_____

Filter – GA Based Approach to Feature Selection for Classification

Amit Saxena
Dept. of CSIT
Guru Ghasidas Vishwavidyalaya
Bilaspur, India
amitsaxena65@rediffmail.com

Madan Madhaw Shrivas
Dept. of CSIT
Guru Ghasidas Vishwavidyalaya
Bilaspur, India
Madhaw_m@yahoo.com

Abstract— This paper presents a new approach to select reduced number of features in databases. Every database has a given number of features but it is observed that some of these features can be redundant and can be harmful as well as and can confuse the process of classification. The proposed method applies filter attribute measure and binary coded Genetic Algorithm to select a small subset of features. The importance of these features is judged by applying K-nearest neighbor (KNN) method of classification. The best reduced subset of features which has high classification accuracy on given databases is adopted. The classification accuracy obtained by proposed method is compared with that reported recently in publications on twenty eight databases. It is noted that proposed method performs satisfactory on these databases and achieves higher classification accuracy but with smaller number of features.

Keywords- *Data Mining, Genetic algorithm (GA), K-nearest neighbor (KNN), Information Gain, Gain Ratio, ReliefF, Chi-Squared, Correlation Feature Selection (CFS),Classification, Feature Selection, Leave-one-out cross validation (LOOCV).*

_____***** *****_____

I. INTRODUCTION

Due to growth of computer science and information technology the dimension of database are increasing day by day in every industry and organizations. The high dimension [1] data creates the problem in learning process of data mining and machine learning task such as data analysis, information retrieval processing, and data/pattern classification process and degenerate the performance of learning tasks. While designing of dataset, where some data is relevant and some data is irrelevant, feature selection play very important role to identify the relevant and irrelevant features. Feature selection is a technique to remove the irrelevant feature from the data to improve the performance of data mining and machine learning techniques. In machine learning and data mining application feature selection [2] and classification [3] techniques are used for data analysis and data identification process. Both these techniques are known as the supervised techniques. Feature selection is a preprocessing step that is used before classification process to computationally effective and efficient model. Feature selection techniques are used to identify the relevant features and remove irrelevant, redundant, noisy and harmful features from high dimension dataset or original set of features. It is also helpful to improve the classification performance, reduced computation time and deled understandable model. Supervised feature selection means that the class of the pattern is given in advance. These methods include three types of searching strategies i.e. filter method [4], [5], [6], [7] wrapper methods [8], [9], [10] and embedded methods [11]. Filter methods select the most discriminative features with highest ranking from the dataset without using any learning algorithm. Wrapper method uses the intended learning/ classification algorithm itself to evaluate quality of important features. Embedded models perform features selection in the process of model construction [11]. Both wrapper and embedded are classifier dependent method. Feature selection is used in many applications such as machine learning [5], data mining [3],[6],[12] Pattern recognition [13] gene selection from microarray data [7] text categorization [14], multimedia information retrieval [15], [16], image processing [17], signal processing [18], and statistics [19], by using different feature selection techniques to reduce the

irrelevant and redundant features where there are huge amount of dataset and improve the classification accuracy of the model. Some authors have also used various evolutionary computation (EC) techniques for effectively solving feature selection problems which utilize the optimization process. The examples of these include genetic algorithms (GAs) [20],[21],[22] genetic programming (GP) [23], particle swarm optimization (PSO) algorithms [24], ant colony optimization (ACO) [25], harmony search [26], simulated annealing (SA) [27], [28] and differential evolution (DE) [29].These techniques are well-known and are important for their good global search ability.

The main motive of this paper is to reduce the dimension of the dataset to achieve the better classification accuracy with minimal number of features. In this paper we have proposed Filter-GA based approach for feature selection called GAFFS method. We chose new five popular filter based feature ranking and feature selection technique such as Information gain (InfoGainAttributeEval), Gain ratio (GainRatioAttributeEval), ReliefF (ReliefFAttributeEval), Chi-square (ChiSquaredAttributeEval) and Correlation feature selector (cfsSubsetEval) for selecting most relevant attributes from real world datasets using weka (Waikato Environment for Knowledge Analysis) [30] software platform which selects the top rank features based on its important and again Genetic algorithm (GA) is used for feature selection while fitness of the chromosomes taken as classification accuracy obtained by the classifier such as K-nearest neighbor (KNN) are used as a base classifier. Some standard real world datasets are selected to evaluate the proposed algorithm. The simulation results show that the proposed algorithm is superior in terms of both classification accuracy and smaller number of features.

The rest of this paper is organized as follows: Section II shows the literature review part of the Classification, Feature selection with evolutionary algorithm related Works, Section III presents preliminaries for the Basic Concepts of Genetic algorithm, Information Gain, Gain Ratio, ReliefF, Chi-Squared, Correlation feature selector (CFS) method and K-nearest neighbor (KNN) method. In Section IV proposed model is explained by algorithm and model. The datasets on which experiments are performed are shown in Section V. Section VI explains the experiments performed on the datasets with proposed model. In Section VII explanation and discussion of

_____

_____

obtained results are provided. In the last section VIII, conclusion of the whole paper is presented with future scope.

II. RELATED WORK

Recently, many authors have used different techniques for classification of data in which evolutionary algorithms are popularly used. They also used different feature selection techniques to reduce the number of features and enhance the efficiency performance of the models. A. Unler, A. Murat, [31] have proposed a discrete particle swarm optimization approach for feature selection in binary classification problems. L.-Yeh Chuang et al., [32] have presented catfish binary particle swarm optimization (CatfishBPSO) algorithm in which the catfish effect is applied to perform feature selection and improve the performance of binary particle swarm optimization (BPSO). The K-nearest neighbor (KNN) methods with Leave-one-out cross validation (LOOCV) was used to evaluate the quality of the solutions. CatfishBPSO was applied and compared to 10 classification problems taken from the literature. L. Chuang et al., [33] have presented a hybrid algorithm (CBPSOL) for selecting optimal feature subsets efficiently. This algorithm is based on CBPSO and local search. The 1-nearest neighbor (1-NN) method with leave-one-out cross-validation as a classier is used for evaluating classification accuracies. B. Sahu, D. Mishra, [34] have proposed a novel feature selection approach for the classification of high dimensional cancer microarray data, which used filtering technique such as signal to noise ratio (SNR) score and optimization technique as Particle swarm Optimization (PSO). K-nearest neighbor (KNN), Probabilistic Neural Network (PNN) and Support vector machines (SVM) are used as evaluators and leave one out cross validation approach is used for validation. M. Susana et al., [35] have proposed a modified binary particle swarm optimization (MBPSO) approach for feature selection with the simultaneous optimization of SVM kernel parameter setting, applied to mortality prediction in septic patients. H. Banka, S. Dara, [36] have presented a Hamming distance based binary PSO algorithm for feature selection and classification in gene expression data. The experimental results validate that the proposed HDBPSO performs better using Hamming distance as proximity measure for this problem. The experimental results on three benchmark datasets vis-á-vis colon cancer, defused B-cell lymphoma and leukemia data are evaluated by means of classification accuracies and validity indices as well. Indriyani et al., [9] have proposed a feature selection strategy based on Naive Bayes Multinomial (NBM), Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) approach for Arabic Document Classification. G. Haixing, [37] have proposed a novel ensemble algorithm named of BPSO-Adaboost-KNN, which is designed to solve multiple class imbalanced data problems. This model uses BPSO to select key features of datasets so that the classifier can ignore more noise. Considering traditional classifiers gain a poor performance when facing imbalanced data, generate the Adaboost-KNN classifier by using boosting-by-resample strategy. Another contribution is in employing a novel measure AUC area as the criteria for selecting optimal sub-feature set. M. Masoud Javidi et al., [10] have proposed a wrapper feature selection algorithm for classification that is based on chaos theory, binary particle swarm optimization, and local search. In the proposed algorithm, the nearest neighbor algorithm is used for the evaluation phase. A. Moayedikia et al., [26] have introduced a novel feature selection approach called SYMON which uses symmetrical uncertainty and harmony search for high dimensional imbalanced class datasets. SYMON has a two stage algorithm, the first stage, feature weighting, measures the features' weights (or importance). In the second stage, known as feature selection, the top k features are selected based on their weights. W. Srisukkham et al., [38] proposed two modified BBPSO algorithms for feature optimization to enhance intelligent decision support system for acute lymphoblastic leukaemia (ALL) classification. S. Kar et al., [39] proposed a PSO adaptive KNN based gene selection technique to distinguish a small subset of useful genes that are sufficient for the desired classification purpose of microarray data and also proposed a heuristic for selecting the optimal values of K efficiently, guided by the classification accuracy. This proposed technique of finding minimum possible meaningful set of genes is applied on three benchmark microarray datasets, namely the small round blue cell tumor (SRBCT) data, the acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) data and the mixed-lineage leukemia (MLL) data. A. Saxena et al., [40] have proposed four approaches for feature selection in an unsupervised manner by using genetic algorithms. These methods do not use the class label information but select a set of features using a task independent criterion that can preserve the geometric structure (topology) of the original data in the reduced feature space. These approaches are tested on six real data sets with dimensionality varying between 9 and 60. The selected features are found to be excellent in terms of preservation topology (inter-point geometry), cluster structure and classifier performance.

