RESEARCH ARTICLE OPEN ACCESS

Leave one out cross validated Hybrid Model of Genetic Algorithm and Naïve Bayes for Classification Problem

Amit Saxena *, Madan Madhaw Shrivas**
*(Department of Computer Science and Information Technology (CSIT), Guru Ghasidas vishwavidyalaya, Bilaspur, Chhattisgarh, India.Email: amitsaxena65@rediffmail.com)
** (Department of Computer Science and Information Technology (CSIT), Guru Ghasidas vishwavidyalaya, Bilaspur, Chhattisgarh, India.Email: 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 confuse the process of classification. The proposed method first applies a binary coded genetic algorithm to select a small subset of features. The importance of these features is judged by applying Naïve Bayes (NB) 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 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, Classification, Feature selection, Naïve Bayes (NB), Genetic Algorithm (GA), Leave-one-out cross validation (LOOCV).

I. INTRODUCTION

Data mining [1] is an inter-disciplinary field for research. It is one of the most important areas in the field of research that is used for discovering, evaluating and finding new interesting patterns from real data set. There are various applications of data mining technique like classification [2], association mining rule [3], clustering [4] and pattern recognition [5]. Classification plays an important role in data mining application that can be applied in real world application for classification of various data such as large volume of data (large number of features/attributes) [6]. The supervised learning strategy emphasizes on building models that are able to assign new instances to one of the well defined classes [7]. Feature selection [8] is one of the important concepts that is used for reducing the features from data set and still improve the performance of model. Feature selection not only improves accuracy of model but also reduces the computation time of the model. In this paper, we used Genetic algorithm (GA) based Naïve Bayes (NB) feature selection and classification for the datasets ionosphere, Australian, Wine, Sonar, German, Statlog Heart, Wisconsin Diagnostic Breast Cancer (WDBC), Wisconsin Prognostic Breast Cancer (WPBC).

The rest of this paper is organized as follows: Section II summarizes the Naïve Bayes (NB) related Works, Section III presents preliminaries for the Basic Concepts of GA and NB based classification. Section IV provides proposed method; Section V lists datasets in brief. In Section VI, experimental study is provided whereas in Section VII, results and discussions are given. Finally the paper presents conclusions with future discussions.

II. RELATED WORKS

Various authors have worked in the field of classification of data using Naïve Bayes (NB) classification techniques. They have also used feature selection techniques to reduce the features and improve the performance of models. Santosh Kumar , G. Sahoo [9] proposed NB and GA based algorithm for classification of heart disease to improve the accuracy. Guozhong Fenga et. al. [10] proposed a latent selection augmented (LSAN) NB classifier for Feature subset selection for classification of text data. XuZhang [11] presented a novel learning-free image classification technique i.e. NBCR, under the framework of NB Nearest-Neighbor (NBNN) and collaborative representation, where non- negative sparse coding, low-rank matrix recovery and collaborative representation are jointly employed to obtain more robust and discriminative representation. Chang-Hwan Lee [12] suggested a new paradigm of assigning weights in classification learning that is value weighting method. The suggested method is implemented in the context of NB learning, and optimal weights of feature values are calculated using a gradient approach. The performance of NB learning with value weighting method is compared

