Multi-Tier Hybrid Feature Selection by Combining Filter and Wrapper for Subset Feature Selection in Cancer Classification

Bibhuprasad Sahu*

Department of CS&IT, North Orissa University, Baripada – 757003, Orissa, India;
prasadnikhil176@gmail.com

Abstract

Objectives: We have implemented a bio-inspired algorithm, particle swarm optimization method of miRNA subset selection to indentify the irreverent and redundant miRNAs for proper assessment of cancer diagnosis. **Methods/Statistical Analysis:** In this study we develop a creative multitier framework for subset selection to improve accuracy of caner classification. In the first tier we have used different filter methods to rank miRNAs according to their class relation then using union operator we have created a combinational model (Second tier) which consist of top ranked features of individual filter methods. Here the miRNAs are indentified according to their ranking with the threshold value defined. In third tier (feature pre selection model) improvised competitive swarm optimization algorithm is used to generate feasible optimal subset from the generated weighted miRNA in second tier to detect the biomarker gene for cancer detection. To minimize the gap between exploration and exploitation we have used Mamdani Fuzzy interference system. All selected genes from the fourth tire (feature reselection) is classify with classifier such as KNN. **Findings:** The objective has successfully achieved by implementing improvised competitive swarm optimization technique. Experimental result demonstrated that the proposed ICSO-KNN performs better than other method like PSO, PCA and PSO-KNN. ICSO-KNN outperformed with less error and larger amount of new solutions. **Application/Improvements:** We have four tier frameworks as an efficient feature selection algorithm which outperforms better. This approach may help to use any other metaheuristic feature selection to solve multimodal subset problem.

Keywords: Filter, Mamdani, Wrapper, ICSO, KNN

1. Introduction

The impact of miRNA is important for growth of cancer in human body. MiRNA are the small RNAs which are more then million no's available in the human body, but all of them are not responsible for growth of such dangerous disease.

So identification of irreverent ones is a tough task for due to the high dimension of data. To improve the classification accuracy we have proposed a new method for better classification of the cancer. Because identification of the related subset of cancer genes which is called biomarker genes for the better treatment of the patient. For that reason we have considered Subset feature selection as preprocessing technique for classification of cancer. Basically if the irrelevant and redundant miRNA feature present, then applied classification algorithm takes more time complexity and space complexity for such high dimension data. So to reduce such we should consciously apply the different preprocessing techniques to avoid or minimize from the dataset. Choosing of preprocessing techniques even if we remove small amount miRNA the information gene should not get loose. So after preprocessing the data is get available with fewer no of miRNA for betterment extraction of biomarkers. Feature subset algorithm is the best way to remove irrelevant and redundant miRNA[1]. Different subset selection algorithms are proposed, exhaustive search is one which evaluates each

and every subset and find out the best feature set. The disadvantage of exhaustive search is having exponential time complexity[2-3]. So we can assume that this searching technique is suitable for small and medium size datasets but not for high dimension datasets. For high dimension data we can consider approximate feature selection techniques like filter, wrapper and embedded methods. Filter method can be used in preprocessing phase to select high rank gene using a learning algorithms.

In wrapping the learning algorithm is wrapped with search algorithm to find a subset for detection of high performance subset feature set. There are two basic types of wrapper algorithms such as sequential selection algorithms and metaheuristic algorithms[4]. Sequential selection algorithm removes features until it reaches the maximum value of objective function. Sequential forward and sequential backward are the two different methods of sequential selection[5]. A metaheuristic algorithm evaluates different subsets based on optimization of the objective function[6-8]. Different types of algorithms belong to metaheuristics nature such as Particle swarm optimization[9], Genetic Algorithms (GA)[10], Ant Colony Optimization (ACO)[11], Competitive Swarm Optimization (CSO)[12-13]. For solving continuous space problems PSO is the first option for solving the problem. Two types of PSO techniques such as Continuous PSO and Binary PSO has been proposed for various feature selection problems, it proved that CPSO performs better[14-16]. Basically for selecting feature subsets are computationally efficient by filter algorithm but it suffers with feature interaction problem. It means the optimized feature depend its interaction with other features. So the interaction between the features in the miRNA dataset may be various types such as two ways, three ways and multi ways.

It is noteworthy to discuss that importance of the miRNA varies when the feature is combine with another. In the year of 1989 the researchers started solving of feature subset selection with metaheuristics algorithms. Due to large dimension of data this approach is not suitable till 2007. Most of the researchers used GA as a metaheuristics algorithms for selection of near optimal feature subsets. In[17] proposed for selection of a feature subset a Genetic algorithm with multiple populations is used where each and every neighbors share their knowledge (Solutions). A hybrid GA was developed by[18] with combination of local search algorithm. A hybrid genetic local search algorithm was proposed by[19] with combination if KNN with feature weighting. To improve the classification accuracy GASVM

a hybridized model was designed by[20]. Similar different research subset feature selection with GA was described on[21-22]. For feature subset selection different metaheurastic algorithms are used bur Particle swarm optimization (Binary PSO and Continuous PSO) used by many of the researchers to solve problem. In case of Continuous PSO the feature with higher threshold value λ is selected otherwise feature not selected. To improve the classification accuracy a hybrid PSO-SVM model was proposed by[23] for feature selection. To solve feature selection problems[24-26] used a combination of chaotic maps with BPSO. To achieve highest featured genes and to avoid the premature problem[27-29] proposed advance BPSO algorithm by adjusting the local and global optimum. Similar different research subset feature selection with PSO was described on. A hybrid algorithm Ant Colony Optimization (ACO) and Artificial Bee Colony (ABC) was proposed by[30] to indentify the best ant of colony by exploit the bees where each bee considers their food source by the ants. A hybrid algorithm ABC and DE (differential Evolution) was proposed by[31] for best feature selection. A novel graph representation ACO by[32] which represents the features are represents as node in graph. From both nodes one represents selection of feature and other represents removing the feature. An ACO algorithm was proposed by[33] where the fitness of individual ant was defined by its classification accuracy.

