International Journal of Soft Computing 11 (4): 255-269, 2016
ISSN: 1816-9503

Improving Classification Performance by Using Feature Selection with Resampling

[1]Raya Ismail, [1]Sherihan Abuelenin and [2]Ahmed Aboelfetouh
[1]Department of Computer Science,
[2]Department of Information System, Mansoura University, Mansoura, Egypt

Abstract: Feature selection methods tend to identify the most relevant features for classification and can be categorized as either subset selection (wrapper) methods or ranking (filter) methods. The main purpose of this stusy is to prove that a feature selection preprocessing step could enhance classifiers performance by eliminating redundant features. The proposed method consists of three stages; the first refines sample space domain by resample filtering, the second minimizes feature space by applying subset evaluation algorithm and the third measures the goodness of the resulting set of features using different classifiers. Tow experiments carried out on the data sets from UCI repository. The proposed method is evaluated by measuring the accuracy, number of selected features, precision, recall, f-measure, ROC area, time to build model, error rate and relative absolute error. Tests are done on two main types of classifiers Naive Bayes and its variance NBTREE, NBNET and J48 with other tree classifiers Random Forest, BFTREE.

Key words: Feature selection; classification; resampling, NBNET, BFTREE

INTRODUCTION

In machine learning, feature selection is the process of selecting a subset of relevant features for using in the model construction. Feature selection techniques are important for three reasons: simplification of models to make them easier to interpret by researchers/users, shorter training timesand enhanced generalization by reducing over fitting. Feature selection plays an important role in building classification systems, it can not only reduce the dimension of data, but also lower the computation consumption and so that it can gain good classification performance Xie and Wang (2011). Any machine learning can perform classification using a set of features. In the past years the machine learning applications of pattern recognition in the domain of features have expanded from tens to hundreds of features or attributes used in the applications of different fields. Several techniques are presented to solve the problem of reducing irrelevant and redundant attributes. Feature selection helps in reducing computation requirement, understanding data and predictor performance Chandrashekar and Shin (2014). The selection of relevant object-features-efficient sampling of training data are frequently encountered issues when applying supervised classification techniques in different analysis. For both rule-based classification and the training of supervised algorithms it remains crucial to identify truly relevant featuress and disregarded irrelevant attributes that may deteriorate the classification.

Dozens of feature selection methods have been developed in the last decades, the main categories of feature selection approaches can be classified into: filter, wrapper and embedded methods. Selecting appropriate features is an important step in both machine learning and data mining processes, feature selection helps in understanding data, reducing the effect of high dimensionality, improving the classifier performance and reducing computation requirements Chandrashekar and Sahin (2014).

In filter methods a suitable ranking criterion is used to rank the features and a threshold is used to remove features below the threshold, so relevance issue has to be raised to answer the question:" how do we measure the relevance of anattribute". Essentially, the feature is relevant if it can be independent of the input data but cannot be independent of the class label Chandrashekar and Sahin (2014). The main advantage of filter methods is easy and fast to implement. The characteristics of filter methods are as follows: including redundant features considering the features independently ignoring some features which as group have strong discriminatory and individually are weak and the filtering procedure is independent of the classifying method. Correlation Criteria and Mutual Information (MI) are the most filter methods used. On the other handin wrapper methods feature selection process uses the classifier accuracy to evaluate the performance of each subset,since evaluating

Corresponding Author: Raya Ismail, Department of Computer Science, Mansoura University, Mansoura, Egypt

subsets becomes hard problem sub optimal subsets are found using different search strategies to find a subset heuristically.

In wrappers, the goal is to maximize the objective function (classifier accuracy), some of the search algorithms used to achievethat goal such as Sequential Selection Algorithms and Heuristic Selection Algorithms Chandrashekar and Sahin, (2014). Wrappers have the advantage of achieving greater classification accuracy than filters but it consumes more time and obtains a feature subset that is biased towards the classifier used Bermejo *et al*.(2012). Embedded methods are another type of feature subset selection, where the feature selection process is done inside the induction algorithm itself, attempting to jointly or simultaneously train both a classifier and a feature subset Kotsiantis. The maingoal of feature selection is to find the best subset consisting of n features chosen from the total m features. Critical problem for many feature selection methods is that an exhaustive search strategy has to be applied to seek the best subset among all the possible $\binom{m}{n}$ feature subsets which has computational complexity.

To evaluate the best subset of features,various classification algorithms can be used such as Support Vector Machine (SVM), Random Forest, J48, K Nearest Neighbor and Naïve Bayes.

In classification approach, Support Vector Machines (SVM) is widely used in many machine learning tasks due to its perfect performance in generalization of different classification problems such as in text classification, disease diagnosis, voice recognition and so on, SVM algorithm based on maximizing the margin hyper plane to handle classification in nonlinear manner using kernel function Xie and Wang, (2011).

In the field of probability, In (Patil and Sherekar, 2013) the Naive Bayes classifier uses Bayesian theorem to predict new labels as in the following:

$$P(h_1|x_i) = \frac{P(x_i|h_1)P(h_1)}{P(x_i|h_1)P(h_1) + P(x_i|h_2)P(h_2)} \quad (1)$$

Where:

$P(h_1/x_i)$ = Posterior probability and

$P(h_1)$ = The prior probability of hypothesis

In generaln different hypothesis, the equation will be as follows:

$$P(x_i) = \sum_{j=1}^{n} P(x_i|h_j)P(h_j) \quad (2)$$

$$P(h_1|x_i) = \frac{P(x_i|h_1)P(h_1)}{P(x_i)} \quad (3)$$

NB classifier reduces the cost of computation by assuming that the impact of a feature value on a certain class is independent of the values of other features. Other variances of NB are Naïve Bayes Tree (NBTREE) and Bayes Net (BNET), they adopts the same idea of probability where the NBTREE classifier builds tree and uses the NB at each leaf and BNET generates class estimator based on analyzing the training data (Srimani and Koti, 2013).

Decision trees are big family of classifiers where trees are constructed using different algorithms to solve classification in binary and multi classification problems. J48 is a decision tree classifier using C4.5 algorithm to build a binary tree and apply it to each tuple in the dataset Patil and Sherekar (2013). This classifier tries to find the attribute with highest information gain that will be more helpful in classifying the data instances (Win and Khaing, 2014). Other tree classifiers like Best First Tree (BFTREE) that builds a binary tree using Best-First decision tree classifier and Random Forests (RF) from the name we can tell it construct random trees with different tree algorithms, random tree means that the classifier construct a tree with random M attributes at each node Srimani and Koti (2013).

Wrapper methods use the predictor as a black box and the predictor performance as the objective function to evaluate the variable subset. A number of search algorithms can be used to find which maximizes the objective function which is the classification performance. The wrapper methods are classified into: Sequential Selection Algorithms and Heuristic Search Algorithms Chandrashekar and Sahin (2014). Wrapper methods are better in defining optimal features rather than simply relevant features and they do that by allowing for the specific biases and heuristics of the learning algorithm and the training set. Predefined learning algorithm needed in wrapper methods to identify the relevant features. But, it takes more time comparing filter method Vanaja and Kumar (2014). Three main steps in wrapper algorithm:

- Subset generation: generating subsets of features by a search procedure, the total number of nominee subsets is where N is the number of original set of features
- Subset evaluation: evaluating each subset produced by the generating procedure using certain evaluation criterion and comparing it with the last previous best subset if it is better (according to the criterion), then it replaces
- Stopping criteria: a feature selection process may stop under one of these criteria
- Reaching a certain predefined number of iterations
- Selecting a predefined number of features
- in case addition (or deletion) of features fails to find a better subsets

- Obtaining an optimal subset according to the evaluation criterion (Bidgol and Parsa, 2012)

Sequential Forward Floating Selection (SFFS) as shown in works in bottom-to-top manner, starting with an empty set of attributes; the single best attribute of the original attributes is determined and added to the set. In each iteration the best one attribute of the remaining original is added to the set Xie and Wang (2011), The starting point of sequential search can be an empty set which is then successively built up or the starting point can be the complete set of features, Y, both these methods are generally suboptimal and suffer from the so called "nesting effect" there for this rezone, the floating search method of feature selection is presented by Pudil. The Sequential Floating Forward Selection (SFFS) algorithm is more flexible than SFS because it adds another step which excludes one feature at a time from the subset obtained in the first step and evaluates the new subsets byChandrashekar and Sahin (2014).