III. PRELIMINARIES

A. Genetic Algorithm

It was developed and first proposed by John Holland in 1975 [41]. It is robust and stochastic search method with a large amount of implicit parallelism. GA is based on the principle of natural genetics and the evolutionary theory of genes. The algorithm starts by initializing a population of potential solutions encoded into string called chromosomes. Each solution has some fitness value based on which the fittest parents that would be used for reproduction are found (survival of the fittest). The new generation is created by applying genetic operators such as selection (based on natural selection to create the mating pool), crossover (exchange of information among parents) and mutation (sudden small change in a parent) on selected parent's .Thus the quality of the population is improved as the number of generation's increases. The process continues until some specific criterion is met or the solution convergence to some optimized value [42]. Genetic algorithm has basic three operators:

1) Selection: This operator is responsible for selection of parents for creation of new offspring. It mimics the process of natural selection and the survival of the fittest of Darwinian evolution theory. In these processes, an intermediate population, called mating pool, is generated by copying the chromosomes from the parent population. Usually, the number of copies a chromosome receives in the mating pool is taken to be directly proportional to its fitness value. Only the selected chromosomes in the mating pool take part in the subsequent genetic operations like crossover To change the default, adjust the template as follows. and mutation. Among the several available selection methods, roulette wheel selection, stochastic

_____

_____

universal sampling and binary tournament selection are three widely used techniques [42].

2) Crossover: crossover is one of the main genetic operators that combine (mates) two chromosomes (parents) to produce a new chromosome (offspring). The idea behind crossover is that the new chromosomes may be better than both parents if they take the best characteristics from each of the parents. Crossover occurs during evolution according to a user-definable crossover probability. Some popular crossover methods are single – point crossover, two – point crossover and uniform crossover [42].

3) Mutation: mutation is a genetic operator that alters one or more gene values in a chromosome from its initial state. This can result in entirely new gene values being added to the gene pool. With these new gene values, the Genetic Algorithm may be able to arrive at a better solution than was previously possible. Mutation is an important part of the genetic search as it helps to prevent the population from stagnating at any local optimum. Mutation occurs during evolution according to a user definable mutation probability. This probability should usually be set fairly low (0.01 is a good first choice). If it is set too high, the search will turn into a primitive random search. A commonly used mutation operator for binary chromosomes is bit-flip mutation (i.e. convert 0 to 1 or vice versa), where each bit of a chromosome is subjected to mutation with the mutation probability and if the bit is selected to be mutated, it is just flipped [42], [43]. A more complete description about Genetic Algorithm can be found in [43].

B. Information Gain

The Information Gain filter is one of the most popular univariate methods of evaluating attributes. This filter evaluates the features according to their Information Gain and consults a single feature at a time. It provides an orderly classification of all the features, and then a threshold is required to select a certain number of them according to the order obtained [44], [45], [46].

C. Gain Ratio

Gain ratio aims to maximize the information gain of feature and minimize the number of its value. Gain ratio is the ratio between the information gain and intrinsic value defined as the following equation (4).

$$\text{Gain Ratio} = \frac{\Delta_{info}}{-\sum_{j=1}^{k} \frac{|C_v|}{|C|} \log_2 \frac{|C_v|}{|C|}} \quad (1)$$

Where C is a collection of samples and Cv is the subset of collection C for attribute which has the value of v. k is the number of attribute values, information gain of features

$$\Delta_{info} = \text{Entropy}(C) - \sum_{j=1}^{k} \frac{|C_v|}{|C|} \text{Entropy } c_v. \text{ [45], [47].}$$

D. ReliefF

The filter ReliefF is an extension of the original Relief algorithm. The original Relief works by randomly sampling an instance from the data and then locating its nearest neighbor from the same and opposite classes. The values of the attributes of the nearest neighbors are compared to the sampled instance and used to update relevance scores for each attribute. The rationale is the useful attribute should differentiate between instances from different classes and have the same value for

instances from the same class. ReliefF adds the ability of dealing with multiclass problems and is also more robust and capable of dealing with incomplete and noisy data. This approach may be applied in all situations that have low bias, includes interaction among features and may capture local dependencies which other method miss [5], [46], [48].

E. Chi-Squared

Chi-Squared attribute evaluation evaluates a feature by computing the chi-squared statistic of the feature with respect to the class label. First the hypothesis H0 is assumed as the two features are unrelated. Then it is tested using the following equation (5).

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2)$$

Where O_{ij} is the observed frequency and E_{ij} is the expected (theoretical) frequency, the larger the value of X^2 is, the more evidence to show that the hypothesis H_0 is true, [45], [49].

F. Correlation based feature selection (CFS)

Correlation feature selection (CFS) is a simple type of multivariate filter algorithm that ranks feature subsets according to a correlation based heuristic evaluation function. The bias of the evaluation function is toward data subsets that consist of features that are highly correlated with the class and uncorrelated with each other, so redundant and irrelevant features should be screened out [4], [45], [46].

G. K-nearest neighbour (KNN) method

K-nearest neighbor (KNN) are used as a classifier for classification purpose in machine learning, data mining and pattern recognition, which use distance metrics to predict those classes of instances that still are not seen. It is also known as "instance based learning" and "lazy learner". In this approach an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (where k is some user specified constant). If k=1 then the object is simply assigned to the class of the single nearest neighbor. This approach is more suitable for numerical data; also it can deal with discrete value [46].