From the above literature survey we understand that many researchers used multi objective optimization problem due to reduce classification error rate and number of features. Last four years PSO with multi objective feature subset used by many researchers. A multi objective ACO algorithm was proposed by[34] where ACO is used to reduce both classification error rate and number of features as described above. Multi objective feature with classification accuracy will performance optimized by using DE[35-36]. As per the literature study we analyze that none of the methods provide optimal feature subset with high dimensional dataset. So above approaches are advisable for low and medium dimension data. As per survey we conclude that most of the approaches are two tiers.

Where in first tier different filter algorithms are used to find out the best features and then a ranking approach is applied, the second tier wrapper approach is applied to the selected high ranked features.In[37-38] use two tier approaches for generating optimal feature subset from

different datasets. In two tier approach the ranked feature pool (by ranking) has not consider the importance of wrapper. Proposed model of tier 1 (Preliminary Screening) we have applied different filter approaches for subset feature selection and considered the best ranked features of individual. In tier 2 (Combinational model) For classification point of view we have combine all high ranked features of individual selection methods applied in tier 1. In the tier 3 (Feature Pre selection Model) we have applied improvised competitive swarm optimization algorithm with the current swarm to achieve the best optimized result. As per my concern no such method is applied by any researcher using feature ranked obtain by using ensemble filter with competitive swarm optimization algorithm. In this study we have multi tier subset selection technique with combination of multiple filter and multiple wrapper techniques to find out the comparative study of cancer classification accuracy. Experiment analysis indicates that proposed multi tier technique provide better performance in term minimization of miRNAs, error rate, accuracy of classification. The overall concept of this paper is summarized below.

A multi tier hybrid feature selection is purposed. In the tier 1 of the purposed model different filters are such as Mutual Information Feature Selection (MIFS)[39], Joint Mutual Information (JMI)[40], Max-Relevance Min-Redundancy (MRMR)[41], Interaction Capping (ICAP)[42], Conditional Infomax Feature Extraction (CIFE)[43-44], and Double Input Symmetrical Relevance (DISR) are used to generate the miRNA pool by detecting the top 10% rank features using union operator. Tier 2 we have calculated the weights of each miRNA based on its rank. The main motto of the tier 2 presents a filter algorithm to calculate the weight vector. In tier 3 of the purposed model we have implemented a wrapper algorithm to find out the optimal features which will provide better information about the disease. As we know top ranked features may provide better opportunity to indentify the biomarker genes from the large miRNA datasets. As per the literature survey we only few researches are done with ensemble filter wrapper using weight rank approach. The layout of this research article is presented as follows. The proposed model, experimental analysis of the proposed method, has been detailed in section 2 and 3 respectively. Finally section 4 and 5 presents the experimental analysis and conclusion drawn from this research work respectively.

2. Fundamental Concept of Applied Optimization Technique

In this section, explain the concept of competitive swarm optimization.

2.1 CSO Algorithm

Till now many researchers has taken their approaches to PSO but these are unable to solve the large scale optimization problems. The performance enhancement of the most existing PSO methods is based on complex algorithm enactment which numerously increases the computational complexity. The present PSO techniques endeavor to alter gbest and pbest value, which makes the large scale optimization performance enhancement constrained. In[9] author has introduced CSO method as the effective solution to the large scale optimization problem in which the particles keep learning from the competitors in a random manner without considering the gbest and pbest. In each phase the swarm is arbitrarily partitioned into two different groups and from each group a pair wise competition is performed. After every execution the winner is considered for the next iteration and the other one updates its position and velocity by gaining the values from the winner one.

$$v_{los}^{itc+1} = r_1^{itc} v_{los}^{itc} + R_2^{itc}\left(x_{win}^{itc} - x_{los}^t\right) + \emptyset R_3^{itc}\left(\bar{x}^{itc} - x_{los}^{itc}\right)$$

$$x_{los}^{itc+1} = x_{los}^{itc} + v_{los}^{itc+1}$$

Where itc is the iteration counter R_1^{itc}, R_2^{itc}, R_3^{itc} e three randomly generated vectors within $[0, 1]^n$, x_{win}^{itc}, x_{los}^{itc} denote the winner particle and the loser particle, respectively, x^{-itc} denotes the mean position of current swarm in iteration t, and φ controls the influence of .

2.2 CSO Algorithm

1. If (c<-0)
2. For all particle $p_i^{itc} = <x_i^{itc}, v_i^{itc}>$ in swarm p^{itc} do
3. Initialize the position x_i^{itc} and velocity v_i^{itc}
4. end for
5. while termination criteria is not meet do
6. for all particle p_i^{itc} do
7. calculate fitness $f(x_i^{itc})$
8. end for
9. while $p^{itc}=\emptyset$ do
10. randomly choose two particles p_{r1}^{itc} and p_{r2}^{itc} from p^{itc}

11. if $f\left(x_{r1}^{itc}\right)$ is better than $f\left(x_{r2}^{itc}\right)$ then

12. $p_{win}^{itc} \leftarrow p_{r1}^{itc}$, $p_{los}^{itc} \leftarrow p_{r2}^{itc}$

Else

13. $p_{win}^{itc} \leftarrow p_{r2}^{itc}$, $p_{los}^{itc} \leftarrow p_{r1}^{itc}$