with that of other state-of-the-art methods for a number of datasets. Jia Wu et.al. [13] presented a new Artificial Immune System (AIS) based self-adaptive attribute weighting method for NB classification, namely AISWNB. It uses immunity theory in Artificial Immune Systems to search optimal attribute weight values, where self-adjusted weight values will alleviate the conditional independence assumption and help to calculate the conditional probability in an accurate way. Bilal, M. et.al. [14] have used three different classification models for text classification using Waikato Environment for Knowledge Analysis (WEKA). This research was conducted on Roman Urdu opinion mining by using these classification algorithms i.e. Naïve Bayes, Decision Tree and KNN. Devesh Kumar et.al. [15] proposed comparative study of calculation of classification error with classical PCA technique and adaptive method. They have used Naïve Bayes Classifier for calculating the classification error of each feature vector instead of considering K largest Eigen-value as in PCA. Dewan Md. Farid [16] introduced two independent hybrid mining algorithms to improve the classification accuracy rates that are decision tree (DT) and naive Bayes (NB) classifiers for the classification of multi-class classification problems. Pablo Bermejo [17] proposed hybrid model as NB with incremental wrapper FSS algorithms. Optimization for incremental selection algorithms when using classifiers that allows incremental construction when adding variables, as is the case of Naive Bayes. Khadija Al-Aidarous et. al. [18] presented a new variant of NB classifier which involves weighting of attributes using rough set analysis. The method makes use of two rough set measures. Rough set dependency measure is used to analyze the dependencies among the attributes and the rough set significance measure is used to reflect the importance of each attribute in the final produced classifier. The experimental results using benchmark data sets are promising and they suggest that the proposed Rough Naïve Bayes (RNB) usually gives the most accurate results compared to the other approaches. Li Liu et. al. [19] proposed a method of feature selection using the AdaBoost algorithm for action recognition. This method is extended of selecting the most discriminative features using the AdaBoost algorithm to the human-action recognition task. Damrongrit Setsirichok [20] presented the classification of blood characteristics by a C4.5 decision tree and naive Bayes classifier. The aim is to classify eighteen classes of thalassaemia abnormality, which have a high prevalence in Thailand, and one control class by inspecting data characterised by a complete blood count (CBC) and hemoglobin typing and a multilayer perceptron for thalassaemia screening. In this article, a thalassaemia classification problem in Thailand is investigated. The aim is to identify whether the human subject is a person with abnormal haemoglobin, a person with thalassaemia trait, a thalassaemic patient or a normal person using complete blood count (CBC) and haemoglobin typing data. Liwei Fan et.al. [21] presented a sequential feature extraction approach for naïve Bayes classification of microarray data. Mukherjee et. al. [22] proposed Feature Vitality Based Reduction Method to identify important reduced input features. They have used one of the efficient classifier i.e. naive bayes on reduced datasets for intrusion detection. Jingnian Chen et. al. [23] presented two feature evaluation metrics (CDM and MOR) to reduce features and this features are applied on the

NB classifier for multi-class text collections.

III. PRELIMINARIES

There are certain terms that are briefed here.

A. Genetic Algorithm

Genetic algorithm was first proposed by John Holland in 1975 [24]. 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 operators such as selection (based on natural genetic 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 [25].

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 and mutation. Among the several available selection methods, roulette wheel selection, stochastic universal sampling and binary tournament selection are three widely used techniques [25].

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 [25].

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, 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 [25], [26],[27]. A more complete description about Genetic Algorithm can be found in [26], [27].

B. Naïve Bayesian Classification

In machine learning, Naive Bayesian Classification is a family of a simple probabilistic classifier based on the Bayes theorem (or Bayes's rule) with Naive (Strong) independence assumption between the features. It is one of the most efficient and effective classification algorithms and expresses a supervised learning method along with statistical method for classification [18]. Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. In other words Naïve Bayesian classifiers assume that there are no dependencies amongst attributes. This assumption is called class conditional independence. It is made to simplify the computations involved and, in this sense, is considered "naïve. It is particularly suited when the dimensionality of the inputs is high. When we want more competent output, as compared to other methods output we can use Naïve Bayes implementation. Naïve Bayesian is used to create models with predictive capabilities. This classifier is also called Simple Bayesian Classifier, idiot Bayes, simple Bayes, or independent Bayes.[27],[28][29].

Let X be a data tuple. In, Bayesian terms, X is considered "evidence". Let H be some hypothesis, such as that the data tuple X belongs class C. P(H|X) is the posterior probability, of H conditioned on X. In contrast, P(H) is the prior probability, or a priori probability, of H. Bayes' theorem is

$$P(H|X) = \frac{P(X|H)\,P(H)}{P(X)} \quad \ldots\ldots\ldots (1)$$

Similarly, P(X|H) is the posterior probability of X conditioned on H. P(X) is the prior probability of X [16][27].