IV. PROPOSED METHOD

In this paper, we have proposed the Filter-GA based approach to feature selection for classification method with Leave-one-out cross validation (LOOCV) [50], [51] to improve the classification accuracy of supervised dataset i.e. the datasets that contain classes. In both, the training set and the test set, K-nearest neighbor (KNN) technique with n fold cross-validation is employed to evaluate the classification accuracy. Firstly we use some popular ranking based filter attribute selection method, search for new informative features already available in Weka data mining tools such as Information Gain, Gain Ratio, ReliefF, Chi-Squared, Correlation Feature Selection (CFS) and secondly we apply Genetic Algorithm for feature selection with KNN classification evaluator. In this process, we have first used the features of 28 datasets of UCI and KEEL repository then ranked the features of datasets using above five attribute selection methods. We have selected the most important top ranked features (like 10%, 20% or so on) of the dataset and combined these selected features of dataset in one set. Hence we find best new features sub set (a reduced feature set) by using different ranking based feature method. We again

_____

_____

apply the GA feature selection technique on the new optimized set to achieve the better classification accuracy with even smaller feature subset. We use genetic algorithm for feature selection. In this process, first initialize randomly each chromosome by filling 0 or 1 in its genes to represent presence of a feature (1) or absence of a feature (0). Number of genes (say n) in each chromosome is number of features in dataset. The value of each gene is checked if it is 1 then corresponding feature is collected and kept in an array. In this manner a subset of features is obtained. Now this subset of features makes a reduced dataset and its goodness is checked by using K-nearest neighbor (KNN) classifier with LOOCV techniques. This process is repeated for every chromosome in the population. The best chromosome (chromosome that gives best classification accuracy) is retained after running GA for a given number of generations using selection, crossover and mutation operations or when satisfactory classification accuracy is obtained. In this manner a subset of features is obtained. Now this subset of features makes a reduced dataset and its goodness is checked using K-nearest neighbor (KNN) classifier with LOOCV techniques. This process is repeated for every chromosome in the population. The best chromosome (i.e. a chromosome that gives best classification accuracy) is retained after running GA for a given number of generations using selection, crossover and mutation operations or when satisfactory classification accuracy is obtained. Figure 1 shows the complete algorithm for the proposed method and Figure 2 shows model for the proposed method.

$$Accuracy = \frac{Number\ of\ samples\ correctly\ classified\ in\ test\ data}{Total\ number\ of\ samples\ in\ the\ test\ data} \times 100\% \qquad (3)$$

- Get data set
- D [m, n] (Where m number of patterns, n number of features)

// use filter based attribute selection measure (Information Gain, Gain Ratio, ReliefF, Chi-Squared, Correlation based feature selection(CFS)

- Combined feature set D [m, n] (Where m number of patterns, n number of features)

// use Genetic Algorithm for feature selection and K-NN as base classifier to evaluate goodness of each chromosomes

- Initialize chromosomes randomly with binary value at each genes

 (number of genes is equal to number of features)

For 1: number of iteration

 a. Calculate fitness of each chromosome using KNN classifier (With LOOCV cross validation technique)
 b. Create new child using three operators :
 - Selection
 - Crossover
 - Mutation

End

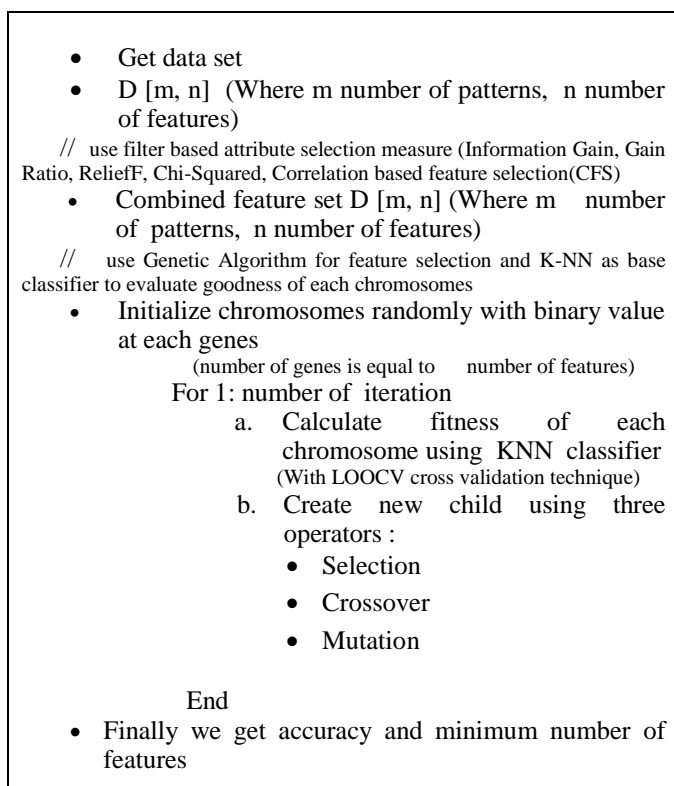- Finally we get accuracy and minimum number of features

Figure 1. Algorithm for proposed method

V. DATASETES

We performed experiments on the real life datasets with all features listed with brief summary given in Table I. The

datasets are collected from two repositories, namely the University of California, Irvine (UCI) Machine Learning repository [52] and the KEEL dataset repository [53]. Table I. contains serial number (S.No.), Dataset name, total number of features (excluding class attributes), total number of instances (records or rows), Missing values, total number of Classes and the corresponding repository. Table I. has Twenty eight datasets namely Audiology, Dermatology, Spambase, Arrhythmia, Ionosphere etc. For each dataset, we carried out a Leave one out cross validation. The outline of these databases can be seen in Table I.
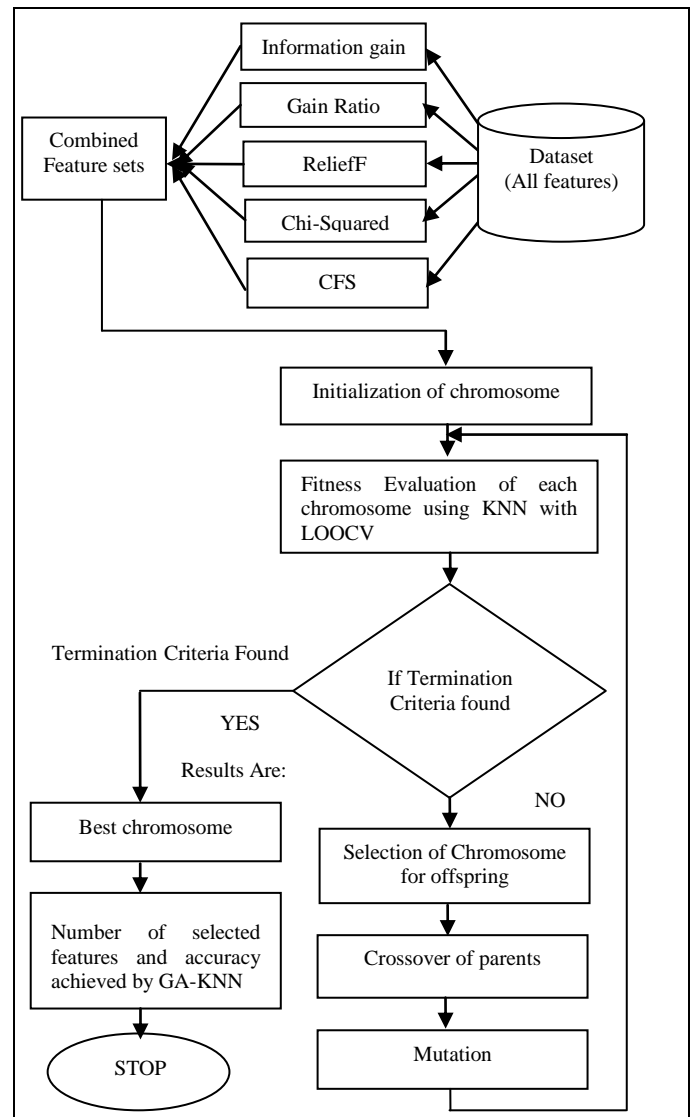

Figure 2. Model of proposed method

VI. EXPERIMENTS

We performed our experiments on Intel i5 Processor with 4 GB RAM and 500 GB hard disk. Matlab is used for development of code and experiment. Also GA Matlab code [54] is used for this study. In this paper first we have used some standard attribute selection measure such as Information Gain, GainRatio, ReliefF, Chi-square, Correlation based attribute evaluator (CFS) to predict the most relevant features available in the Weka Software version 3.6 .12 [30],[55]. We have converted the datasets into csv file then this file was loaded into

_____

Weka explorer. We used leave-one-out cross validation technique to obtain robust classification accuracy. In this method the whole reduced dataset (combined features set) is decomposed in ten folds each having equal number of patterns (the last fold having remaining number of patterns if number of patterns is not a multiple of 10). One of these ten folds is used for testing while all other are used in training phase. Average of accuracy obtained on each fold gives accuracy of our model. In this experiment we performed KNN classifier with Genetic Algorithm based experiment.