14. end if

15. $v_{los}^{itc+1} = R_1^{itc} v_{los}^{itc} + R_2^{itc}\left(x_{win}^{itc} - x_{los}^{itc}\right) + \phi R_3^{itc}\left(\overline{x}^{itc} - x_{los}^{itc}\right)$

16. $x_{los}^{itc+1} = x_{los}^{itc} + x_{los}^{itc+1}$

17. $p^{itc+1} \leftarrow p^{itc+1} U\left\{p_{win}^{itc}, p_{los}^{itc}\right\}$

18. $p^{itc} \leftarrow p^{itc} \setminus \left\{p_{r1}^{itc}, p_{r2}^{itc}\right\}$

19. end while

20. itc = itc + 1

21. end while

3. The Proposed Approach

Algorithm 1 described the overall description of the proposed model. The individual stage description is stated below.

3.1 Tier 1: Preliminary Screening

1. Using different filter approaches find out the best featured miRNA.
2. The best feature can be identifying using ranking approach.

3.2 Tier 2: Combinational Model

1. Using Union Operator combine the selected feature by different filter methods.
2. Apply Rankers algorithms to predict weight of miRNA (Threshold Vector).

3.3 Tier 3: Feature Pre Selection Model

1. Using Fuzzy interference system Calculate the Value of r1.
2. Apply Improvised competitive swarm optimization algorithm(A proposed model) with current Swarm.

3.4 Tier 4: Feature Reselection

We have considered K - Nearest Neighbors (K-NN), SVM as classification algorithm. K-NN (Instance based machine learning) where k stands for no of training instances Figure 1.

4. Tier-1: Preliminary Screening

From literature survey different researchers applied filter approaches with ranker for best feature subset selection. But the used filter may give poor performance for another dataset. Due to lack of knowledge regarding the dataset it is difficult to choose best ranking based filter for a certain dataset. In the above situation we have use one filter-based ranker for feature subset selection with trial-and-error runs to consider a better filter algorithm[45]. As we know feature subset selection is a computationally expensive problem, it may suffer with high resource consumption. To prevent above disadvantage and to reduce miRNAs ranking variability we have proposed a more effective and robust filter algorithm for miRNA selection which integrates several well-known existing ensemble algorithms[46]. Here we have applied an ensemble to choose the miRNAs having top rank of each filter.

4.1 Combinational Model

In tier 2 of the proposed model used to find out the weight vector of the different features of the considered dataset. The ranked features of different classifiers are combined by evaluating the mean and the score of feature d is calculated by the mean of the ranking score of the selected miRNA ranking list. Implementing the min-max method[47] we try to normalize miRNA ranking value with 0.1 to 0.2.

4.2 Feature Pre Selection Model

In the pre selection model we have used competitive swarm optimization algorithm, during the execution of the proposed algorithm, the value of all particles are considered between [0,1]. And the proposed swarm technique performs the search operation in the continuous search.

For this algorithm we have considered a threshold parameter for mapping the solution with in continuous space to binary miRNA. Mamdani fuzzy interference is used to manage the balance between exploration and exploitation by keeping eye on the value of r_1. Here we use purpose mamdani interference system with 2 inputs and 1 output[48-52]. Normalized current iteration (It) and normalized diversity of swarm in the decision tree are two parameters for fuzzy system. So it is clear that iteration of the input 1 it explores search space at beginning and later it become convergence by lapse of iteration. So Itc for
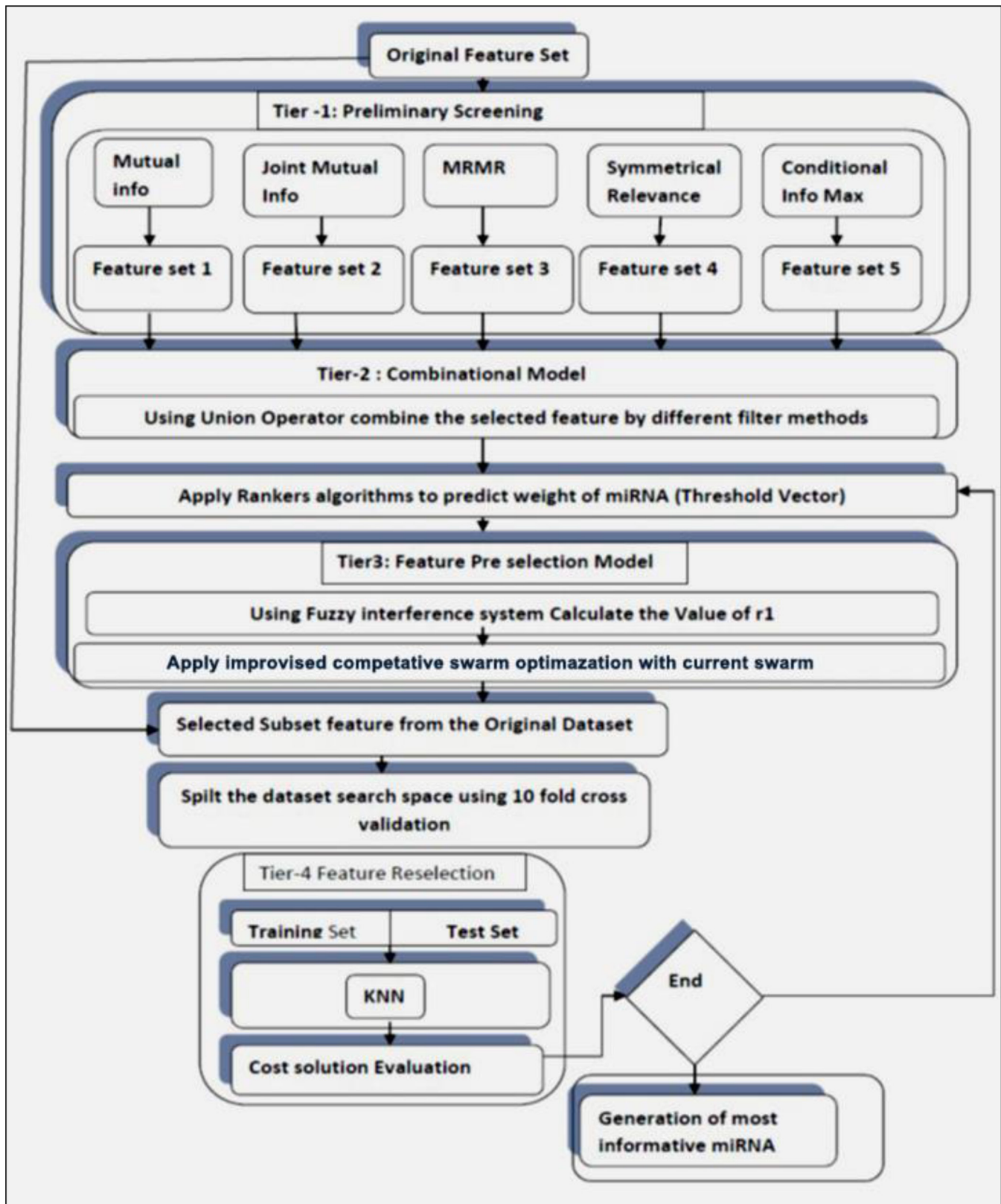
Figure 1. Block diagram of the multi-tier architecture.