The naïve Bayesian classifier, or simple Bayesian classifier, works as follows:

1. Let D be a training set of tuples and their associated class labels. As usual, each tuple is represented by an n-dimensional attribute vector, X = (x_1, x_2... x_n), depicting n measurements made on the tuple from n attributes, respectively, A_1, A_2... A_n.

2. Suppose that there are m classes, C_1, C_2... C_m. Given a tuple, X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is, the naïve Bayesian classifier predicts that tuple X belongs to the class C_i if and only if

$P(C_i|X) > P(C_j|X)$ for $1 \le j \le m, j \ne i$.

Thus we maximize $P(C_i|X)$. The class C_i for which $P(C_i|X)$ is maximized is called the maximum posteriori hypothesis. By Bayes' theorem (1).

$$P(C_i|X) = P(X|C_i)P(C_i)\,/P(X) \quad \ldots\ldots (2)$$

3. As P(X) is constant for all classes, only $P(X|C_i)P(C_i)$ need be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, that is, $P(C_1) = P(C_2) = \cdots = P(C_m)$, and we would therefore maximize $P(X|C_i)$. Otherwise, we maximize $P(X|C_i)P(C_i)$. Note that the class prior probabilities may be estimated by $P(C_i) = |C_i,D|/|D|$, where $|C_i,D|$ is the number of training tuples of class C_i in D.

4. Given data sets with many attributes, it would be extremely computationally expensive to compute $P(X|C_i)$. In order to reduce computation in evaluating $P(X|C_i)$, the naive assumption of class conditional independence is made. This presumes

that the values of the attributes are conditionally independent of one another, given the class label of the tuple (i.e., that there are no dependence relationships among the attributes). Thus,

$P(X|C_i) = \prod_{k=1}^{n} P(xk|Ci)$ ….. (3)

$= P(x_1|C_i)\times P(x_2|C_i)\times\cdots\times P(x_n|C_i)$.

We can easily estimate the probabilities $P(x_1|C_i)$, $P(x_2|C_i)$,..., $P(x_n|C_i)$ from the training tuples. Recall that here x_k refers to the value of attribute A_k for tuple X. For each attribute, we look at whether the attribute is categorical or continuous-valued. For instance, to compute $P(X|C_i)$, we consider the following:

(a) If A_k is categorical, then $P(x_k|C_i)$ is the number of tuples of class C_i in D having the value x_k for A_k, divided by $|C_i,D|$, the number of tuples of class C_i in D

(b) If A_k is continuous-valued, then we need to do a bit more work, but the calculation is pretty straightforward. A continuous-valued attribute is typically assumed to have a Gaussian distribution with a mean μ and σ or o standard deviation, defined by

$$g(x,\mu,o) = \frac{1}{\sqrt{2\pi}o} e^{-\frac{(x-u)2}{2\pi 2}} \quad \ldots\ldots\ldots\ldots (4)$$

$P(x_k|Ci) = g(x_k, \mu C_i , oC_i)$ …… ….(5)

5. In order to predict the class label of X, $P(X|C_i)P(C_i)$ is evaluated for each class C_i. The classifier predicts that the class label of tuple X is the class Ci if and only if

$P(X|Ci)P(C_i) > P(X|C_j)P(C_j)$ for $1 \le j \le m, j \ne i$. … (6)

In other words, the predicted class label is the class C_i for which $P(X|C_i)P(C_i)$ is the maximum.

In theory, Bayesian classifiers have the minimum error rate in comparison to all other classifiers. However, in practice this is not always the case, owing to inaccuracies in the assumptions made for its use, such as class conditional independence, and the lack of available probability data. Bayesian classifiers are also useful in that they provide a theoretical justification for other classifiers that do not explicitly use Bayes' theorem. For example, under certain assumptions, it can be shown that many neural networks and curve-fitting algorithms output the maximum posteriori hypothesis, as does the naïve Bayesian classifier. [16],[27],[28],[29].