VII. RESULTS AND DISCUSSION

The proposed method is applied over the datasets listed in the Table I. Results of the experiment are displayed in the Table II. Table II has four columns. As shown in Table II, for GAFFS method, Audiology dataset reduces to 5 relevant features out of 69 features and classification accuracy is 71.50. Dermatology dataset has 10 relevant features out of total 34 numbers of features and classification accuracy is 97.21. Spambase dataset has 20 selected features out of 57 numbers of features and classification accuracy is 85.75. Arrhythmia dataset has 14 selected features out of 279 features and classification accuracy is 62.39. Ionosphere dataset has 6 selected features out of 33 features and classification accuracy is 94.87. Image Segmentation dataset has 6 selected features out of 19 features and classification accuracy is 91.43. Hepatitis dataset has 6 selected features out of 19 features and classification accuracy is 92.50. Lung cancer dataset has 7 selected features out of 56 features and classification accuracy is 78.13. Breast cancer dataset has 4 selected features out of 9 features and classification accuracy is 78.70. German credit dataset has 8 selected features out of 20 features and classification accuracy is 70.10. Lymphography dataset has 8 selected features out of 18 features and classification accuracy is 83.11. Mushroom dataset has 6 selected features out of 22 features and classification accuracy is 98.28. Vehicle dataset has 8 selected features out of 18 features and classification accuracy is 72.22. Vote dataset has 1 selected feature out of 16 features and classification accuracy is 95.63. Breast-w dataset has 5 selected features out of 31 features and classification accuracy is 93.50. Cmc dataset has 7 selected features out of 9 features and classification accuracy is 48.68. Car dataset has 3 selected features out of 6 features and classification accuracy is 72.69. Credit-Australian dataset has 2 selected features out of 14 features and classification accuracy is 85.51. Ecoli dataset has 6 selected features out of 7 features and classification accuracy is 79.17. Glass dataset has 6 selected features out of 9 features and classification accuracy is 74.30. Hill-valley dataset has 16 selected features out of 100 features and classification accuracy is 63.04. Heart-c dataset has 3 selected features out of 13 features and classification accuracy is 56.90. Hayes-roth dataset has 3 selected features out of 4 features and classification accuracy is 84.38. Iris dataset has 2 selected features out of 4 features and classification accuracy is 95.33. Liver-disorder dataset has 2 selected features out of 6 features and classification accuracy is 65.80. Pima dataset has 5 selected features out of 8 features and classification accuracy is 70.57. Tic-Toc-Toe dataset has 5 selected features out of 9 features and classification accuracy is 81.63. tae dataset has 3 selected features out of 5 features and classification accuracy is 64.90.

In Table III we compared results obtained by proposed method with the results obtained by other methods. Table III

has eight columns, First column contains S.No., Second column contains dataset name, third column contains total number of features, Fourth column divided in two column which contains accuracy and number of features obtained by ACOFSS+mRMR method [56]. Fifth column divided in two columns which contains accuracy and number of features obtained by DTRSM+PSO method [57]. Sixth column contains accuracy obtained by AM (mbc) method [58]; Seventh column divided in two columns which contain accuracy and number of features obtained by PSOFFS method [59]. Eight columns are also divided in two columns for accuracy and number of features obtained by proposed method GAFFS. In case of Audiology dataset, the classification accuracy obtained by the proposed method GAFFS is 71.50 (5) and it is compared with accuracy 86.6 (38) obtained by ACOFSS+mRMR proposed by A. Khan & A. Rauf Baig [56], accuracy 73.28 (16) obtained by DTRSM+PSO proposed by S. Chebrolu, G. Sriram Sanjeevi [57] and accuracy 71.50 (6) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets () shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Dermatology dataset, the classification accuracy obtained by the proposed method GAFFS is 97.21 (10) and it is compared with accuracy 97.3(20) obtained by ACOFSS+mRMR method proposed by A. Khan & A. Rauf Baig [56], accuracy 97.21 (23) obtained by DTRSM+PSO methods proposed by S. Chebrolu, G. Sriram Sanjeevi [57], accuracy 85.31 () obtained by AM (mbc) methods proposed by Z. Liang et al. [58] and accuracy 97.21 (12) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets () shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In the case of Spambase dataset the classification accuracy obtained by the proposed method GAFFS is 85.75 (20) and it is compared with accuracy 91.4 (37) obtained by ACOFSS+mRMR method proposed by A. Khan & A. Rauf Baig [56] and accuracy 91.92 (18) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In the case of Arrhythmia dataset the classification accuracy obtained by the proposed method GAFFS is 62.39 (14) and it is compared with accuracy 79.2 (120) obtained by ACOFSS+mRMR method proposed by A. Khan & A. Rauf Baig [56] and accuracy 61.50 (16) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets () shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Ionosphere dataset the classification accuracy obtained by the proposed method GAFFS is 94.87 (6) and it is compared with accuracy 93.2 (14) obtained by ACOFSS+mRMR method proposed by A. Khan & A. Rauf Baig, [56] , accuracy 92.42 obtained by AM (mbc) methods proposed by Z. Liang et al. [58] and accuracy 93.73 (8) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Image segmentation dataset the classification accuracy obtained by the proposed method GAFFS is 91.43 (6) and it is compared with accuracy 89.1 (9) obtained by ACOFSS+mRMR method proposed by A khan & A. Rauf Baig [56], and accuracy 90.95 (4) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets () shows number of features used to

206

_____

TABLE I.   A Brief Description of the Datasets Used in This Experiment

S. No.	Dataset Name	NOF	NOI	MV	NOC	Data Source
1	Audiology	69	226	No	24	UCI
2	Dermatology	34	358(366)	Yes	6	Keel
3	Spambase	57	4597	No	2	Keel
4	Arrhythmia	279	452	No	16	UCI
5	Ionosphere	33	351	No	2	Keel
6	Image Segmentation	19	2310	No	7	UCI
7	Hepatitis	19	80(155)	Yes	2	Keel
8	Lung cancer	56	32	No	3	UCI
9	Breast cancer	9	277(286)	Yes	2	Keel
10	German credit	20	1000	No	2	Keel
11	Lymphography	18	148	No	4	Keel
12	Mushroom	22	5644(8124)	Yes	2	Keel
13	Vehicle	18	846	No	4	Keel
14	Vote	16	232(435)	Yes	2	Keel
15	Breast-w	31	569	No	2	UCI
16	Cmc	9	1473	No	3	Keel
17	Car	6	1728	No	4	Keel
18	Credit-australian	14	690	No	2	Keel
19	Ecoli	7	336	No	8	Keel
20	Glass	9	214	No	7	Keel
21	Hill-valley	100	606	Yes	2	UCI
22	Heart-c	13	297(303)	No	5	Keel
23	Hayes-roth	4	160	No	3	Keel
24	Iris	4	150	No	3	Keel
25	Liver-disorder	6	345	No	2	Keel
26	Pima	8	768	No	2	Keel
27	Tic-Tac-Toe	9	958	No	2	Keel
28	Tae	5	151	No	3	Keel