Sequential forward floating selection algorithem (Penget *et al.*, 2010):

$K = 0$; $X_k = \phi$; $Y_k = U$;
While the stop criteria have not been fulfilled {
  $Y = \text{argmax}_{c \in y_k} J(X_k \cup \{c\})$// to find the best features set
$X_{k+1} = X_k \cup \{Y\}$; $k = k+1$ // inclusion
  $X\text{argmax}_{c \in y_k} = j(X_{k+1} - \{c\})$ // to find the least significant features
  While $J(X_{k+1}) > J(X_k)$ {
$X_k = m X_{k+1} - \{x\}$;
    K=k-1;// conditional exclusion
    $X = J(\text{argmax}_{c \in y_k} J(X_k - \{c\}))$

The main purpose of this paper is to propose an efficient resampling method that could improve the feature selection process and produce new subset of features to get good classifier performance.

Literature review: Liu *et al.* (2004) introduced active feature selection and selective sampling approaches by using KD-tree to partition data and select instance from those partitions, aiming to improve the performance of filter model settings. They Used data variance as selective sampling and apply it to Relief feature selection algorithm and use Distance Performance Measure to observe that this method save significant amount of time and improve nearest neighbor classifier accuracy on numeric data.

Way *et al.* (2010) evaluated the performance of different combinations of classifiers and simulate the effect of the training sample size on feature selection methods in a simulation study, using three feature selection algorithms, Sequential Forward Selection (SFS), Sequential Forward Floating Selection (SFFS) and Principal Component Analysis (PCA) with two types of classification algorithms, Support Vector Machine (SVM) and Fishers Discriminant analysis all those methods were investigated for 15-100 sample size per class. In result they find that no superior performance in these combinations under the conditions of this study, but the performance of SVM with Radial Kernel was better than with or computable to that with Polynomial Kernel.

Sebastien investigated context of landslide inventory mapping from VHR satellite images. They used Random Forest with active learning heuristic to improve the spatial coverage sampling performance by selecting sample batches in spatial neighborhoods with a high variance of the Vote-entropy. This study combined feature selection approach with iterative routine and QBC sampling to deal with three important aspects cost-efficient, selecting relevant features and class-imbalance. In the result the method achieve good accuracy using small training set.

Srimani and Koti (2013) proposed a procedure that uses resampling approach to solve two important issues when dealing with a combination of mutual information and forward selection in a hybrid feature selection method. The first issue is the number of features to be chosen when forward selection is applied. Second determining suitable relevance criterion estimator which can be number of units or a kernel width in a prediction model, a number of neighbors or a number of bins in a non-parametric relevance estimator, tests have been done on a synthetic dataset and real-world examples.

Naseriparsa *et al.* (2014) used two phase method based on derive a secondary dataset from the original and select reliable features.In the first phase the author increases the samples of minority class by SMOTE technique and apply sample filtering on the resulting samples in the second phase he produces a hybrid procedure by using information gain and genetic search in subset evaluation to find the best subset of attributes, this method decrease the classification errors for five classifiers (Logistic, Multilayer perceptron, Best First Tree, JRIP and Naïve Bayes).

Sasikala *et al.* (2014) proposed a multi Filtration Feature Selection method (MFFS) that based on maximizing the classification accuracy to build a model after adjusting the "variance coverage" parameter, he uses feature subset selection and ranking methods in order to provide a small set of optimal features in four stage procedure first feature extraction, second feature selection third feature ranking and fourth classification, the method tested on 22 medical datasets to help physicians in medical prescriptions.

MATERIALS AND METHODS

Most of recent studies in the field of combining resampling filter with feature selection methods used the hybrid approach in feature selection. In this study, the

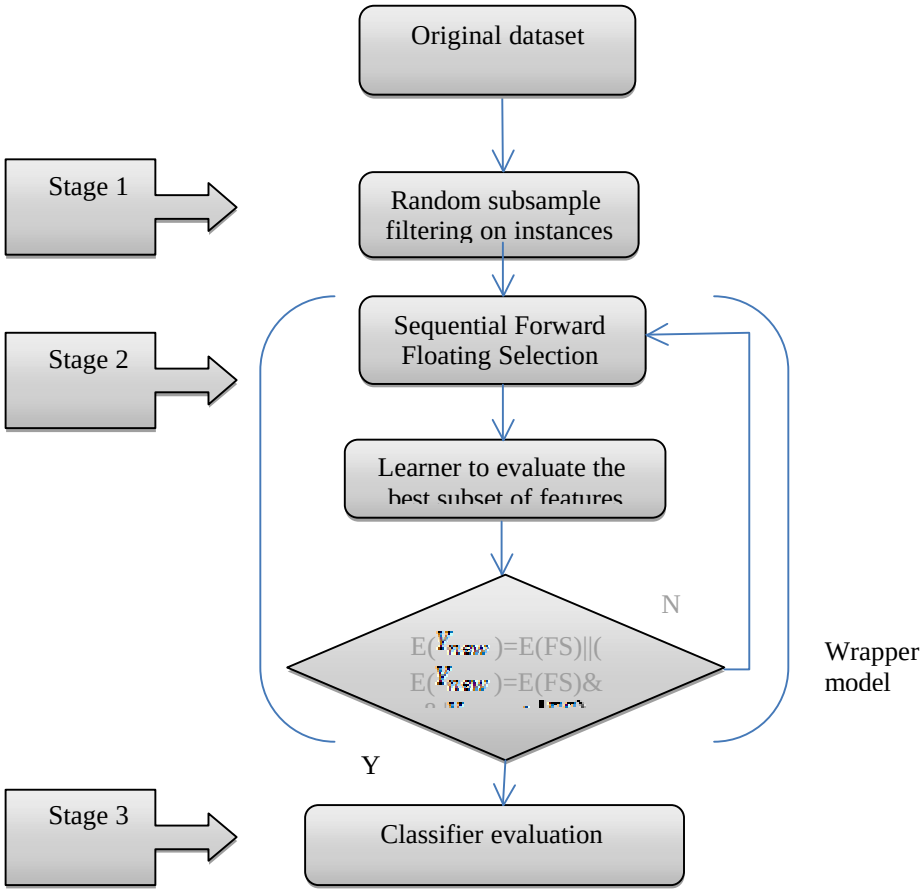

Fig. 1: Flows char of the proposed method

proposed method is based on the wrapper approach without ranking, since wrappers are more specific in finding relevant features for each classifier and combine it with filter on instances to reduce the search space in each stage.

In the proposedmethod, the sample domain filtering is applied to remove some instances that could decrease the classifier accuracy. Then a wrapper approach is used to eliminate the irrelevant features by employing the SFFS as search method and classifiers accuracy as evaluator. In this case two dimensions of the dataset will be invoked in the process to gain significant performance with much less features than the original feature space.

The proposed algorithem:2
Functions and Variables
X - The instance data,
Y - The list of labels (features),
n- Number of labels,
D[X, Y] - randomly sampled dataset instances,
E- Evaluation function (NB, J48) to be maximized,
FS- feature subsets,
Per-sample size
Input: D[X, Y]
Output: Y_{best}
Begin

1. Resampling (X,Y, per)
2. FS: = [all possible subsets]
3. For I = 1 to | FS |
4. Y_{new} SFFS (E,FS [i])= SFFS (E, FS[i])
5. If E (Y_{new}) =E (FS[i]) || (E (Y_{new}) =E (FS[i]) && |Y_{new} | < |FS[i]|)
6. then Y_{best} :=Y_{new}
7. End if
8. End for
9. Out put

End

A procedural description of the proposed algorithm as:
Use random sampling to select the instances of the training data that will be used in feature selection stage

[D] = (X, Y, per)

Where:

X = The instance data
Y = The list of labels (features), per is the sample size in percentage

Execute a wrapper model with SFFS using classifier subset evaluators. SFFS [naïve Bayes (X_{new} Y) or SFFS J48 (X_{new} Y)]. Test $_{new}$ subset if it is the best one for the learner then save it as Y_{best} Fig. 1.