fuzzy balancer evaluated as ratio of no of lapse iteration and maximum no of iteration. The range of (Itc) between (0,1). The diversity can be calculated with respect to variance of the miRNA.

$$Di = \frac{4}{n} \sum_{dim=1}^{n} variance(x_1^{dim}, x_2^{dim}, \ldots, x_N^{dim})$$

Where n is the no of miRNAs, N is the no of particles and variance of different particles with dimension dim. Here we have considered the upper bound is 1 and lower bound is 0 when the particles are converging to single solution. Like the iteration the diversity value also ranges between (0,1). The following rule based is proposed to calculate r_1 for fuzzy inference system. The rule's are set as VH=1, High=0.5, Med=0.2, Low=0.1, Vlow (VL)=0.01.

1. If (Itc) is Low (L) and diversity(D_i) is low then r_1=Vhigh(VH),
2. If (Itc) is L and (D_i) is Med(M),then r_1=High(H),
3. If (Itc) is L and (D_i) is H, then r_1=H,
4. If (Itc) is M and (D_i) is H, then r_1=H,
5. If (Itc) is M and (D_i) is M, then r_1=M,
6. If (Itc) is M and (D_i) is H, then r_1=L,
7. If (Itc) is H and (D_i) is L, then r_1=M,
8. If (Itc) is H and (D_i) is M, then r_1=L, and
9. If (Itc) is H and (D_i) is H, then r_1=Very low (Vl).

The cost evaluation function plays a major role for generating the optimal solution of a given problem. In our research we have consider K- Nearest Neighbors (K-NN), SVM as classification algorithm. K-NN (Instance based machine learning) where k stands for no of training instances.

4.3 Algorithm Details

4.3.1 Tier-1: Preliminary Screening

1. Using different filter approaches find out the best featured miRNA.
2. The best feature can be identifying using ranking approach.

4.3.2 Tier-2: Combinational Model

1. Using Union Operator combine the selected feature by different filter methods
2. Apply Rankers algorithms to predict weight of miRNA (Threshold Vector).

4.3.3 Tier 3: Feature Pre Selection Model

1. Using Fuzzy interference system Calculate the Value of r1, and
2. Applying improvised Competitive swarm optimization.

$$itc \leftarrow 0, \ Fittness \ Value(FV) \leftarrow \varphi$$

For all particle $p_i^{itc} = x_i^{itc}, \ v_i^{itc}$ in swarm p^{itc} do

Initialize position x_i^{itc} and velocity v_i^{itc}

End for

While termination criteria not meet do

End for

For all particles p_i^{itc} do

$cost_i^{itc} \leftarrow p$

For all $distance_a^{i,itc} \in x_i^{itc}$, a=1,2,….N do

if $distance_a^{i,itc} >$ Thershold (λ)

$cost_i^{itc} \leftarrow cost_i^{itc}$ U a

end if

end for

if $\left(\left(cost_i^{itc}, \star\right) \in FV\right)$

extract $f\left(disance_i^{itc}\right)$ from FV

else

calculate fitness $f\left(disance_i^{itc}\right)$ by n fold cross validation

$FV \leftarrow FV \ U \left\{cost_i^{itc}, \ f\left(cost_i^{itc}\right)\right\}$

end if

end for

$p^{itc+1} \neq \emptyset$ do

Randomly choose two different particles $p_{r1}^{itc}, p_{r2}^{itc}$ from p^{itc}

If $f\left(x_{r1}^{itc}\right)$ is better as compare to $f\left(x_{r2}^{itc}\right)$ then

$$p_{win}^{itc} \leftarrow p_{r1}^{itc}, \ p_{los}^{itc} \leftarrow p_{r2}^{itc}$$

Else

$$p_{win}^{itc} \leftarrow p_{r2}^{itc}, \ p_{los}^{itc} \leftarrow p_{r1}^{itc}$$

end if

$v_{los}^{itc+1} = R_1^{itc} v_{los}^{itc} + R_2^{itc}\left(x_{win}^{itc} - x_{los}^{itc}\right) + \emptyset R_3^{itc}\left(\bar{x}^{itc} - x_{los}^{itc}\right)$

$x_{los}^{itc+1} = x_{los}^{itc} + x_{los}^{itc+1}$

$p^{itc+1} \leftarrow p^{itc+1} \ U \left\{p_{win}^{itc}, p_{los}^{itc}\right\}$

$p^{itc} \leftarrow p^{itc} \setminus \left\{p_{r1}^{itc}, p_{r2}^{itc}\right\}$

end while

itc = itc + 1

end while

4.4 Tier 4: Feature Reselection

For classification we have considered to well-known classification techniques such as K-NN. To evaluate the threshold cost of for K-NN 10 fold cross validation we have to follow following algorithm.