NOF: Number of features, NOI: Number of instances, MV: Missing Value in datasets, NOC: Number of class

obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features.In case of hepatitis dataset the classification accuracy obtained by the proposed method GAFFS is 92.50 (6) and it is compared with accuracy 90.3 (8) obtained by ACOFSS+mRMR method proposed by A. Khan & A. Rauf Baig [56], accuracy 83.99 (5) obtained by DTRSM+PSO methods proposed by S. Chebrolu, G. Sriram Sanjeevi [57], accuracy 81.12 obtained by AM (mbc) methods proposed by Z. Liang et al. [58] and accuracy 92.50 (5) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Lung cancer dataset the classification accuracy obtained by the proposed method GAFFS is 78.13 (7) and it is compared with accuracy 88.9 (24) obtained by ACOFSS+mRMR method proposed by A. Khan & A. Rauf Baig [56] and accuracy 82.73 (9) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Breast cancer dataset the classification accuracy obtained by the proposed method GAFFS is 78.70 (4) and it is compared with accuracy 72.27 (4) obtained by DTRSM+PSO methods proposed by S. Chebrolu, G. Sriram Sanjeevi [57], accuracy 74.02 obtained by AM (mbc) methods proposed by Z. Liang et al. [58] and accuracy 78.70 (4) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of German credit dataset the classification accuracy obtained by the proposed method GAFFS is 70.10 (8) and it is compared with

accuracy 75.29 (6) obtained by DTRSM+PSO methods proposed by S. Chebrolu, G. Sriram Sanjeevi [57] and accuracy 70.70 (9) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Lymphography dataset, the classification accuracy obtained by the proposed method GAFFS is 83.11 (8) and it is compared with accuracy 85.09 (12) obtained by DTRSM+PSO methods proposed by S. Chebrolu, G. Sriram Sanjeevi [57] and accuracy 82.43 (9) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Mushroom dataset, the classification accuracy obtained by the proposed method GAFFS is 98.28 (6) and it is compared with accuracy 98.88 (3) obtained by DTRSM+PSO methods proposed by S. Chebrolu, G. Sriram Sanjeevi [57] and accuracy 100 (18) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Vehicle dataset, the classification accuracy obtained by the proposed method GAFFS is 72.22 (8) and it is compared with accuracy 57.48 (7) obtained by DTRSM+PSO methods proposed by S. Chebrolu, G. Sriram Sanjeevi [57] and accuracy 72.22 (6) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy   obtained   using   r_n   number   of

TABLE II.   CLASSIFICATION ACCURACY (IN PERCENTAGE) AND NUMBER OF FEATURES SELECTED BY RESPECTIVE MODEL FOR PROPOSED METHOD (GAFFS)

S. No.	Dataset Name	NOF	Proposed (GAFFS) Method	
			ACC	*NORF*
1	Audiology	69	71.50	5
2	Dermatology	34	97.21	10
3	Spambase	57	85.75	20
4	Arrhythmia	279	62.39	14
5	Ionosphere	33	94.87	6
6	Image Segmentation	19	91.43	6
7	Hepatitis	19	92.50	6
8	Lung cancer	56	78.13	7
9	Breast cancer	9	78.70	4
10	German credit	20	70.10	8
11	Lymphography	18	83.11	8
12	Mushroom	22	98.28	6
13	Vehicle	18	72.22	8
14	Vote	16	95.63	1
15	Breast-w (wdbc)	31	93.50	5
16	Cmc	9	48.68	7
17	Car	6	72.69	3
18	Credit-australian	14	85.51	2
19	Ecoli	7	79.17	6
20	Glass	9	74.30	6
21	Hill-valley	100	63.04	16
22	Heart-c	13	56.90	3
23	Hayes roth	4	84.38	3
24	Iris	4	95.33	2
25	Liver-disorder	6	65.80	2
26	Pima	8	70.57	5
27	Tic-Tac-Toe	9	81.63	5
28	Tae	5	64.90	3

NOF: Number of features, ACC: Accuracy, NORF: No. of Reduced features

features. In case of vote dataset, the classification accuracy obtained by the proposed method GAFFS is 95.63 (1) and it is compared with accuracy 92.85 (4) obtained by DTRSM+PSO methods proposed by S. Chebrolu, G. Sriram Sanjeevi [57] and accuracy 95.17 (3) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features .In case of Breast-w dataset the classification accuracy obtained by the proposed method GAFFS is 93.50 (5) and it is compared with accuracy 94.73 obtained by AM (mbc) methods proposed by Z. Liang et al. [58] and accuracy 93.50 (5) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Cmc dataset, the classification accuracy obtained by the proposed method GAFFS is 48.68 (7) and it is compared with accuracy 47.07 ( ) obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 48.54 (6) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of car dataset, the classification accuracy obtained by the proposed method GAFFS is 72.69 (3) and it is compared with accuracy 77.26 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 72.69 (3) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Credit-australian dataset the classification accuracy obtained by the proposed method GAFFS is 85.51 (2) and it is

compared with accuracy 86.37 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 85.51 (1) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of E-coli dataset the classification accuracy obtained by the proposed method GAFFS is 79.17 (6) and it is compared with accuracy 79.14 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 79.97 (6) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Glass dataset, the classification accuracy obtained by the proposed method GAFFS is 74.30 (6) and it is compared with accuracy 96.36 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 74.30 (6) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Hill-valley dataset, the classification accuracy obtained by the proposed method GAFFS is 63.04 (16) and it is compared with accuracy 52.12 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 62.38 (22) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Heart-c dataset, the classification accuracy obtained by the proposed method GAFFS is 56.90 (3) and it is compared with accuracy 54.28 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 56.23 (3)

_____

TABLE III.   COMPARISON OF ACCURACY OBTAINED BY DIFFERENT METHODS

S. No.	Dataset Name	Total NOF	ACOFSS+mRMR Method		DTRSM+PSO Method		AM(mbc) Method	PSOFFS Method		Proposed (GAFFS) Method	
			ACC	*NOF*	*ACC*	*NOF*	*ACC*	*ACC*	*NOF*	*ACC*	*NOF*
1	Audiology	69	86.6	38	73.28	16	-	71.50	6	**71.50**	5
2	Dermatology	34	97.3	20	97.21	23	85.31	97.21	12	**97.21**	10
3	Spambase	57	91.4	37	-	-	-	91.92	18	85.75	20
4	Arrhythmia	279	79.2	120	-	-	-	61.50	16	**62.39**	14
5	Ionosphere	33	93.2	14	-	-	92.42	93.73	8	**94.87**	6
6	Image Segmentation	19	89.1	9	-	-	-	90.95	4	**91.43**	6
7	Hepatitis	19	90.3	8	83.99	5	81.12	92.50	5	**92.50**	6
8	Lung cancer	56	88.9	24	-	-	-	82.43	9	78.13	7
9	Breast cancer	9	-	-	72.27	4	74.02	78.70	4	**78.70**	4
10	German credit	20	-	-	75.29	6	-	70.70	9	70.10	8
11	Lymphography	18	-	-	85.09	12	-	82.43	9	**83.11**	8
12	Mushroom	22	-	-	98.88	3	-	100	18	98.28	6
13	Vehicle	18	-	-	57.48	7	-	72.22	6	**72.22**	8
14	Vote	16		-	92.85	4		95.17	3	**95.63**	1
15	Breast-w	31	-	-	-	-	94.73	93.50	5	**93.50**	5
16	Cmc	9	-	-	-	-	47.07	48.54	6	**48.68**	7
17	Car	6	-	-	-	-	77.26	72.69	3	**72.69**	3
18	Credit-Australian	14	-	-	-	-	86.37	85.51	1	**85.51**	2
19	E-coli	7	-	-	-	-	79.14	79.97	6	79.17	6
20	Glass	9	-	-	-	-	96.36	74.30	6	**74.30**	6
21	Hill-valley	100	-	-	-	-	52.12	62.38	22	**63.04**	16
22	Heart-c	13	-	-	-	-	54.28	56.23	3	**56.90**	3
23	Hayes -Roth	4	-	-	-	-	82.88	84.38	3	**84.38**	3
24	Iris	4	-	-	-	-	88.87	95.33	3	**95.33**	2
25	Liver-disorder	6	-	-	-	-	46.93	70.02	3	65.80	2
26	Pima	8	-	-	-	-	74.81	70.57	4	69.79	5
27	Tic-Tac-Toe	9	-	-	-	-	98.19	82.15	6	81.63	5
28	tae	5	-	-	-	-	32.12	50.1	2	**64.90**	3