IV. PROPOSED METHOD

In this study, we propose the combination of genetic algorithm and Naïve Bayes (NB) with Leave one out cross validation (LOOCV) [30] to improve the classification accuracy of unbalanced data set that contains classes. We use genetic algorithm for feature selection. In this process, first initialize each chromosome by filling 0 or 1 in its genes. Number of genes (say n) in each chromosome is equal to 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 Naïve Bayes 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.

Figure1 shows the complete algorithm for the proposed method and Figure 2 shows model for the proposed method.

$$\text{Accuracy} = \frac{\text{Number of samples correctly classified in test data}}{\text{Total no.of samples in the test data}} \text{ X } 100\%$$

- D [m, n] (Where m number of patterns, n number of features)
  // use Genetic Algorithm for feature selection and NB as classifier to evaluate goodness of each chromosomes
- Initialize chromosomes randomly with binary value at each gens
  (number of genes is equal to number of features)
  For 1: number of iteration
  a. Calculate fitness of each chromosome using NB classifier (with LOOCV cross validation technique)
  b. Create new child using three operators
  - Selection
  - Crossover
  - Mutation

  End
- Finally we get accuracy and number of features

Fig.1. Algorithm for proposed method

V. DATASETS

We performed experiments on the datasets listed in Table I. These datasets are collected from University of California, Irvine (UCI) repository [31]. Table I has four columns, first column contains dataset name, second column contains total number of instances (records or rows), third column contains total number of features (excluding class

attribute) and fourth column contains Classes. Table I has eight datasets namely Ionosphere, Australian, Wine, Sonar, German, Heart, Wisconsin Diagnosis breast cancer, (WDBC) datasets, Wisconsin Prognostic Breast Cancer (WPBC). The outline of these databases can be seen in Table I.

TABLE I. A BRIEF DESCRIPTION OF THE DATASETS USED IN THIS EXPERIMENT.

Dataset name	Total no. of	Total No. of	Classes
Ionosphere	351	34	2
Australian	690	14	2
Wine	178	13	3
Sonar	208	60	2
German	1000	24	2
Heart	270	13	2
WDBC	569	30	2
WPBC	198	33	2

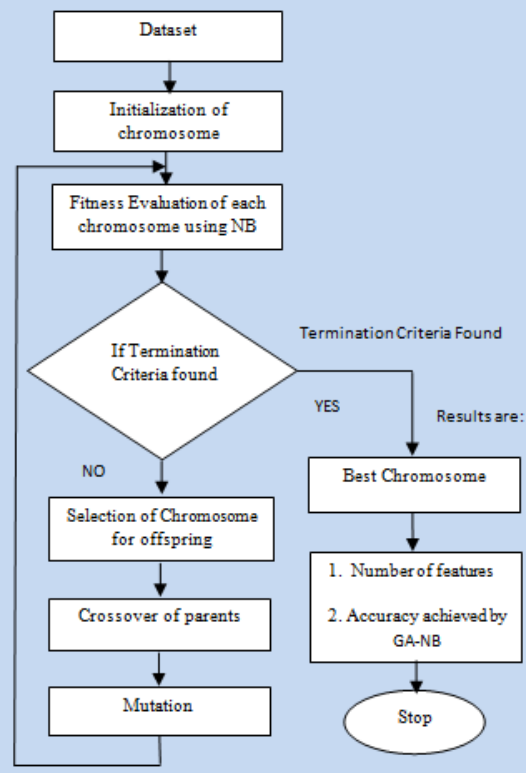

Fig. 2. Model of proposed method

VI. EXPERIMENTS

We performed our experiments on Intel i3 Processor with 2 GB RAM and 500 GB hard disk. MATLAB is used for development of code and experiment. Also GA MATLAB code [32] is used for this study. We used leave one out cross validation technique to obtain robust classification accuracy. In this method the whole dataset 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.