4.4.1 Input Parameters

Vector x and threshold Vector
Cost=Null
For d=1 to n do
If (distance of the particle > threshold value)
Cost = Cost + distance
End if
End for
Calculate the cost for K-NN for 10 fold cross validation.
Output: Calculate the cost Vector.

5. Experimental Analysis

In this section we have evaluated the accuracy and effectiveness for the proposed algorithm to reduce the error rate of classification and minimization of no of miRNA features. Experimental section we have presented the numerical result of the proposed algorithm with comparison with different miRNA subset selection algorithms.

5.1 Dataset and Experimental Execution Settings

For experimental study we have consider two cancer datasets such as lung and melanoma from Gene Expression Omnibus (GEO) which is easily downloadable and publicly available[53]. We have not done any type preprocessing on mentioned datasets. We have spitted the dataset in to two parts such as training set and test set with ratio of 70-30%.

6. Dataset Details

Using Matlab 2015b we have implemented and tested some algorithms with K-NN classifier with k=10 and some are implemented on java SE8 using weka data mining tool .To evaluate the performance of the proposed algorithm with 2S-GA[54], 2S-HGA[39], 2S-PSO[38], Xue1-PSO, Xue2-PSO, Xue3-PSO, and Xue4-PSO[35], GA, the original CSO algorithm[25], the original PSO, DE[55], ABC-DE[31], ACO-FS[32], ACO-ABC[30], GSA[56], BQIGSA[57]. For Matlab experimental execution the population is set to 100 and

Table 1. Datasets description

Dataset Details	miRNA nos	Normal Patients nos.	No of classes	No of Patients
Lung	886	19	2	17
Melanoma	864	12	2	35

fitness function is set to 20000. We have also considered the value of c1 and c2 to 1.49 and the value of w is set to 0.7928. As per the requirement we have consider following parameter sets for program execution. For Xue's algorithm, the threshold parameter λ is set to 0.6 and 0.5 canonical PSO and original PSO respectively. The parameters for other experiment used from as per the original work done by the different authors. To achieve statistical results we have run each algorithm 40 times independently Table 1.

7. Result Study

Accuracy of a classifier depends on various criteria but the main motto is to maximize the generation capability (high accuracy and low error rate).To reduce over lifting problem we consider the average performance analysis of all algorithms. For this we have use Wilcox on rank sum test, where the symbols "+","-","≈" represents inferior to, superior or equal to of algorithms presented in Table 2. It is noteworthy to express that our proposed algorithm (ICSO-KNN) performs in term of statistical misclassification elimination (low) as compared with others. Low ranked features (miRNA) eliminated by the searching process and indentify only the high ranked ones. Table 3 presents average positive predictive values of various algorithms. Table 4 represents the statistical no of indentified miRNA compared with different miRNa sub selection algorithms. Proposed algorithm performs superiority because the searching methodology reduces irrelevant miRNAs. From the table analysis it is clearly understood, the performance of PSO depends on none of miRNAs initialized during the first generation. But the purposed methodology not concern about the initialization and it use near optimal miRNA feature subset.

Table 5 represents frequently selected miRNA. As we are mainly focusing on ensemble based filter technique, to achieve this we made a result analysis with individual filter approach and KNN against proposed method. It is presented on Table 6.

Table 2. Error rate (average) of different machine learning algorithm

Dataset	Lung	Melanoma
ICSO kNN	0.1931	0.1605
(Proposed algorithm)	-0.0159	-0.0179
GA	0.3312 +	0.2981 +
	-0.0393	-0.0296
CSO	0.2607 +	0.2350 +
	-0.0236	-0.0284
PSO	0.3515 +	0.3173 +
	-0.0442	-0.0342
DE	0.3227 +	0.3092 +
	-0.0371	-0.0307
ABC-DE	0.3073 +	0.3217 +
	-0.0481	-0.0374
ACOFS	0.3406 +	0.3686 +
	-0.0361	-0.0391
ACO-ABC	0.3204 +	0.2730 +
	-0.0382	-0.0311
GSA	0.3372 +	0.3511 +
	-0.0423	-0.0368
BQIGSA	0.3007 +	0.2761 +
	-0.0299	-0.0273
Xue1-PSO	0.3603 +	0.3154 +
	-0.049	-0.0491
Xue2-PSO	0.3249 +	0.2920 +
	-0.0405	-0.0301
Xue3-PSO	0.3618 +	0.3064 +
	-0.0506	-0.0429
Xue4-PSO	0.3242 +	0.2796 +
	-0.0345	-0.0307
2S-GA	0.2785 +	0.2532 +
	-0.0272	-0.0291
2S-HGA	0.2804 +	0.2699 +
	-0.0316	-0.0254
2S-PSO	0.2523 +	0.2382 +
	-0.0205	-0.0232

