclude that it finds more informative set of feature subsets. The suggested algorithm is also compared with several combinations of feature selection criteria for multi-objective function [Spolaor et al., (2010)] and MOEA/D [Paul and Das, (2015)]. The presentation of the proposed algorithm regarding accuracy is outstanding with top rank. MO-Micro-CHC stands at par with the top approach, (MOGA-FS with AC+LS) out of total twelve approaches tested. Looking at the accuracy and reduction rate trade-off, the performance of the suggested algorithm is quite satisfactory. For future work, we will use multi-objective parallel genetic search algorithms for multi-objective feature/instance selection.

References

- [1] Ahuja, J.; Ratnoo, S. D. (2015): Feature Selection using Multi-objective Genetic Algorithm: A Hybrid Approach. ACM Recommender System 2012 Workshop on Recommendation Utility Evaluation: Beyond RMSE, Dublin, Ireland, 14(1), pp. 26–37.
- [2] Anusha, M.; Sathiascelan, J. G. R. (2015): Feature Selection Using K-Means Genetic Algorithm for Multi-objective Optimization. Procedia Computer Science, 3rd International Conference on Recent Trends in Computing (ICRTC), 57, pp. 1074–1080.
- [3] Cano, J. R.; Herrera, F.; Lozano, M. (2003): Using evolutionary algorithms as instance selection for data reduction in KDD: an experimental study. IEEE Transactions on Evolutionary Computation, 7(6), pp. 561–575.
- [4] Cano, J. R.; Herrera, F.; Lozano, M. (2005): Stratification for scaling up evolutionary prototype selection. Pattern Recognition Letters, 26(7), pp. 953–963.
- [5] Coello, C. A. C. C.; Pulido, G. T. (2001): A Micro-Genetic Algorithm for Multiobjective Optimization. International Conference on Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 126–140.
- [6] Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. (2002): A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. Transaction Evolutionary Computation, 6(2), pp. 182–197.
- [7] Derrac, J.; García, S.; Herrera, F. (2009): A First Study on the Use of Coevolutionary Algorithms for Instance and Feature Selection. Hybrid Artificial Intelligence Systems, Lecture Notes in Computer Science, E. Corchado, X. Wu, E. Oja, Á. Herrero, and B. Barua, eds., Springer Berlin Heidelberg, pp. 557–564.
- [8] Derrac, J.; García, S.; Herrera, F. (2010): A Survey on Evolutionary Instance Selection and Generation. International Journal of Applied Metaheuristic Computing, 1(1), pp. 60–92.
- [9] Derrac, J.; Triguero, I.; García, S.; Herrera, F. (2012): Integrating Instance Selection, Instance Weighting, and Feature Weighting for Nearest Neighbor Classifiers by Coevolutionary Algorithms. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 42(5), pp. 1383–1397.
- [10] Forman, G. (2003): An Extensive Empirical Study of Feature Selection Metrics for Text Classification. Journal of Machine Learning Research, 3, pp. 1289–1305.
- [11] García, S.; Derrac, J.; Cano, J.; Herrera, F. (2012): Prototype Selection for Nearest Neighbor Classification: Taxonomy and Empirical Study. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(3), pp. 417–435.
- [12] García-Pedrajas, N.; Castillo, J. A. R. del; Ortiz-Boyer, D. (2009): A cooperative coevolutionary algorithm for instance selection for instance-based learning. Machine Learning, 78(3), pp. 381–420.
- [13] Grandchamp, E.; Abadi, M.; Alata, O. (2015): An Hybrid Method for Feature Selection Based on Multiobjective Optimization and Mutual Information. Journal of Informatics and Mathematical Sciences, 7(1), pp. 21–48.