NOF: Number of features, ACC: Accuracy and The bold value denote the highest and similar value of proposed method

obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Hayes roth dataset, the classification accuracy obtained by the proposed method GAFFS is 84.38 (3) and it is compared with accuracy 82.88 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 84.38 (3) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Iris dataset, the classification accuracy obtained by the proposed method GAFFS is 95.33 (2) and it is compared with accuracy 88.87 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 95.33 (3) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Liver-disorder dataset the classification accuracy obtained by the proposed method GAFFS is 65.80 (2) and it is compared with accuracy 46.93 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 70.02 (3) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Pima dataset, the classification accuracy obtained by the proposed method GAFFS is 69.79 (5) and it is compared

with accuracy 74.81 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 70.57 (4) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of Tic-Tac-Toe dataset, the classification accuracy obtained by the proposed method GAFFS is 81.63 (5) and it is compared with accuracy 98.19 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 82.18 (6) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features. In case of tae dataset, the classification accuracy obtained by the proposed method GAFFS is 64.90 (3) and it is compared with accuracy 32.12 obtained by AM (mbc) methods proposed by Z. Liang et al., [58] and accuracy 50.1 (2) obtained by PSOFFS proposed by A. Saxena, et al. [59] where values in brackets ( ) shows number of features used to obtain this accuracy i.e. c_a (r_n) means classification accuracy obtained using r_n number of features.

It is evident that the classification accuracy obtained by proposed method on twenty eight datasets is better than that obtained by other methods as shown in Table III. The results by the proposed method are shown as par with the others in the graph chart as well in Figure 3, Figure 4, Figure 5, Figure 6 and Figure 7. The classification accuracy obtained by proposed method is shown by Figure 3 in a green bar and it is taller in each of the eight dataset compared

209

_____

_____

to other methods shown by blue (ACOFSS+mRMR method) color. The classification accuracy obtained by proposed method is shown by Figure 4 in a green bar and it is taller in each of the nine dataset compared to other methods shown by red (DTRSM+PSO method) colors. The classification accuracy obtained by proposed method is shown by Figure 5 in a green bar and it is taller in each of the eighteenth dataset compared to other methods shown by purpal (AM (mbc) method) colors. The best classification accuracy obtained by proposed method is shown by Figure 6 in a green bar and it is taller in each of the twenty-eighth datasets compared to other methods shown by orange (PSOFFS method) colors. The best classification accuracy obtained by proposed method is shown by Figure 7 in a green bar and it is taller in each of the twenty-eighth datasets compared to other methods shown by blue (ACOFSS+mRMR method), red (DTRSM+PSO method), purple (AM (mbc) method) ,and orange (PSOFFS method) colors.
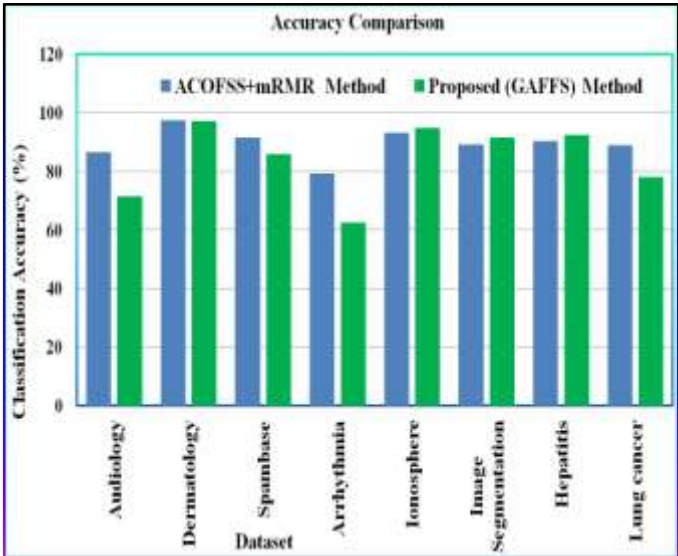
Figure 3.   Graphical representation for Comparison of accuracy obtained by ACOFSS+mRMR and GAFFS
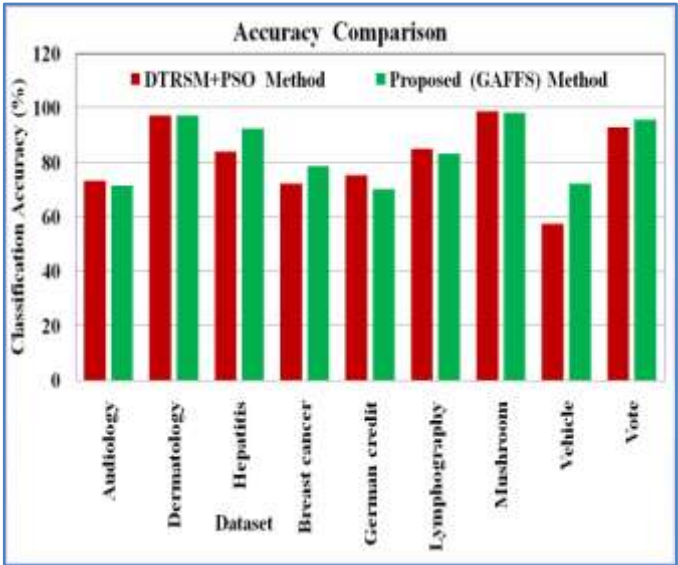
Figure 4.   Graphical representation for Comparison of accuracy obtained by DTRSM+PSO and GAFFS
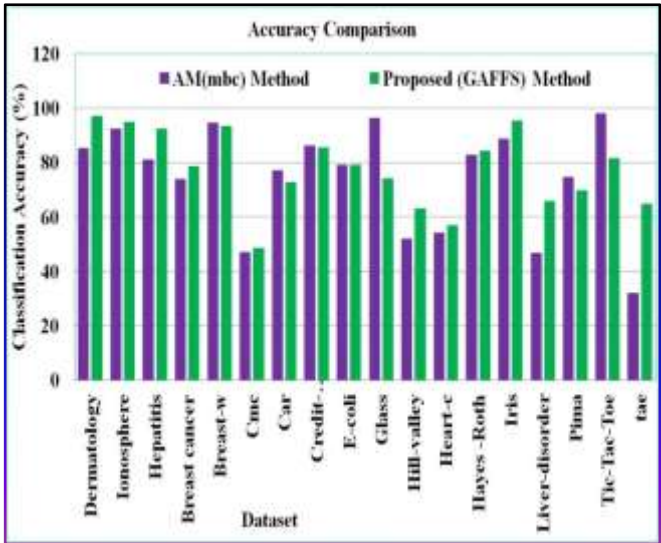