Table 3. Positive predictive value (average) machine learning algorithm

Dataset	Lung	Melanoma
ICSO kNN	0.8852	0.9217
	-0.0317	-0.0253
GA	0.7021 +	0.7326 +
	-0.0437	-0.0491
CSO	0.7419 +	0.7147 +
	-0.0379	-0.0572
PSO	0.7112 +	0.7492 +
	-0.0512	-0.042
DE	0.6826 +	0.7718 +
	-0.0417	-0.0325
ABC-DE	0.6991 +	0.7930 +
	-0.0454	-0.0458
ACOFS	0.6592 +	0.7504 +
	-0.0459	-0.0429
ACO-ABC	0.6811 +	0.8271 +
	-0.0357	-0.0342
GSA	0.6927 +	0.7906 +
	-0.0448	-0.0394
BQIGSA	0.7284 +	0.8985 ≈
	-0.0521	-0.0294
Xue1-PSO	0.6268 +	0.7892 +
	-0.0457	-0.0426
Xue2-PSO	0.6995 +	0.8427 +
	-0.0452	-0.0368
Xue3-PSO	0.6817 +	0.8244 +
	-0.0481	-0.0401
Xue4-PSO	0.7393 +	0.8625 +
	-0.0479	-0.0429
2S-GA	0.7692 +	0.8751 +
	-0.0472	-0.0284
2S-HGA	0.7824 +	0.8901 +
	-0.0385	-0.0372
2S-PSO	0.7924 +	0.9063 ≈
	-0.0322	-0.0371

8. Conclusion

The main focus of this study is to identify small feature subset from high dimensional feature dataset. The objective has successfully achieved by implementing improvised competitive swarm optimization technique. Experimental result demonstrated that the proposed ICSO-KNN performs better than other method like PSO, PCA and PSO-KNN. ICSO-KNN outperformed with less error and larger amount of new solutions. The most

Table 4. No. of MiRNAs selected by machine learning algorithm

Dataset	Lung	Melanoma
ICSO kNN	9.74	15.21
	-5.36	-7.82
GA	35.62 +	41.65 +
	-18.52	-19.23
CSO	14.27 +	20.52 +
	-13.43	-10.29
PSO	30.44 +	37.93 +
	-19.17	-21.16
DE	31.15 +	35.19 +
	-15.66	-18.8
ABC-DE	47.91 +	42.01 +
	-20.41	-21.13
ACOFS	50.14 +	40.24 +
	-23.62	-17.91
ACO-ABC	32.81 +	36.74 +
	-19.6	-19.05
GSA	52.83 +	41.10 +
	-25.12	-24.16
BQIGSA	30.05 +	37.35 +
	-17.91	-16.87
Xue1-PSO	29.37 +	36.13 +
	-17.44	-18.65
Xue2-PSO	17.64 +	23.18 +
	-21.49	-15.75
Xue3-PSO	28.14 +	34.92 +
	-19.62	-16.35
Xue4-PSO	X 24.02 +	30.44 +
	-16.71	-17.32
2S-GA	15.67 +	22.47 +
	-12.74	-10.79
2S-HGA	17.05 +	24.10 +
	-14.25	-12.32
2S-PSO	16.31 +	21.52 +
	-13.27	-11.26

important feature of the purposed algorithm is that it is not rely on the initialization for selecting feature. We have four tier frameworks as an efficient feature selection algorithm which outperforms better. The next aim is to use any other metaheuristic feature selection to solve multi-modal subset problem.

Table 5. Frequent identified MiRNAs by proposed algorithm

Lung	Melanoma
hsa-let-7d	hsa-miR-17
hsa-miR-423-5p	hsa-miR-664
hsa-let-7f	hsa-miR-145
hsa-miR-140-3p	hsa-miR-422a
hsa-miR-25	hsa-miR-216a
hsa-miR-98	hsa-miR-186
hsa-miR-195	hsa-miR-601

Table 6. Error rate (average) (proposed algorithm *vs* existing)

Dataset	Lung	Melanoma
ICSO KNN (Proposed algorithm)	0.1871	0.1537
	0.0159	0.0179
MI-k-NN	0.1822 −	0.2173 +
	0.0121	0.0198
JMI-k-NN	0.2274 +	0.2394 +
	0.0213	0.0219
MRMR-k-NN	0.2182 +	0.2201 +
	0.0195	0.0214
ICAP-k-NN	0.2418 +	0.1994 +
	0.0269	0.0201

9. References

1. Chandrashekar G, Sahin F. A survey on feature selection methods, Computers and Electrical Engineering. 2014;40(1): 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024.
2. Gheyas IA, Smith LS. Feature subset selection in large dimensionality domains, Pattern Recognition. 2010; 43(1):5–13. https://doi.org/10.1016/j.patcog.2009.06.009.
3. Computers and Intractability; A Guide to the Theory of NP-Completeness. Date accessed: 1990. Retrieved from https://dl.acm.org/citation.cfm?id=574848.
4. Xue B, Zhang M, Browne WN, Yao X. A survey on evolutionary computation approaches to feature selection, IEEE Transactions on Evolutionary Computation. 2016; 20(4):606–26. https://doi.org/10.1109/TEVC.2015.2504420.
5. Pudil P, Novovicová J, Kittle J. Floating search methods in feature selection, Pattern Recognition Letters. 1994; 15(11):1119–25. https://doi.org/10.1016/0167-8655(94) 90127-9.
6. Introduction to Machine Learning. Date accessed: 15/07/2017. Retrieved from https://towardsdatascience. com/introduction-to-machine-learning-db7c668822c4.

7. Talbi EG. Metaheuristics: From Design to Implementation. John Wiley & Sons; 2009. p.1–618. https://doi.org/10.1002/9780470496916.ch1, https://doi.org/10.1002/9780470496916.