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, first column contains dataset name, second columns contains number of selected features in the dataset, third columns contains total number of features and fourth column contains accuracy by proposed method. Table II contains results for eight datasets named Ionosphere, Australian, Wine, Sonar, German, Stalog Heart, Wisconsin Diagnosis breast cancer (WDBC) datasets, Wisconsin Prognostic Breast Cancer (WPBC). As shown in table II, Ionosphere dataset reduces to 9 relevant features out of 34 features and classification accuracy is 92.59.Australian datasets has 9 relevant features out of 14 features and classification accuracy is 87.68. Wine datasets has 10 relevant features out of 13 features and classification accuracy is 99.44. Sonar datasets has 26 relevant features out of total 60 numbers of features and classification accuracy is 80.29. German datasets has 14 selected features out of 24 numbers of features and classification accuracy is 76.50. Stalog Heart cancer has 10 selected features out of 13 features and classification accuracy is 86.30. Wisconsin Diagnosis breast cancer (WDBC) has 12 selected features out of 30 features and classification accuracy is 97.89, Wisconsin Prognostic Breast Cancer (WPBC) has 5 selected features out of 33 features and classification accuracy is 79.80.

TABLE II. RESULTS FOR PROPOSED METHOD

Dataset name	No. of selected Features in the dataset	Total No. of Features	Accuracy by proposed method
Ionosphere	9	34	**92.59**
Australian	9	14	**87.68**
Wine	10	13	**99.44**
Sonar	26	60	**80.29**
German	14	24	**76.50**
Heart	10	13	**86.30**
WDBC	12	30	**97.89**
WPBC	5	33	**79.80**

In Table III we compared results obtained by proposed method with results obtained by other methods. Tables III has five columns, First column contains dataset name, second column contains accuracy obtained by proposed method, third column contains accuracy obtained Simultaneous feature selection and weighting based method [33], Fourth column contains accuracy obtained by AIS

based attribute weighted Naïve Bayes method (AISWNB) [13] and fifth column contains accuracy obtained by JSDA method [34]. In case of Ionosphere dataset the classification accuracy obtained by the proposed method GA-NB (LOOCV) is 92.59 and it is compared with accuracy 88.31 obtaind by Simultaneous feature selection and weighting based method proposed by Sujoy Paul, Swagatam Das [33] and accuracy 90.69 obtained by AISWNB method proposed by Jia Wu [13]. In case of Australian dataset the classification accuracy obtained by the proposed method GA-NB (LOOCV) is 87.68 and it is compared with accuracy 84.64 obtained by Simultaneous feature selection and weighting based method proposed by Sujoy Paul, Swagatam Das [33] and accuracy 85.13 obtaind by AISWNB method proposed by Jia Wu [13]. In case of Wine dataset the classification accuracy obtained by the proposed method GA-NB (LOOCV) is 99.44 and it is compared with accuracy 96.05 obtained by Simultaneous feature selection and weighting based method proposed by Sujoy Paul, Swagatam Das [33] .In case of Sonar dataset the classification accuracy obtained by the proposed method GA-NB (LOOCV) is 80.29 and it is compared with accuracy 76.76 of AISWNB based method proposed by Jia Wu. et.al [13]. In case of German dataset the classification accuracy obtained by the proposed method GA-NB (LOOCV) is 76.50 and it is compared with accuracy of 71.30 obtained by feature selection and weighting based method proposed by Sujoy Paul, Swagatam Das [33] and accuracy 75.80 obtained by AISWNB method proposed by Jia Wu [13]. In case of Heart dataset the classification accuracy obtained by the proposed method GA-NB (LOOCV) is 86.30 and it is compared with accuracy 80.00 obtained by Simultaneous feature selection and weighting based method proposed by Sujoy Paul, Swagatam Das [33] and accuracy 83.52 obtained by AISWNB method proposed by Jia Wu [13]. In case of WDBC dataset the classification accuracy obtained by the proposed method GA-NB (LOOCV) is 97.89 and it is compared with accuracy 94.06 obtained by Simultaneous feature selection and weighting based method proposed by Sujoy Paul, Swagatam Das [33] and accuracy 93.81 obtained by JSDA method proposed by Kong, Heng[34], In case of WPBC dataset the classification accuracy obtained by the proposed method GA-NB (LOOCV) is 79.80 and it is compared with accuracy 64.00 obtained by JSDA based method proposed by Kong, Heng [34] .