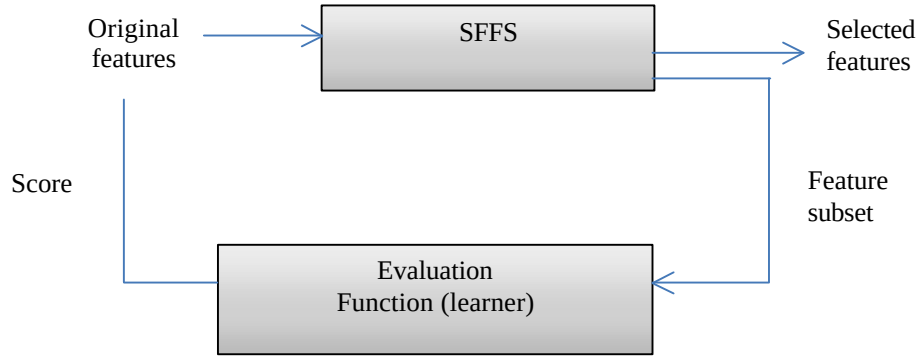

Fig. 2: Feature selection process

The proposed method contains three main stages.The full flowchart of the proposed method is shown in Fig. 1 the individual phases are described in details in the following.

or without replacement.The sampling with replacemen **Stage 1-resampling filter:** In this stage a Resampling filtration on dataset instances is applied, this filter uses random sampling where instances are sampled using either a random number table or a random number generator. In random sampling process the instances are included in a list called sampling frame, the list ordered sequentially. The instances can be filtered with Resampler in two ways with t is noticed that ismore efficient in 80% of sample size where the total number of samples is where M is the number of original instances and m is the number of instances to be sampled.

Stage 2-feature selection process: In this stage, the sequential forward floating selection is used as a search method.The classifier accuracy is used as evaluation function to determine which feature will be included in the resulting features set as shown in Fig. 2. In the next stage of evaluation the new features set will be invoked to measure the new accuracy.

Stage 3- classifier evaluation: The resulting subset of feature of each classifier is used to measure the accuracy of the one used, in more detailed.If Naïve Bayes is used as evaluator in the feature selection process then the resulting features are used in Naïve Bayes classification process and similarly on J48 classification. Choosing Naïve Bayes classifier is not only because it's quick and reliable but also it's cheap and has less computation than other classifiers. J48 decision tree classifier which is based on building a tree and branch depending on the attribute with the highest information gain has shown very good efficiency in classifying the used datasets.

RESULTS AND DISCUSSION

Test conditions: Testing study with different datasets has been executed, to analyze the performance of the proposed method in the following perspective:

- Classifier accuracy
- Number of selected features
- The impact of the new subset of features on different classifiers measurements like (precision, recall, f-measure, ROC area and time to build model)
- Error rate and relative absolute error

Datasets, test set-up and objectives for the experiments are described below.

Datasets: To evaluate the proposed method two experiments are executed; the first one used a collection of datasets from UCI machine learning repository (Peng *et al*., 2010) for different purposes and the second experiment used a medical purposes datasets also from UCI machine learning repository including Heart diseases datasetwhich contain four datasets from different regions, datasets used to test the proposed method are summarized in Table 1-11.

Test set-up: The experiments are countedwith following properties system, Intel core i3, 4 GB Ram, 400 GB hard drive a Windows 7(64 bit) operating system. The proposed method is implemented using Weka, Weka is a java based tool used in the field of machine learning and data mining. The input to the system is given as (ARFF) file, Attribute Relation File format. So all used datasets have been changed to the ARFF or CSV format.

On each data set, the sampling filter is applied with 80% sample size then run the feature selection algorithm twice, once with J48 and second with NB as evaluator and get the subset of features as output to use the later in classification task.

Table 1: Characteristics of first experiment datasets

Dataset	#Instances	#Features	#classes
Ionosphere	351	35	2
Adult	1198	15	2
labor	57	17	2

Table 2: Some performance evaluation metrics of the proposed method with NB on Ionosphere dataset

Classifier	Precision	Recall	F-measure	ROC area	Time to build model in seconds
NB	0.93	0.926	0.927	0.979	0.001
BNET	0.928	0.926	0.924	0.989	0.03
NBTREE	0.97	0.968	0.968	0.975	0.33 (#lea f=2, size=3)

Table 3: Some performance evaluation metrics of the proposed method with J48 on Ionosphere dataset

Classifier	Precision	Recall	F-measure	ROC area	# of leaves	Size oft ree	Timetobuild model in sec
J48	1	1	1	1	18	35	0.06
BFTREE	0.986	0.986	0.986	0.988	15	29	0.16
RANDOM FOREST	0.996	0.996	0.996	1	-	10	0.09

Table .4: Some performance evaluation metrics of the proposed method with NB on Adult dataset

Classifier	Precision	Recall	F-measure	ROC area	Time to build model in seconds
NB	0.802	0.815	0.783	0.782	0.001
BNET	0.831	0.839	0.82	0.789	0.01
NBTREE	0.831	0,839	0.82	0.789	0.17 (#leaf =1, size =1)

Table .5: Some performance evaluation metrics of the proposed method with J48 on Adult dataset

Classifier	Precision	Recall	F-measure	ROC area	# of leaves	Size of tree	Time to build model in sec
J48	0.954	0.954	0.953	0.965	145	192	0.08
BFTREE	0.945	0.946	0.945	0.954	53	105	0.91
RANDOM FOREST	0.994	0.994	0.994	1	-	10	0.14

Table. 6: Some performance evaluation metrics of the proposed method with NB on Labor dataset

Classifier	Precision	Recall	F-measure	ROC area	Time to build model in seconds
NB	1	1	1	1	0.001
BNET	0.94	.933	.932	.9	0.01
NBTREE	1	1	1	1	0.09 (#leaf=1, size=1)

Table 7: Some performance evaluation metrics of the proposed method with J48 on Labor dataset

Classifier	Precision	Recall	F-measure	ROC area	# of leaves	Size of tree	Timetobuild model in sec
J48	1	1	1	1	6	9	0.001
BFTREE	0.944	0.933	0.934	0.996	6	11	0.02
RANDOM FOREST	0.944	0.933	0.934	0.991	-	10	0.02

Table.8 Characteristics of the second experiment datasets

Dataset name	Instances	Features	Classes
Heart diseases	781	14	5
Cleveland	303	14	5
Hungarian	294	14	5
Long-beach	200	14	5
Switzerland	123	14	5
Missidor	1151	20	2
WDBC	569	32	2

Table 9: Some performance evaluation metrics of the proposed method with NB on Heart Disease dataset

Classifier	Precision	Recall	F-measure	ROC area	Time to build model in seconds
NB	0.695	0.688	0.685	0.856	0.01
BNET	0.619	0.659	0.627	0.859	0.03
NBTREE	0.79	0.79	0.784	0.929	1.52 (# leaves = 18, size = 35)

Table 10: Some performance evaluation metrics of the proposed method with J48 on Heart Disease dataset

Classifier	Precision	Recall	F-measure	ROC area	# of leaves	Size of tree	Time to build model
J48	0.911	0.912	0.911	0.989	97	193	0.07
BFTREE	0.892	0.891	0.887	0.97	89	177	0.28
RANDO MFOREST	0.986	0986	0.985	1	-	10	0.19

Table .11: Some performance evaluation metrics of the proposed method with NB on Cleveland dataset

Classifier	Precision	Recall	F-measure	ROC area	Time to build model in seconds
NB	0.745	0.74	0.737	0.893	0.001
BNET	0.663	0.682	0.651	0.883	0.02
NBTREE	0.64	0.678	0.657	0.883	0.03 (leaf = 37, size =73)

Classification models: NB is the first classifier in this study. NBis a simple probabilistic classifier, the presence (or absence) of a particular feature of class is not related to the presence (or absence) of other features. In the experiments on this classifier the resulting set of features are tested on other Naïve variances (BNET,NBTREE) and the data used are numeric or nominal and with both types it was fast and reliable.

C4.5 decision tree classifier, while building a tree, j48 ignores the missing values i.e., the value for that item can be predicted based on what known about the attribute values for the other records. J48 is the second classifier in this study.Also the new sub set of features resulted from J48 wrapped method is tested on the other tree classifiers (BFTREE, RANDOMFOREST).