- [14] Guyon, I.; Gunn, S.; Nikravesh, M.; Zadeh, L. A. (Eds.). (2006): Feature Extraction: Foundations and Applications. Studies in Fuzziness and Soft Computing, Springer-Verlag, Berlin Heidelberg.
- [15] Hilario, M.; Kalousis, A. (2008): Approaches to dimensionality reduction in proteomic biomarker studies. *Briefings in Bioinformatics*, 9(2), pp. 102–118.
- [16] Hira, Z. M.; Gillies, D. F. (2015): A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. *Advances in Bioinformatics*, pp. 1–13.
- [17] Jiliang Tang; Salem Alelyani; Huan Liu. (2014): Feature Selection for Classification: A Review. *Data Classification*, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, Chapman and Hall/CRC, pp. 37–64.
- [18] Jović, A.; Brkić, K.; Bogunović, N. (2015): A review of feature selection methods with applications. 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1200–1205.
- [19] Kashyap, H.; Das, S.; Bhattacharjee, J.; Halder, R.; Goswami, S. (2016): Multi-objective Genetic Algorithm setup for feature subset selection in clustering. 3rd International Conference on Recent Advances in Information Technology (RAIT), pp. 243–247.
- [20] Khan, A.; Baig, A. R. (2015): Multi-Objective Feature Subset Selection using Non-dominated Sorting Genetic Algorithm. *Journal of Applied Research and Technology*, 13(1), pp. 145–159.
- [21] Krishnakumar, K. (1990): Micro-Genetic Algorithms For Stationary And Non-Stationary Function Optimization. *Intelligent Control and Adaptive Systems*, International Society for Optics and Photonics, Philadelphia, PA, United States, pp. 289–297.
- [22] Kudo, M.; Sklansky, J. (2000): Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33(1), pp. 25–41.
- [23] Mistry, K.; Zhang, L.; Neoh, S. C.; Lim, C. P.; Fielding, B. (2017): A Micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition. *IEEE Transactions on Cybernetics*, 47(6), pp. 1496–1509.
- [24] Mukhopadhyay, A.; Maulik, U.; Bandyopadhyay, S.; Coello, C. A. C. (2014): Survey of Multiobjective Evolutionary Algorithms for Data Mining: Part I. *IEEE Transactions on Evolutionary Computation*, 18(1), pp. 4–19.
- [25] Olvera-López, J. A.; Carrasco-Ochoa, J. A.; Martínez-Trinidad, J. F.; Kittler, J. (2010): A review of instance selection methods. *Artificial Intelligence Review*, 34(2), pp. 133–143.
- [26] Pappa, G. L.; Freitas, A. A.; Kaestner, C. A. A. (2002): A multiobjective genetic algorithm for attribute selection. *Proceeding 4th International Conference on Recent Advances in Soft Computing (RASC)*, A. Lofti, J. Garibaldi, and R. John, eds., Nottingham Trent University, Nottingham, United Kingdom, pp. 116–121.
- [27] Paul, S.; Das, S. (2015): Simultaneous feature selection and weighting – An evolutionary multi-objective optimization approach. *Pattern Recognition Letters*, 65, pp. 51–59.
- [28] Porebski, A.; Vandenbroucke, N.; Macaire, L. (2010): Comparison of feature selection schemes for color texture classification. 2nd International Conference on Image Processing Theory, Tools and Applications, pp. 32–37.
- [29] Rathee, S.; Ratnoo, S. (2020): Feature selection using multi-objective CHC genetic algorithm. *Procedia Computer Science*, International Conference on Computational Intelligence and Data Science, 167, pp. 1656–1664.
- [30] Rathee, S.; Ratnoo, S.; Ahuja, J. (2017): Instance Selection using Multi-objective CHC evolutionary algorithm. *Proceedings of Third International Conference on ICTCS*, Springer, Udaipur, India, pp. 475–484.
- [31] Rathee, S.; Ratnoo, S.; Ahuja, J. (2019): Simultaneous Instance and Feature Selection using Multi-objective CHC Algorithm. SSRN Scholarly Paper, Social Science Research Network, Rochester, NY, pp. 1–7.
- [32] Ratnoo, S.; Rathee, S.; Ahuja, J. (2018): A Clustering Based Hybrid Approach for Dual Data Reduction. *International Journal of Intelligent Engineering Informatics*, 6, pp. 468–490.
- [33] Saeys, Y.; Inza, I.; Larrañaga, P. (2007): A Review of Feature Selection Techniques in Bioinformatics. *Bioinformatics*, 23 (19), pp. 2507–2517.
- [34] Salappa, A.; Doumpos, M.; Zopounidis, C. (2007): Feature Selection Algorithms in Classification Problems: An Experimental Evaluation. *Optimization Methods Software*, 22(1), pp. 199–212.
- [35] Saroj; Jyoti. (2014): Multi-objective genetic algorithm approach to feature subset optimization. *IEEE International Advance Computing Conference (IACC)*, pp. 544–548.
- [36] Senecal, P. K. (2000): Numerical Optimization Using the Gen 4 Micro-genetic Algorithm Code. UNIVERSITY OF WISCONSIN MADISON.
- [37] Spolaor, N.; Lorena, A. C.; Lee, H. D. (2010): Use of Multiobjective Genetic Algorithms in Feature Selection. *Eleventh Brazilian Symposium on Neural Networks*, IEEE Computer Society, Sao Paulo, Brazil, pp. 146–151.
- [38] Tan, F.; Fu, X.; Zhang, Y.; Bourgeois, A. G. (2008): A genetic algorithm-based method for feature subset selection. *Soft Computing*, 12(2), pp. 111–120.
- [39] Toscano Pulido, G.; Coello Coello, C. A. (2003): The Micro Genetic Algorithm 2: Towards Online Adaptation in Evolutionary Multiobjective Optimization. *Evolutionary Multi-Criterion Optimization*, Lecture Notes in Computer Science, C. M. Fonseca, P. J. Fleming, E. Zitzler, L. Thiele, and K. Deb, eds., Springer Berlin Heidelberg, pp. 252–266.