It is evident that the classification accuracy obtained by proposed method on 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 Fig. 3. The classification accuracy obtained by proposed method is shown by a blue bar and it is taller in each of the eight databases compared to other methods shown by red (feature selection and weighting method), green (AISWNB) and violet (JSDA) colors.

Table III. Comparison of accuracy obtained by different methods

Data set name	Proposed Method	Feature selection and weighting Method	AISWNB	JSDA
Ionosphere	**92.59**	88.31	90.69	-
Australian	**87.68**	84.64	85.13	-
Wine	**99.44**	96.05	-	-
Sonar	**80.29**	-	76.76	-
German	**76.50**	71.30	75.80	-
Heart	**86.30**	80.00	83.52	-
WDBC	**97.89**	94.06	-	93.81
WPBC	**79.80**	-	-	64.00

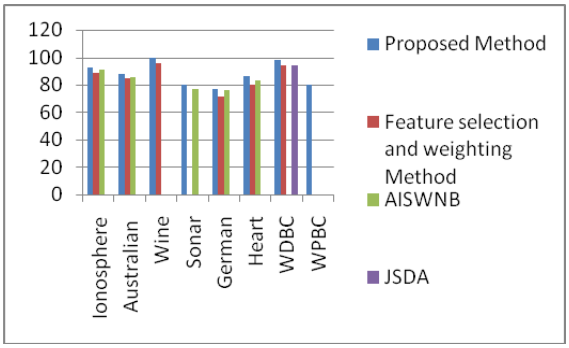

Fig.3. Graphical representation for Comparison of accuracy obtained by different methods

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 Ionosphere, Australian, Wine, Sonar, German, Stalog Heart, Wisconsin Diagnosis breast cancer (WDBC) 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.

REFERENCES

[1]. A. Ghosh, S. Dehuri, (eds): ''Multi-objective Evolutionary Algorithms for knowledge Discovery from Databases''. Springer.verlag, Heidelberg, Germany, 2008

[2]. 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.

[3]. S. Saravanan, T. Christopher, "A study on milestones of association rule mining algorithms in large databases," Int. J. Comput. Appl., vol. 47, No. 3, pp. 12–19, 2012.

[4]. D.W. Van der Merwe and A.P. Engelbrecht, "Data Clustering using Particle Swarm Optimization," Procedings of IEEE Congress on Evolutionary Computation, vol. 1, pp. 215-220, 2003.

[5]. 'Pattern Classification' Duda, Hart, Stork, Wiley 2001.

[6]. Alper Unler, Alper Murat, Ratna Babu Chainnam, "Mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification," Information science, Vol. 181, pp. 4625-4641, Oct. 2011.

[7]. Suresh Chandra Satapathy, J.V.R Murthy, P.V.G.D. Prasad Reddy, B.B. Mishra, P.K. Dash, G. Panda, "Particle swarm optimized multiple regression linear model for data classification," Applied Soft Computing,Vol. 9 , pp. 470-476, march 2009 .

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

[9]. Santosh Kumar , G. Sahoo, Classification of Heart Disease Using Naïve Bayes and Genetic Algorithm, Computational Intelligence in Data Mining-Volume 2, pp 269-282, Springer,2015.

[10]. Guozhong Fenga, Jianhua Guoa , Bing-Yi Jing , Tieli Sun, Feature subset selection using naive Bayes for text classification, Pattern Recognition Letters 65 (2015) 109–115.

[11]. XuZhang, ShijieHao, ChenyangXu, XuemingQian, MengWang , JianguoJiang, Image classification based on low-rank matrix recovery and Naive Bayes collaborative representation, Neurocomputing169(2015)110–118

[12]. Chang-Hwan Lee, A gradient approach for value weighted classification learning in naiveBayes, Knowledge-Based Systems 85 (2015) 71–79

[13]. Jia Wu, Shirui Pan , Xingquan Zhu , Zhihua Cai , Peng Zhang , Chengqi Zhang , Self-adaptive attribute weighting for Naive Bayes classification, Expert Systems with Applications 42 (2015) 1487–1502

[14]. Bilal, M., Israr, H., Shahid, M., Khan, A., Sentiment classification of Roman-Urdu opinions using Navie Baysian, Decision Tree and KNN classification techniques, Journal of King Saud University – Computer and Information Sciences (2015),pp1-31 doi: http://dx.doi.org/10.1016/j.jksuci.2015.11.003.