Test objectives: From the goals stated in the previous sections, the following objectives are established:

- Improving the prediction accuracy for the classifier models, measured by regarding the correct and incorrect instances
- The numbers of features are reduced to achieve much better accuracy in each classifier
- Reducing the running time of unselected instances and featuress

The first two objectives, high accuracy with least features are the main purposes of this paper, which are shown in the following Figures 5-26. Absolutely the method will save the time of the redundant features that is removed and. The tables show the time to build proposed methodfor different classifiers.

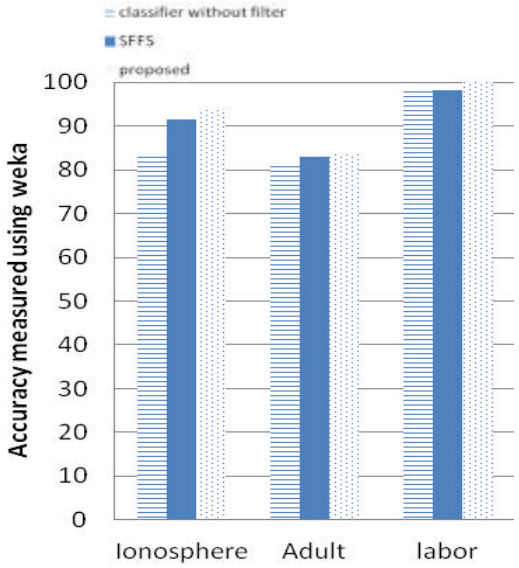

Fig. 3: Naive bayes accuracy of the first experiment

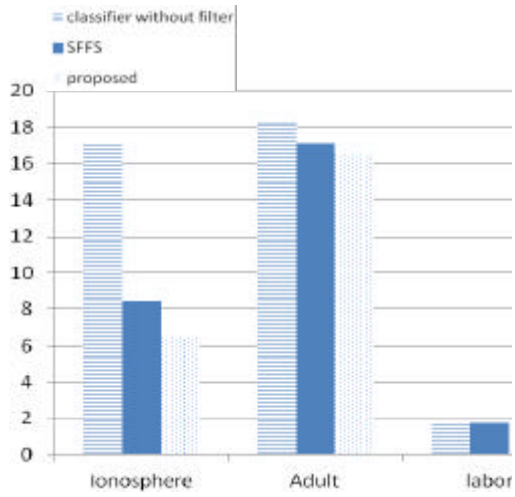

Fig. 4: Naive bayes error rate of the first experiment

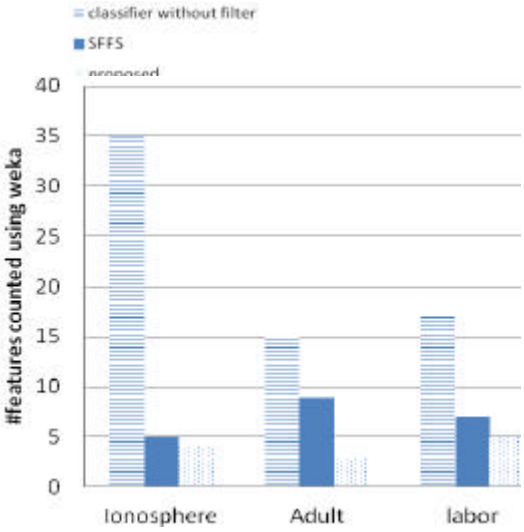

Fig. 5: Naive Bayes accuracy of the first experiment

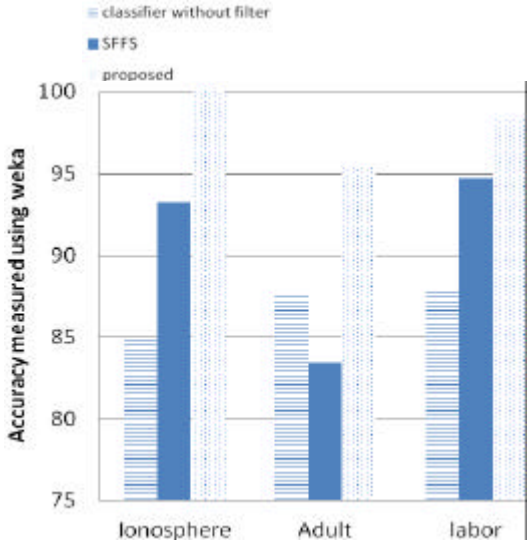

Fig. 6: Naive Bayes error rate of the first experiment

The first experiment: This experiment uses three datasets with different number of features and instances as shown in Table 1. In Fig 3 and 6, the proposed method shows better accuracy in most cases compared with using data without feature selection, wrapper SFFS and the proposed method. In Fig. 5 and 8, the number of features before filtering is compared with the resulted number of features after using the sequential feature selection and using the proposed method, it's obvious that the proposed method reduces the number of features.

Figure 4 and 7 shows a comparison of error rates for NB and J48 and how much it reduces in the proposed method. Figure 10, 11, 13 and 16 introduce the error rates and relative absolute errors for other classifiers compared with NB and J48 using the new subset of features that resulted from the proposed method.

In Fig .9, the proposed method is applied on other variant of Naïve Bayes classifier and the resulting accuracy compared with Naïve Bayes accuracy using the original dataset. The same comparison described in Fig .12 but with J48 and other Decision trees classifiers. Since this experiment uses binary classification the Naïve Bayes and its Variants work better than multi classificationsas in the second experiment.

In this experiment, three different datasets are usedto test the proposed method for two main types of classifiers Bayesian and Trees, forthis part of testing the selected subset of features for each dataset will be illustrated and

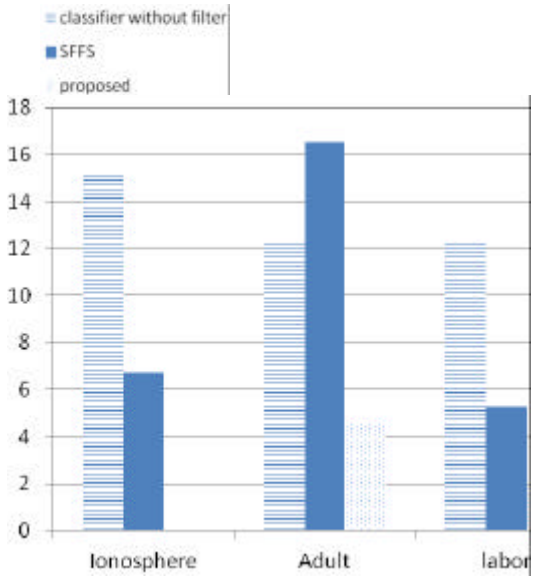

Fig. 7: Numbers of features using NB for the first experiment

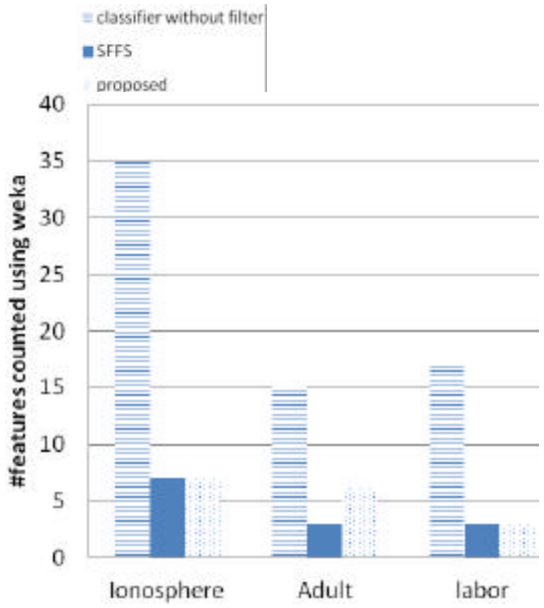

Fig .8: J48 accuracy of the first experiment

the impact of those featuresonthe classifiers measurements (Precision, Recall, F-measure, ROC area and time to build model in seconds) will be described.

The first dataset Ionosphere which contain 35 features described in Table study 1, testedby the proposed method with NBandgenerated new set of 7 features [1,3,4,5,6,14,24] and the impact of those features is calculated on different measurements for three classifiers Naïve Bayes, BNET and NBTREE. The result is shown in Table .2.