8. Dowlatshahi MB, Nezamabadi-Pour H, Mashinchi M. A discrete gravitational search algorithm for solving combinatorial optimization problems, Information Sciences. 2014; 258:94–107. https://doi.org/10.1016/j.ins.2013.09.034.

9. Han KH, Kim JH. Quantum-inspired evolutionary algorithm for a class of combinatorial optimization, IEEE Transactions on Evolutionary Computation. 2002; 6(6): 580–93. https://doi.org/10.1109/TEVC.2002.804320.

10. Banks A, Vincent J, Anyakoha C. A review of particle swarm optimization. Part II: Hybridisation, combinatorial, multicriteria and constrained optimization, and indicative applications, Natural Computing. 2008; ,7(1):109–24. https://doi.org/10.1007/s11047-007-9050-z.

11. Cheng R, Jin Y. A competitive swarm optimizer for large scale optimization, IEEE Transactions on Cybernetics. 2015; 45(2):191–204. https://doi.org/10.1109/TCYB.2014.2322602. PMid: 24860047.

12. Dowlatshahi MB, Rezaeian M. Training spiking neurons with gravitational search algorithm for data classification. In: Proceedings of the Swarm Intelligence and Evolutionary Computation (CSIEC); 2016. p. 53–58.

13. Dowlatshahi MB, Derhami V. Winner determination in combinatorial auctions using hybrid ant colony optimization and multi-neighborhood local search, Journal of AI and Data Mining. 2017; 5(2):169–81.

14. Xue B, Zhang M, Browne WN. Multi-objective Particle Swarm Optimisation (PSO) for feature selection. Proceedings of the 14th annual conference on Genetic and evolutionary computation; 2012. p. 81–88. https://doi.org/10.1145/2330163.2330175.

15. Zhang Y, Gong DW, Cheng J. Multi-objective particles warm optimization approach forcost-based feature selection in classification, IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2017; 14(1):64–75. https://doi.org/10.1109/TCBB.2015.2476796. PMid: 26353379.

16. Siedlecki W, Sklansky J. A note on genetic algorithms for large-scale feature selection, Pattern Recognition Letters. 1989; 10(5):335–47. https://doi.org/10.1016/0167-8655(89)90037-8.

17. LiY, Zhang S, Zeng X. Research of multi-population agent genetic algorithm for features election, Expert Systems with Applications. 2009; 36(9):11570–81.

18. Kabir MM, Shahjahan M, Murase K. A new local search based hybrid genetic algorithm for feature selection, Neuro Computing. 2011; 74(17):2914–28. https://doi.org/10.1016/j.neucom.2011.03.034.

19. Eroglu DY, Kilic K. A novel Hybrid Genetic Local Search Algorithm for feature selection and weighting with an application in strategic decision making in innovation management, Information Sciences. 2017; 405:18–32. https://doi.org/10.1016/j.ins.2017.04.009.

20. Phan AV, Le Nguyen M, Bui LT. Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems, Applied Intelligence. 2017; 46(2):455–69. https://doi.org/10.1007/s10489-016-0843-6.

21. Derrac J, Garcia S, Herrera F. A first study on the use of coevolutionary algorithms for instance and feature selection, Hybrid Artificial Intelligence Systems. 2009; 557–64.

22. Oh I, Lee JS, Moon BR. Hybrid genetic algorithm for feature selection, IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004; 26(11):1424–37. https://doi.org/10.1109/TPAMI.2004.105. PMid: 15521491.

23. Huang CL, Dun JF. A distributed PSO–SVM hybrid system with feature selection and parameter optimization, Applied Soft Computing. 2008; 8(4):1381–91. https://doi.org/10.1016/j.asoc.2007.10.007.

24. Chuang LY, Yang CH, Li JC. Chaotic maps based on binary particle swarm optimization for feature selection, Applied Soft Computing. 2011; 11(1):239–48. https://doi.org/10.1016/j.asoc.2009.11.014.

25. Gu S, Cheng R, Jin Y. Feature selection for high-dimensional classification using a competitive warm optimizer, Soft Computing. 2018; 22(3):811–22. https://doi.org/10.1007/s00500-016-2385-6.

26. WeiJ, Zhang R, Yu Z, Hu R, Tang J, Gui C, Yuan Y. A BPSO-SVM algorithm based on memory renewal and enhanced mutation mechanisms for feature selection, Applied Soft Computing. 2017; 58:176–92. https://doi.org/10.1016/j.asoc.2017.04.061.

27. Bello R, Gomez Y, Garcia MM, Nowe A. Two-step particle swarm optimization to solve the feature selection problem. Seventh International Conference on Intelligent Systems Design and Applications; 2007. p. 691–96. https://doi.org/10.1109/ISDA.2007.101.

28. Tanaka K, Kurita T, Kawabe T. Selection of import vectors via binary particle swarm optimization and cross-validation for kernel logistic regression. International Joint Conference on Neural Networks; 2007. p.12–17.

29. Wang X,Yang J, Teng X, Xia W, Jensen R. Feature selection based on rough sets and particle swarm optimization, Pattern Recognition Letters. 2007; 28(4):459–71. https://doi.org/10.1016/j.patrec.2006.09.003.

30. Sivagaminathan RK, Ramakrishnan S. A hybrid approach for feature subset selection using neural networks and ant colony optimization, Expert System Application. 2007; 33(1):49–60. https://doi.org/10.1016/j.eswa.2006.04.010.

31. Kashef S, Nezamabadi-pour H. An advanced ACO algorithm for feature subset selection, Neuro Computing. 2014; 147:271–79. https://doi.org/10.1016/j.neucom.2014.06.067.