[15]. Devesh Kumar, Ravinder Singh , Abhishek Kumar, Nagesh Sharma, An Adaptive Method of PCA for Minimization of Classification Error Using Naïve Bayes Classifier, Procedia Computer Science 70 (2015) 9 – 15.

[16]. Dewan Md. Farid , Li Zhang , Chowdhury Mofizur Rahman , M.A. Hossain , Rebecca Strachan, Hybrid decision tree and naive Bayes classifiers for multi-class classification tasks, Expert Systems with Applications 41 (2014) 1937–1946.

[17]. Pablo Bermejo , José A. Gámez, José M. Puerta, Speeding up incremental wrapper feature subset selection with Naïve Bayes classifier, Knowledge-Based Systems 55 (2014) 140–147

[18]. Khadija Al-Aidarous, Azuraliza Abu Bakar, Zalinda Othman, Improving Naïve Bayes Classification with Rough Set Analysis, International Journal of Advancements in Computing Technology(IJACT) Volume5, Number13, september 2013.

[19]. Li Liu,LingShao, PeterRockett, Human action recognition based on boosted feature selection and naive Bayes nearest-neighbor classification, Signal Processing 93 (2013) 1521–1530

[20]. Damrongrit Setsirichok, Theera Piroonratana, Waranyu Wongseree, Touchpong Usavanarong,Nuttawut Paulkhaolarn, Chompunut Kanjanakorn, Monchan Sirikong, Chanin Limwongse,Nachol Chaiyaratana,

Classification of complete blood count and haemoglobin typing data by a C4.5 decision tree, a naive Bayes classifier and a multilayer perceptron for thalassaemia screening, Biomedical Signal Processing and Control 7 (2012) 202– 212

[21]. Liwei Fan a, Kim-Leng ,Poh a , Peng Zhou b, A sequential feature extraction approach for naïve bayes classification of microarray data, Expert Systems with Applications 36 (2009) 9919–9923

[22]. Dr. Saurabh Mukherjee, Neelam Sharma, Intrusion Detection using Naive Bayes Classifier with Feature Reduction, Procedia Technology 4 (2012) 119 – 128

[23]. Jingnian Chen , Houkuan Huang , Shengfeng Tian , Youli Qu , Feature selection for text classification with Naïve Bayes, Expert Systems with Applications 36 (2009) 5432–5435.

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

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

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

[27]. Han, J., Kamber M., Data Mining Concepts and Techniques, (Morgan Kaufmann Publishers), 2000 and 2006.

[28]. https://en.wikipedia.org/wiki/Naive_Bayes_classifier

[29]. Cagatay Catal , Ugur Sevim , Banu Diri , Practical development of an Eclipse-based software fault prediction tool using Naive Bayes algorithm, Expert Systems with Applications 38 (2011) 2347–2353

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

[31]. UCI Repository of Machine Learning Databases, Center for Machine Learning and intelligent Systems [Online]. Available :< http://archive.ics.uci.edu/ml/datasets.html >

[32]. Randy L. Haupt,Sue Ellen Haupt, Practicle genetic Algorithms, Wiley - Interscience, A john wiley & sons,Inc., Publication, Copyright © 2004 by John Wiley & Sons, Inc, Published by John Wiley & Sons, Inc., Hoboken, New Jersey. Published simultaneously in Canada.

[33]. Sujoy Paul, Swagatam Das, Simultaneous feature selection and weighting–Anevolutionary multi-objective optimization approach, Pattern Recognition Letters 65 (2015) pp. 51–59

[34]. Kong, Heng, "Breast cancer discriminant feature analysis for diagnosis via jointly sparse learning." Neurocomputing, Elsevier (2015).