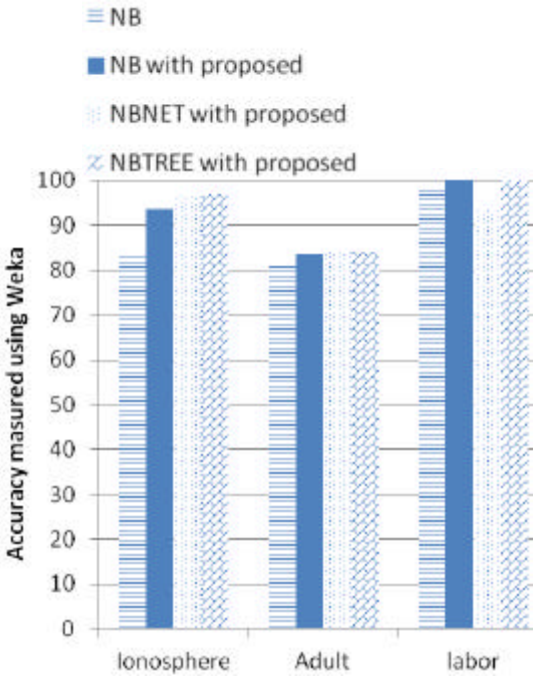

Fig. 9: J48 error rate of the first experiment

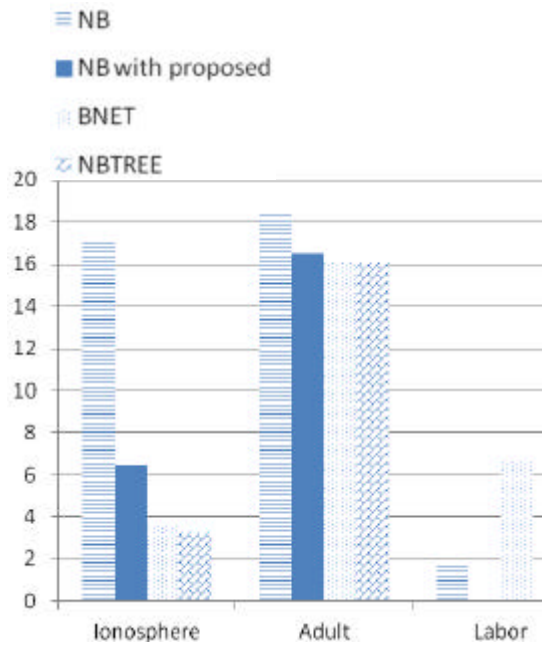

Fig. 10: Numbers of features using J48 for the first experiment

It also produces 7 features [3,12,21,22,23,24,29] when it tested with the proposed method with J48 and the testing results on J48, BFTREE and RANDOMFOREST are described in Table 3sss. The second dataset Adult which contain 15 features described in Table 5, for the

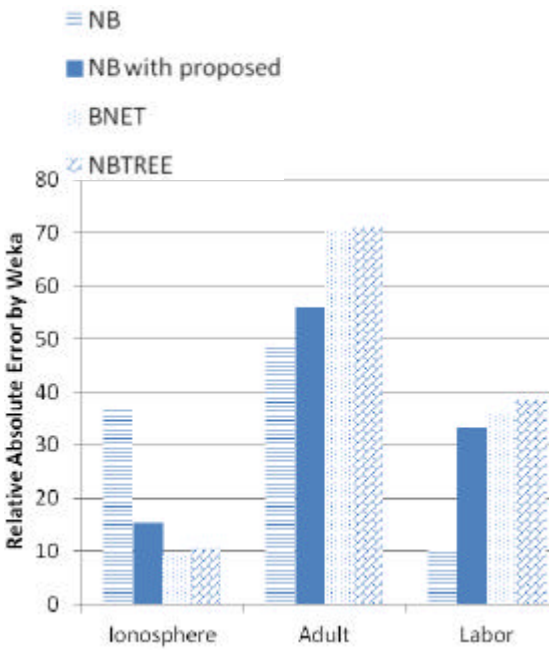

Fig. 11: Classification accuracy of NB and its variants using the proposed method for first experiment

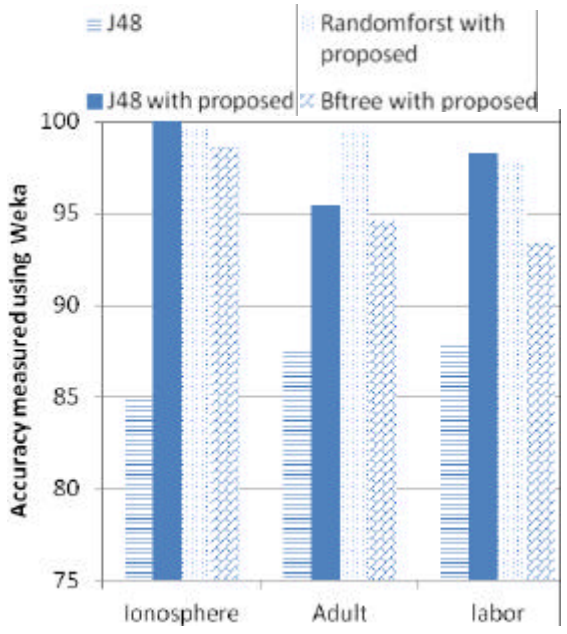

Fig. 12: Error rates of NB and its variants for first experiment

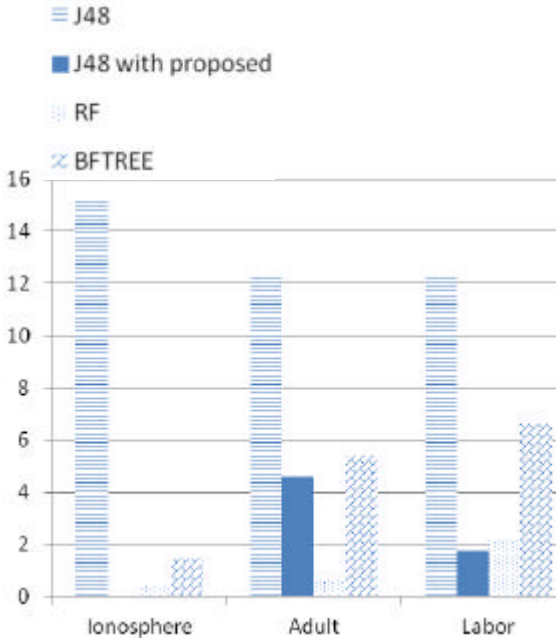

Fig. 13: Relative absolute errors of NB and its variants for the first experiment

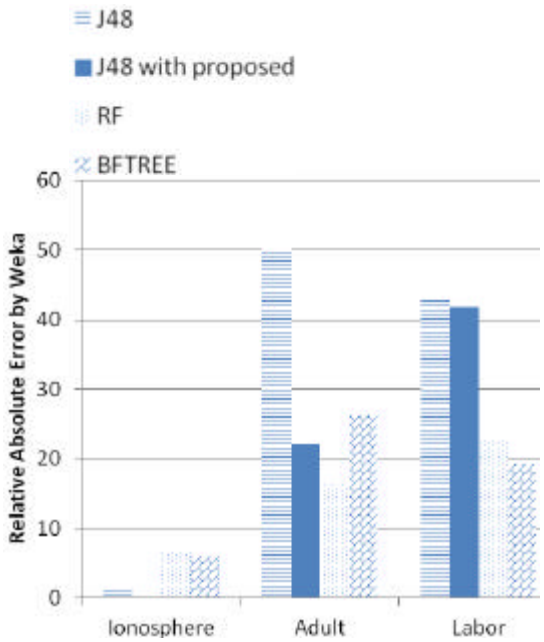

Fig. 14: Classification accuracy of J48 and other trees using the proposed method for the first experiment

proposed method with NB it resultedsubset of 3features [4,11,12] and the impact of those features is calculated on different measurements for three classifiers Naïve Bayes, BNETss and NBTREE.The resultis shown in Table 5. On the other hand, the proposed method with J48 on the Adult dataset produces subset of 6 features [1,3,4,5,7,8] and the testing result on J48, BFTREE and RANDOMFOREST is described in Table 5,6. The last dataset in this experiment is Labor which contain 17 features described in Table .7,the proposed method NB