Figure 5.   Graphical representation for Comparison of accuracy obtained by AM(mbc) and GAFFS
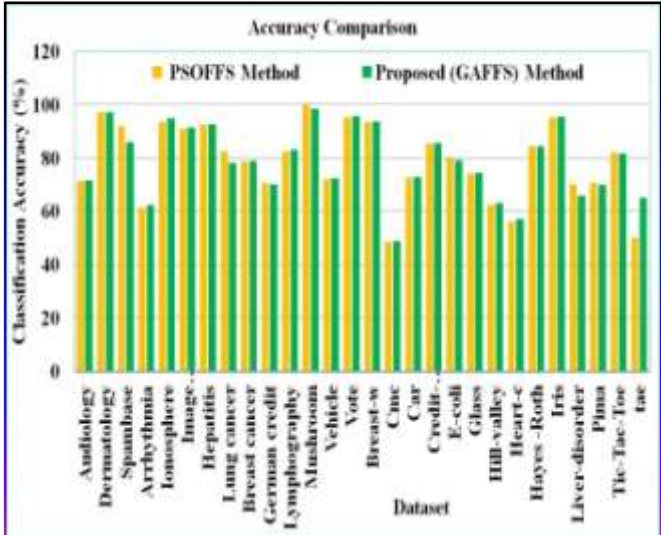

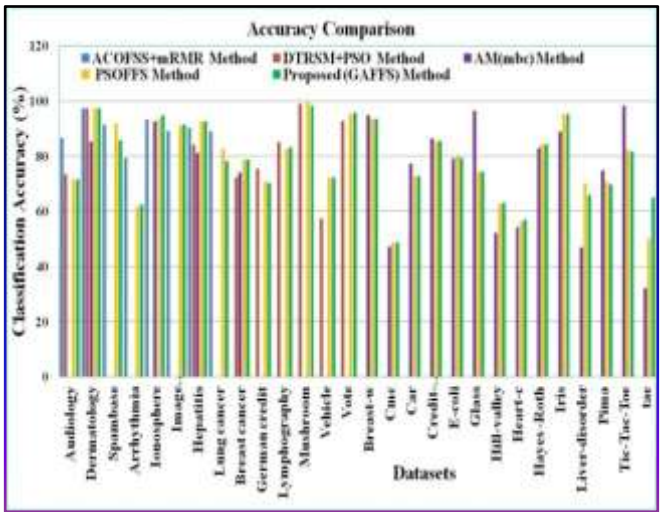Figure 6. Graphical representation for Comparison of accuracy obtained by PSOFFSand GAFFS

Figure 7.   Graphical representation for Comparison of accuracy obtained by ACOFSS+mRMR, DTRSM+PSO, AM(mbc) , PSOFFS and GAFFS

210

_____

VIII. CONCLUSION

In this paper, a new approach is presented to select small number of features from various databases. The leave-one-out cross validation (LOOCV) approach checks the performance of subset of features by taking one folder for testing due to cross validation. In this manner the performance of subset of features on each pattern is evaluated. The Audiology, Dermatology, Spambase, Arrhythmia, Ionosphere, Image Segmentation, Hepatitis, Lung cancer, Breast cancer, German credit, Lymphography, Mushroom, Vehicle, Housevotes (Vote), Breast-w, contraceptive (Cmc), Car, Credit-australian, E-coli, Glass, Hill-valley, Heart-c, Hayes-roth, Iris, Liver-disorder, Pima , Tic-Toc-Toe , Teaching Assistant Evaluation (tae), datasets are used for validation of proposed method. In each case, the classification accuracy which is taken as the measure of goodness of subset of features comes higher than the accuracy claimed by other recently reported techniques. Thus the LOOCV based feature selection method can be applied as another approach to select features. The databases used here have a moderate dimensions, it will be interesting to see the performance of the proposed method on high dimensional databases with other evolutionary algorithms, classifiers and feature ranking methods.

REFERENCES

[1] M. Raymer, W. Punch, E. Goodman, L. Kuhn, A. Jain, "Dimensionality reduction using genetic algorithms," IEEE Trans. Evol. Comput.vol.4, No. 2, pp. 164-171, July 2000.

[2] B. Rajen Bhatt, M. Gopal, "On fuzzy-rough sets approach to feature selection," Pattern Recognition Letters, vol. 26, pp. 965-975, 2005.

[3] T.Sousa, A.Silva, A.Neves, "Particle swarm based data mining algorithms for classification tasks," Parallel Computing, Vol. 30, No. 5, pp.767–783, 2004.

[4] M.A. Hall. Correlation-based feature selection for machine learning. PhD thesis, The University of Waikato, 1999.

[5] M.Robnik- Sikonja, I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," Mach. Learn. Vol. 53, pp. 23-69, 2003.

[6] M. Dash, K. Choi, P. Scheuermann, H. Liu, "Feature selection for clustering-a filter solution," in: Proceedings of the Second IEEE International Conference on Data Mining (ICDM'02), IEEE Computer Society, Washington, DC, USA, 2002, pp. 115–122.

[7] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection," J. Mach.Learn. Res. Vol. 3 pp. 1157–1182, 2003 .

[8] R. kohavi and G.h.john, "Wrappers for feature subset selection", Artificial Intelligence, vol. 97, pp. 273-324, 315-333, 1997.

[9] I., W. Gunawan, A. Rakhmadi, "Filter-Wrapper Approach to Feature Selection Using PSO-GA for Arabic Document Classification with Naive Bayes Multinomial," IOSR Journal of Computer Engineering (IOSR-JCE),Vol. 17, Issue 6, PP. 45-51,Ver. VI (Nov – Dec. 2015), e-ISSN: 2278-0661, p-ISSN: 2278-8727.

[10] M. Masoud Javidi, N. Emami, "A hybrid search method of wrapper feature selection by chaos particle swarm optimization and local search," Turkish Journal of Electrical Engineering & Computer Sciences, vol 24: pp. 3852 – 3861,2016, doi:10.3906/elk-1404-220.

[11] Jianyu Miao, Lingfeng Niu, A Survey on Feature Selection, Procedia Computer Science , vol. 91, pp. 919 – 926, 2016.

[12] J. Han, M. Kamber, "Data Mining Concepts and Techniques," Morgan Kaufmann Publishers, 2000 and 2006.

[13] P. Mitra, C.A. Murthy, S.K. Pal, "Unsupervised feature selection using feature similarity," IEEE Trans. Pattern Anal. Mach. Intell. Vol. 24 , pp.301–312, 2002 .

[14] C.L. Huang, J.F. Dun, "A distributed PSO–SVM hybrid system with features selection and parameter optimization," Appl. Soft Comput. Vol. 8, pp. 1381–1391, 2008.

[15] Y. Liu, F. Dellaert, "A classification based similarity metric for 3D image retrieval," in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp. 800–805, 1998.

[16] M.S. Lew, "Principles of Visual Information Retrieval," Springer-Verlag, London, 2001.

[17] J.S Tsai, W.B Huang, Y.H Kuo, MF Horng, Joint robustness and security enhancement for feature based image watermarking using invariant feature regions," Signal Process., vol. 92, pp. 1431-1445, 2012.

[18] SI Choi, J Oh, CH Choi, C Kim. "Input variable selection for feature extraction in classification problem, Signal Process.,vol. 92: pp. 636-648, 2012.

[19] A Miller., Subset Selection in Regression. 2nd ed. Boca Raton, FL, USA: CRC Press, 2002.

[20] D.J. Monata, L. Davis, "Training feedforward neural networks using genetic algorithms," in: 11th International Joint Conference on Artificial Intelligence, pp. 762–767, 1989.

[21] D. E. Goldberg, "Genetic Algorithms in Search," Optimization and Machine Learning. (Reading, MA, USA: Addison- Wesley, New York, 1989.

[22] A. Saxena, M. Madhaw Shrivas, "Leave one out cross validated Hybrid Model of Genetic Algorithm and Naïve Bayes for Classification Problem," Int. Journal of Engineering Research and Application, ISSN : 2248-9622, Vol. 6, pp.107-114, 2016.

[23] K. Neshatian and M. Zhang, "Improving relevance measures using genetic programming," in European Conference on Genetic Programming (EuroGP 2012), ser. Lecture Notes in Computer Science, vol. 7244. Springer, pp. 97–108, 2012.