32. Ghosh A, Datta A, Ghosh S. Self-adaptive differential evolution for feature selection in hyper spectral image data, Applied Soft Computing. 2013; 3(4):1969–77. https://doi.org/10.1016/j.asoc.2012.11.042.

33. Shunmugapriya P, Kanmani S. A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid), Swarm and Evolutionary Computation. 2017; 36:27–36. https://doi.org/10.1016/j.swevo.2017.04.002.

34. Zorarpacı E, Özel SA. A hybrid approach of differential evolution and artificial bee colony for feature selection, Expert Systems with Applications. 2016; 62:91–103.

35. Vieira S, Sousa J, Runkler T. Multi-criteria ant feature selection using fuzzy classifiers. Swarm Intelligence for Multi-objective Problems in Data Mining; 2009. p.19–36. https://doi.org/10.1007/978-3-642-03625-5_2.

36. Xue B, Zhang M, Browne WN. Particle swarm optimization for feature selection in classification: A multi-objective approach, IEEE Transactions on Cybernetics. 2013; 43(6):1656–71. https://doi.org/10.1109/TSMCB.2012.2227469. PMid: 24273143.

37. Xue B, Fu W, Zhang M. Multi-objective features election in classification: A differential evolution approach. Asia-Pacific Conference on Simulated Evolution and Learning; 2014. p. 516–28. PMid: 25360650.

38. Sahu B, Mishra D. A novel feature selection algorithm using particle swarm optimization for cancer microarray data, Procedia Engineering. 2012; 38:27–31. https://doi.org/10.1016/j.proeng.2012.06.005.

39. Oreski S, Oreski G. Genetic algorithm-based heuristic for features election in credit risk assessment, Expert Systems with Applications. 2014; 41(4):2052–64. https://doi.org/10.1016/j.eswa.2013.09.004.

40. Tan F, Fu XZ, Zhang YQ, Bourgeois A. A genetic algorithm based method for feature subset selection, Soft Computing. 2008; 12(2):111–20. https://doi.org/10.1007/s00500-007-0193-8.

41. Battiti R. Using mutual information for selecting features in supervised neural net learning, IEEE Transactions on Neural Networks. 1994; 5(4):537–50. https://doi.org/10.1109/72.298224. PMid: 18267827.

42. Yang H, Moody J. Data visualization and feature selection: New algorithms for non-gaussian data. Advances in Neural Information Processing Systems; 1999. p. 687–93.

43. Peng H, Long F, Ding C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005; 27(8):1226–38. https://doi.org/10.1109/TPAMI.2005.159. PMid: 16119262.

44. Machine Learning Based on Attribute Interactions. Date accessed: 13/06/2005. http://stat.columbia.edu/~jakulin/Int/jakulin05phd.pdf.

45. Lin D, Tang X. Conditional infomax learning: An integrated framework for feature extraction and fusion European Conference on Computer Vision; 2006. p.68–82.

46. Meyer P, Bontempi G. On the use of variable complementarity for feature selection in cancer classification. Proceedings of the Evolutionary Computation and Machine Learning in Bioinformatics; 2006. p.91–102. https://doi.org/10.1007/11732242_9.

47. Han J, Pei J,Kamber M. Data Mining: Concepts and Techniques, Elsevier; 2011. p. 1–740.

48. Dowlatshahi MB, Nezamabadi-Pour H. GGSA: A grouping gravitational search algorithm for data clustering, Engineering Applications of Artificial Intelligence. 2014; 36:114–21. https://doi.org/10.1016/j.engappai.2014.07.016.

49. Črepinšek M, Liu SH, Mernik M. Exploration and exploitation in evolutionary algorithms: A survey, ACM Computing Surveys. 2013; 45(3):1–35.

50. Rafsanjani MK, Dowlatshahi MB. Using Gravitational search algorithm for find in gnear-optimal base station location in two-tiered WSNs. In: Proceedings of the 3rd International Conference on Machine Learning and Computing; 2011. p. 213–16.

51. Dowlatshahi MB, Derhami V, Nezamabadi-pour H. Ensemble of Filter-Based Rankers to Guide an Epsilon-Greedy Swarm Optimizer for High-Dimensional Feature Subset Selection, Information. 2017; 8(4):1152. https://doi.org/10.3390/info8040152.

52. Li H, Wang J, Du H, Karimi HR. Adaptive sliding mode control for Takagi-Sugeno fuzzy systems and its applications, IEEE Transactions on Fuzzy Systems. 2018; 26(2):531–42. https://doi.org/10.1109/TFUZZ.2017.2686357.

53. Abbasifard MR, Ghahremani B, Naderi H. A survey on nearest neighbor search methods, International Journal of Computer Applications. 2014; 95(25):39–52.

54. Gene Expression Omnibus (GEO). Date accessed: 2017. https://www.ncbi.nlm.nih.gov/geo/.

55. Han X, Chang X, Quan L, Xiong X, Li J, Zhang Z, Liu Y. Feature subset selection by gravitational search algorithm optimization, Information Sciences. 2014; 281:128–46. https://doi.org/10.1016/j.ins.2014.05.030.

56. Barani F, Mirhosseini M, Nezamabadi-pour H. Application of binary quantum-inspired gravitational search algorithm in feature subset selection, Applied Intelligence. 2017; 47(2):304–18. https://doi.org/10.1007/s10489-017-0894-3.

57. Derrac J, García S, Molina D, Herrera F. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, Swarm and Evolutionary Computation. 2011; 1(1):3–18. https://doi.org/10.1016/j.swevo.2011.02.002.