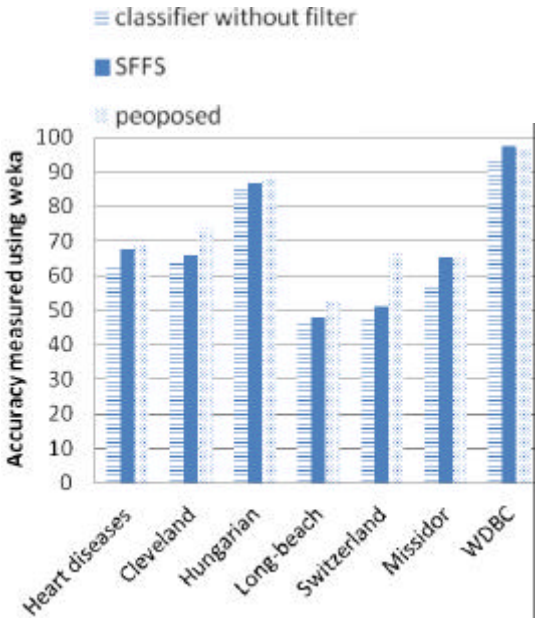

Fig. 15: Error rates of J48 and other trees for the first experiment

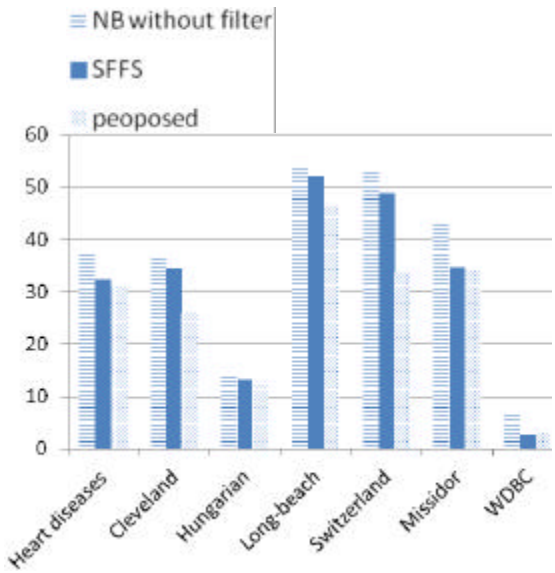

Fig. 16: Relative absolute errors of J48 and other trees for the first experiment

the resulted new subset of 5 features [3,7,10,13,16] and the impact of those features is calculated on different measurements for three classifiers Naïve Bayes, BNET and NBTREE.The result is shown in Table 8. Moreover, the proposed method withJ48on the labor dataset produced subset of 3 features [2,12,14] and the testing

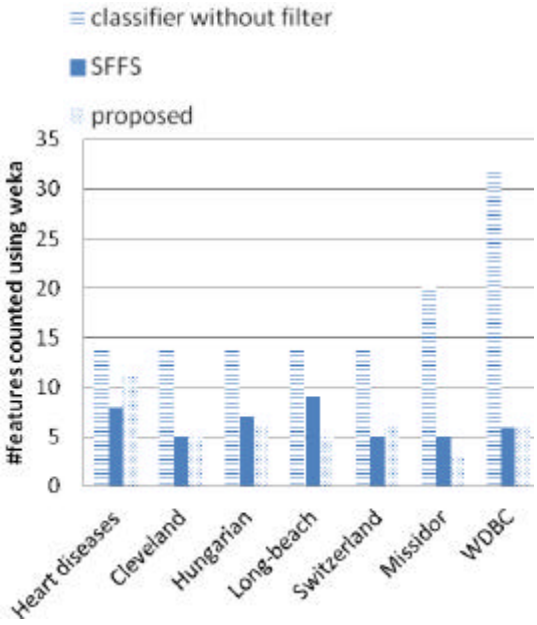

Fig. 17: Number of features using NB for the second experiment

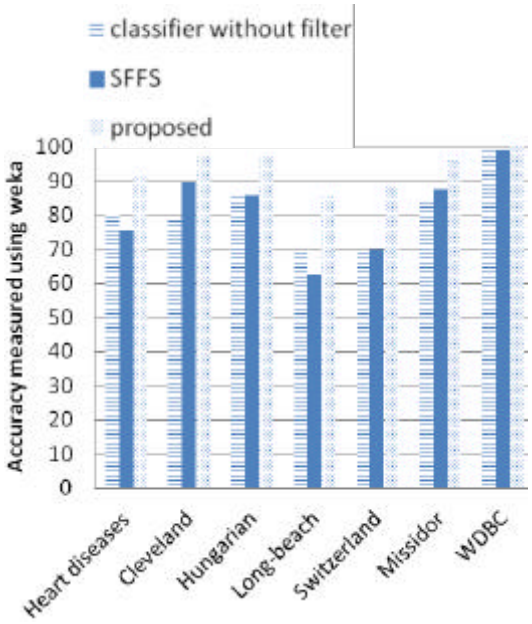

Fig. 18: J48accuracy of the second experiment

result on J48, BFTREE and RANDOMFOREST is described in Table 9. Features atribute are 1-35 of Ionosphere dataset

Features of labor dataset:

- Features
- duration of agreement

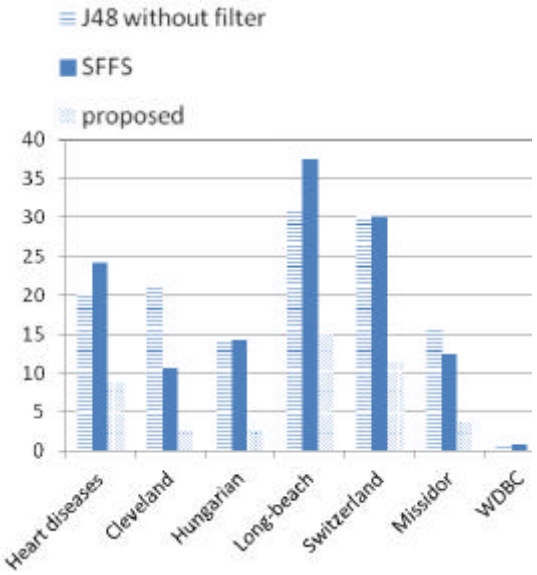

Fig. 19: J48 error rate of the second experiment

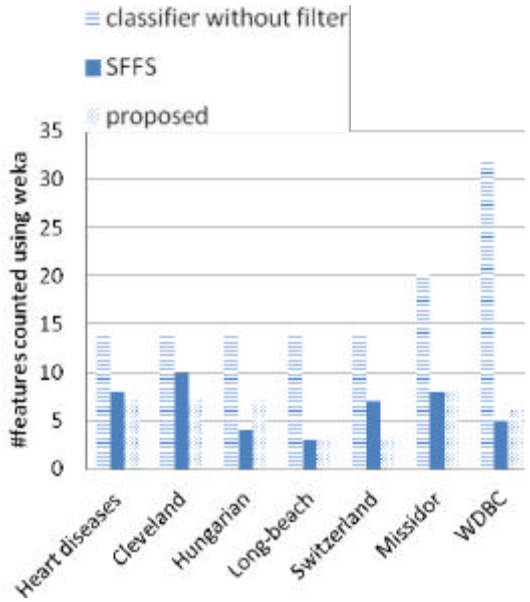

Fig. 20: Number of feature using J48 for the second experiment

- Wage increase in first year of contract
- Wage increase in second year of contract
- Wage increase in third year of contract
- Cost of living allowance
- Number of working hours during week
- Employer contributions to pension plan
- Standby pay
- Shift differential: supplement for work
- Education allowance
- Number of statutory holidays

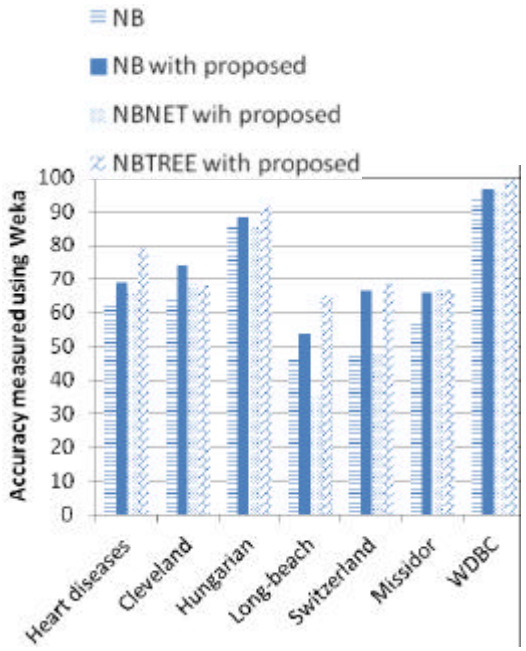

Fig. 21: Classification accuracy of NB and its variants using the proposed method for the second experiment