[24] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimization for feature selection in classification: A multi-objective approach," IEEE Transactions on Cybernetics, vol. 43, no. 6, pp. 1656–1671, 2013.

[25] M.H. Aghdam, N. Ghasem-Aghaee. M.E. Basiri, "Application of ant colony optimization for feature selection in text categorization," in: Proceeding of 5th IEEE Congress on Evolutionary Computation, Hong Kong, 2008.

[26] A. Moayedikia, K.-Leong Ong, Y. Ling Boo, W. GS Yeoh, R. Jensen, "Feature selection for high dimensional imbalanced class data using harmony search," Engineering Applications of Artificial Intelligence, vol. 57, pp. 38–49, 2017.

[27] S. Shaw, W. Kinsner, "Chaotic simulated annealing in multilayer feedforward networks," in: Canadian Conference on Electrical and Computer Engineering, pp. 265–269, 1996.

[28] S.K. Chang, O.A. Mohammed, S. Y Hahn, "Detection of magnetic body using article neural network with modified simulated annealing," IEEE Trans. Magn. Vol. 30 , 3644–3647, 1994

[29] T. Si, S. Hazra, N. Jana, "Artificial neural network training using differential evolutionary algorithm for classification," in: S. Satapathy, P. Avadhani, A.Abraham (Eds.), Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012) held inVisakhapatnam, India, January 2012, Springer, Berlin/Heidelberg, AISC 132, pp. 769–778.

[30] H. Lan. Written; E. Frank, A. Hall, " Data Mining: Practical machine learning tools and techniques, 3 rd Edition", 2011.

[31] A. Unler, A. Murat, "A discrete particle swarm optimization method for feature selection in binary classification problems," European Journal of Operational Research vol. 206 , pp. 528–539, 2010

[32] L.-Yeh Chuang, S. -Wei Tsai, C. -Hong Yang , "Improved binary particle swarm optimization using catfish effect for feature selection," Expert Systems with Applications vol. 38 , pp. 12699–12707, 2011.

211

_____

[33] L. Chuang, C. Yang, J. Li, "Chaotic maps based on binary particle swarm optimization for feature selection.," Appl. Soft Compu, vol. 11, pp. 239-248, 2011.

[34] B. Sahu, D. Mishra, "A Novel Feature Selection Algorithm using Particle Swarm Optimization for Cancer Microarray Data," Procedia Engineering, vol. 38, pp. 27 – 31, 2012.

[35] M Susana, F Vieira, Luís. Mendonc, Gonc¸ alo J. Farinha, João M.C. Sousa, "Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients," Applied Soft Computing ,vol.13, pp. 3494–3504, 2013.

[36] H. Banka, S. Dara, "A Hamming distance based binary particle swarm optimization (HDBPSO) algorithm for high dimensional feature selection, classification and validation," Pattern Recognition Letters, vol. 52 , pp. 94–100, 2015

[37] G. Haixiang , L. Yijing, L. Yanan, L. Xiao, L. Jinling, "BPSO-Adaboost-KNN ensemble learning algorithm for multi-class imbalanced data classification," Engineering Applications of Artificial Intelligence, vol. 49, pp. 176–193, 2016.

[38] W. Srisukkham, L. Zhang,, S. Chin Neoh, S. Todryk, Chee Peng Lim, "Intelligent leukaemia diagnosis with bare-bones PSO based feature optimization," Applied Soft Computing, vol. 56 , 405–419, 2017.

[39] S. Kar, K. Das Sharma, M. Maitra, "Gene selection from microarray gene expression data for classification of cancer subgroups employing PSO and adaptive K-nearest neighborhood technique," Expert Systems with Applications , vol.42 , pp. 612–627, 2015.

[40] A. Saxena, N.R, Pal, M. Vora, "Evolutionary Methods for Unsupervised Feature Selection Using Sammon's Stress Function," Fuzzy Inf. Eng. , 2010, DOI 10.1007/s12543-010-0047-4.

[41] J.H. Holland, Adaptation in Nature and Artificial Systems, University of Michigan Press, Ann Arbor, MI, 1975

[42] U.Maulik, S. Bandyopadhyay, and A. Mukhopadhyay, "Multiobjective Genetic Algorithms for Clustering: Applications in Data Mining and Bioinformatics," Springer, Heidelberg, Berlin, 2011. DOI: 10.1007/978-3-642-16615-0, ISBN: 978-3-642-16614-3 (Print) [CrossRef].

[43] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning." Reading, MA, USA: Addison-Wesley, New York, 1989

[44] T. M. Cover and J. A. Thomas, Elements of information theory, 2nd ed. Hoboken, N.J.: Wiley-Interscience, 2006.

[45] L. Feifei, P. Yongjun , L. Meijing, P. Minghao, K. Ho Ryu, Positive Impression of low-ranking microrn as in human cancer classification, David C. Wyld et al. (Eds) : CCSIT, SIPP, AISC, PDCTA, NLP - 2014 pp. 449–461, 2014. © CS & IT-CSCP 2014 DOI : 10.5121/csit.2014.4239.

[46] V. Bolón-Canedo, N. Sánchez-Maroˇno, A. Alonso-Betanzos, "Distributed feature selection: An application to microarray data classification," Applied Soft Computing, vol. 30 , pp. 136–150, 2015.

[47] T. M. Mitchell, Machine learning, Boston, WCB/McGraw-Hill, 1997.

[48] K. Kira and L. A. Rendell, A practical approach to feature selection, Proceedings of the ninth international workshop on Machine learning, Aberdeen, Scotland, United Kingdom, pp. 249-256,1992.

[49] H. Liu and R. Setiono, Chi2: Feature Selection and Discretization of Numeric Attributes, Proceedings of the Seventh International Conference on Tools with Artificial Intelligence, 1995.

[50] D. Arockia Vanitha, D.D, Venkatesulu M, "Gene Expression Data Classification using support vector machine and mutual information based gene selection, "procedia computer science ,vol 47, pp. 13-21, 2015.

[51] https://www.cs.cmu.edu/~schneide/tut5/node42.html. [Online,Google search] .

[52] University of California at Irvine (UCI), Machine Learning Database, Center for Machine Learning and intelligent Systems, Online Available at :< http://archive.ics.uci.edu/ml/datasets.html>.

[53] Keel dataset repository http://sci2s.ugr.es/keel/category.php?cat=clas) ,[online available].

[54] L. Randy. Haupt,Sue Ellen Haupt, Practicle genetic Algorithms, Wiley - Interscience, A john wiley & sons,Inc., Publication, Published by John Wiley & Sons, Inc., Hoboken, New Jersey. Published simultaneously in Canada.

[55] WEKA,http://www.cs.waikato.ac.nz/ml/weka.

[56] A. Khan & A. Rauf Baig, "Multi-objective feature subset selection using mRMR based enhanced ant colony optimization algorithm (mRMR-EACO)", Journal of Experimental & Theoretical Artificial Intelligence, 3079. 1362-2015, ISSN: 0952-813X .

[57] S. Chebrolu, G. Sriram Sanjeevi, Attribute Reduction in Decision- Theoretic Rough Set Model using Particle Swarm Optimization with the threshold parameters determined using LMS training rule, Procedia Computer Science vol. 57 pp. 527-536, 2015.

[58] Z. Liang, J. Sun, Q. Lin, Z. Du, J. Chen, z. Ming, "A novel multiple rule sets data classification algorithm based on ant colony algorithm," Applied Soft Computing vol. 38, pp. 1000-1011, 2016.

[59] A. Saxena, M. Madhaw Shrivas, Filter – PSO based approach for Feature Selection, International Journal of Advanced Research in Computer Science, Volume 8, No. 5, pp. 2063-2073, May – June 2017.

_____