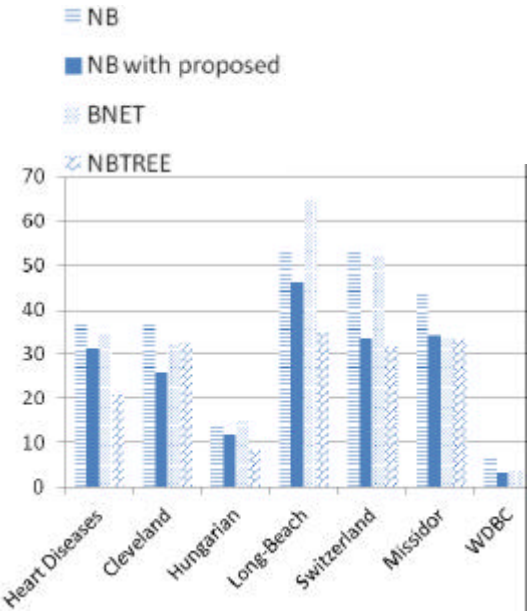

Fig. 22: Error rates of NB and its varients for the second experiment

- Number of paid vacation days
- Employer's help during employee long term disability
- Employers contribution towards the dental plan
- Employer's financial contribution towards the
- Covering the costs of bereavement
- Employer's contribution towards the health plan

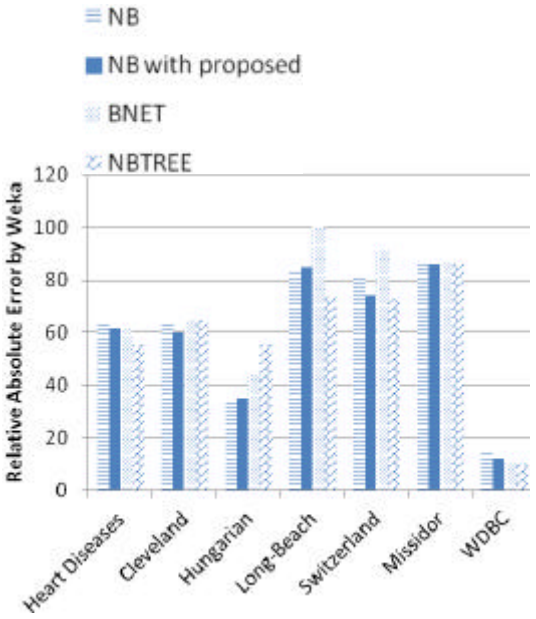

Fig. 23: Relative absolute error of NB and its variants for the second experiment

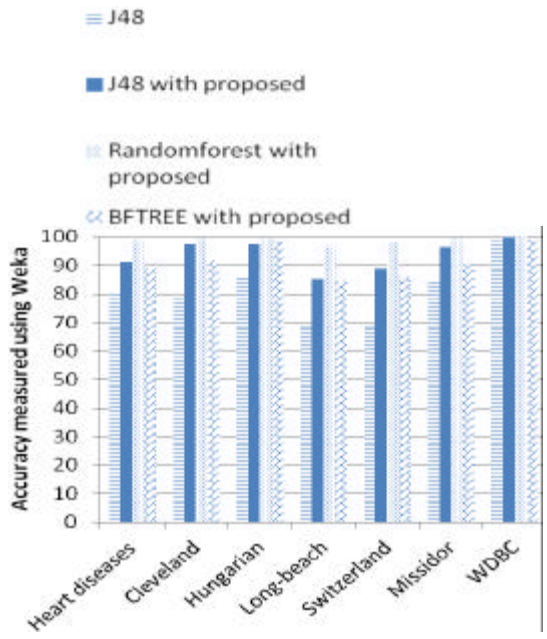

Fig. 24: Classification accuracy of J48 and other trees using the proposed method for the second experiment

The second experiment: In the second experiment, tests are done on the medical datasets in order to get good benefit of the proposed method by improving classification accuracy of different dataset diseases. Three datasets were chosen for different diseases, the

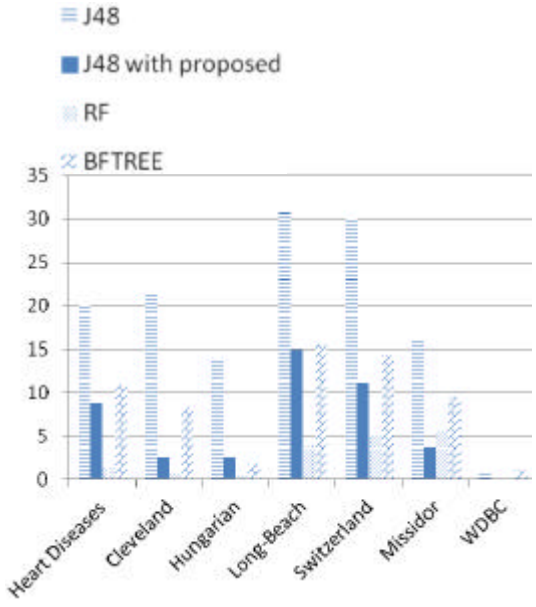

Fig. 25: Error rates of J48 and other for the second experiment

first data set is Heart Diseases database which contains four groups of data from four medical foundations (Cleveland, Hungarian, Switzerland, long beach). The goal is to predicate the presence of heart disease in the patient. The second dataset is Diabetic Retinopathy Debrecen dataset which contains feature extracted from the Missidor image set to predict whether an image contains signs of diabetic retinopathy or not. Finally,the dataset is Wisconsin Diagnostic Breast Cancer features which are computed from a digitalized image of a Fine Needle Aspirate (FNA) of a breast mass. The characteristics of all datasetsare shown in Table 8.

In Fig .15 and Fig .18, the proposed method improves classifiers accuracy for most data set and in the case of Missidor the accuracy reached 96.3043 with 1151 instances. In the features space, the Fig 17. and Fig .20 show how much the features number reduced even when using data set with 32 features. Figures 16 and 21represent that the proposed method reduces the error rates for both NB and J48 classifiers.

In Fig .21 the proposed method is applied on other variant of Naïve Bayes classifier and the resulting accuracy compared with Naïve Bayes accuracy using the original dataset. This comparison is described in Fig .26 but with J48 and other Decision trees classifiers.

In figures 20-26 the error rates and the absolute error are measured for both previous comparisons of NB with its variants and J48 with other tree classifiers.

In this experiment, three main datasets are used to test the proposed method for two main types of classifiers Bayesian and Trees. The first one is Heart diseases

Table 12: Some performance evaluation metrics of the proposed method with J48 on Cleveland dataset

Classifier	Precision	Recall	F-measure	ROC area	# of leaves	Size of tree	Time to build model
J48	0.927	0.926	0.925	0.989	33	65	0.01
BFTREE	0.895	0.893	0.887	0.958	32	63	0.12
RANDOM FOREST	0.988	0.988	0.987	1	-	10	0.03

Table 13: Some performance evaluation metrics of the proposed method with NB on Hungarian dataset

Classifier	Precision	Recall	F-measure	ROC area	Time to build model in seconds
NB	0,88	0.881	0.88	0.889	0.001
BNET	0.848	0.851	0.849	0.855	0.02
NBTREE	0.914	0.915	0.914	0.925	0.17 (leaf=37, size=73)

Table 14: Some performance evaluation metrics of the proposed method with J48 on Hungarian dataset

Classifier	Precision	Recall	F-measure	ROC area	# of leaves	Size of tree	Time to build model
J48	0.975	0.974	0.974	0.975	18	35	0.001
BFTREE	0.983	0.983	0.983	0.998	21	41	0.02
RANDOM FOREST	0.996	0.996	0.996	1	-	10	0.04

Table 15: Some performance evaluation metrics of the proposed method with NB on Long-beach dataset

Classifier	Precision	Recall	F-measure	ROC area	Time to build model in seconds
NB	0.533	0.538	0.519	0.739	0.001
BNET	0.122	0.35	0.181	0.5	0.01
NBTREE	0.632	0.65	0.636	0.866	0.18 (leaf=10, size=19)

Table 16: Some performance evaluation metrics of the proposed method with J48 on Long-beach dataset

Classifier	Precision	Recall	F-measure	ROC area	# of leaves	Size of tree	Time to build model
J48	0.85	0.85	0.849	0.978	37	73	0.01
BFTREE	0.824	0.844	0.833	0.982	43	85	0.04
RAN DO MFOREST	0.964	0.963	0.961	0.997	-	10	0.05

Table 17: Some performance evaluation metrics of the proposed method with NB on Switzerland dataset

Classifier	Precision	Recall	F-measure	ROC area	Time to build model in seconds
NB	0.656	0.663	0.636	0.791	0.001
BNET	0.338	0.48	0.377	0.645	0.09
NBTREE	0.69	0.684	0.655	0.855	0.01 (leaf = 4, size = 7

Table 18: Some performance evaluation metrics of the proposed method with J48 on Switzerland dataset

Classifier	Precision	Recall	F-measure	ROC area	# of leaves	Size of tree	Time to build model
J48	0.864	0.888	0.872	0.975	20	39	0.001
BFTREE	0.835	0.857	0.841	0.965	20	39	0.01
RANDO MFOREST	0.980	0.98	0.98	0.996	-	10	0.02

dataset which is tested with all its instances and tested again in four separated datasets (Cleveland, Hungarian, Long-beach and Switzerland) as described in Table 11. In this part, the selected subset of features for each dataset will be illustrated and how do those features affect the classifiers measurements (Precision, Recall, F-measure, ROC area and time to build model in seconds).

The first dataset Heart diseases which contain 14 features described in Table .12, for the proposed method with NB this dataset resulted new subset of11 features[1,3,4,5,6,8,9,10,11,12,13].The impact of those features is calculated on the different measurements for three classifiers Naïve Bayes, BNET and NBTREE.The result is shown in Table .9 On the other hand, the proposed method with J48 classifier produces 7 features[1,2,3,4,5,7,8], and the testing results on J48, BFTREE and RANDOMFOREST is described in Table 10.

Cleveland is subset of Heart diseases dataset(has the same original set of features). Forthe proposed method with NB, it generated new subset of 5 features [6,9,10,12,13] and the impact of those features is calculated on different measurements for three classifiers Naïve Bayes, BNET and NBTREE are shown in Table 15. Moreover, the proposed method with J48 on the Cleveland dataset produced subset of 7 features [3,4,5,8,9,11,12], the testing result on J48, BFTREE and RANDOMFOREST is described in Table 12.

Hungarian which is another subset of Heart diseases dataset has the same original set of features. in the proposed methodwith NB, it resulted new subset of 6 features [1,5,9,10,11,13] and the impact of those features is calculated on different measurements for three classifiers Naïve Bayes, BNET and NBTREE are shown in sTable .17. The proposed method with J48on the Hungarian dataset produced subset of 7 features [1,2,3,4,5,9,10], and the testing result on J48, BFTREE, and RANDOMFOREST is described in Table 4.

Long-beach is another subset of Heart diseases dataset. For the proposed method with NB,it resulted new subset of 5 features [1,2,3,4,11] and the impact of those features is calculated on different measurements for three classifiers Naïve Bayes, BNET and NBTREE are shown in Table 15. The proposed method with J48 on the Long-beach dataset produced subset of only 3 features [1,5,8], the testing results on J48, BFTREE and RANDOM FOREST are described in Table 16.

Switzerland which is the last subset of Heart diseases dataset (the same original set of features) for the proposed method with NBit resulted new subset of 6 features [7,8,9,10,12,13] and the impact of those features is calculated on different measurements for three classifiers Naïve Bayes, BNET and NBTREE are shown in Table .17. Moreover, the proposed method with J48 on the Switzerland dataset produced subset of only 3 features [4,8,10] the testing results on J48, BFTREE, and RANDOMFOREST are described in Table .18

Missidor (Diabetic Retinopathy Debrecen) dataset which contain 20 features described in Table .23, tested

Table 19: Some performance evaluation metrics of the proposed method with NB on Missidor dataset

Classifier	Precision	Recall	F-measure	ROC area	Time to build model in seconds
NB	0.667	0.658	0.654	0.686	0.01
BNET	0.691	0.667	0.658	0.681	0.01
NBTREE	0.691	0.667	0.658	0.681	0.05 (leaf=1, size=1)

Table 20: Some performance evaluation metrics of the proposed method with J48 on Missidor dataset

Classifier	Precision	Recall	F-measure	ROC area	# of leaves	Size oftree	Time to build model
J48	0.854	0.842	0.841	0.935	58	115	0.06
BFTREE	0.925	0.924	0.924	0.971	83	165	0.11
RANDOMFOREST	0.993	0.993	0.993	1	-	10	0.15

Table 21: Some performance evaluation metrics of the proposed method with NB on WDBC dataset

Classifier	Precision	Recall	F-measure	ROC area	Time to build model in seconds
NB	0.967	0.967	0.967	0.988	0.01
BNET	0.96	0.96	0.96	0.988	0.01
NBTREE	0.996	0.996	0.996	0.9997	0.33 (leaf=8, size=15)

Table 22: Some performance evaluation metrics of the proposed method with J48 on WDBC dataset

Classifier	Precision	Recall	F-measure	ROC area	# ofleaves	Size oftree	Time to build model
J48	1	1	1	1	15	29	0.02
BFTREE	0.989	0.989	0.989	.999	13	25	0.03
RANDOMFOREST	1	1	1	1	-	10	0.04

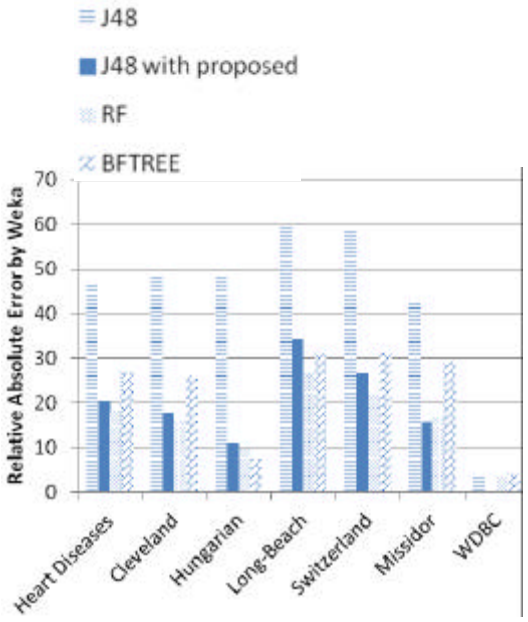

Fig. 26: Relative absolute errors of J48 and other trees for the second experiment

using the proposed method with NB and generateda new subset of 3 features [2,8,16]. The impact of those features is calculated on different measurements for three classifiers Naïve Bayes, BNET and NBTREE are shown in Table .19, the proposed method with J48 on this dataset produced subset of 8 features [1,2,4,9,11,12,13,14]. The testing results on J48, BFTREE and RANDOMFOREST are described in Fig. 26

WDBC (Wisconsin Diagnostic Breast Cancer) dataset which contain 32 featuresdescribed in Table .20, testedusing the proposed method with NBand produced new subset of 6 features [4,10,17,19,24,25].The impact of those features is calculated on different measurements for three classifiers Naïve Bayes, BNET and NBTREE are shown in Table .21. The proposed method with J48 produced subset of 6 features [4,9,26,27,28,29] and the testing results on J48, BFTREE and RANDOMFOREST are described in Table 22.

Features of heart diseases dataset:

- Features
- Age in years
- Sex (1 = male; 0 = female)
- Chest pain type
- Resting blood pressure
- Serum cholesterol in mg/dl
- Fasting blood sugar > 120 mg/dl
- Resting electrocardiographic results
- Maximum heart rate achieved
- Exercise induced angina (1 = yes; 0 = no)
- ST depression induced by exercise relative to rest
- The slope of the peak exercise ST segmentTable .12 (continued)
- Number of major vessels (0-3) colored by fluoroscopy
- Thal: 3 = normal; 6 = fixed defect; 7 = reversible defect
- Diagnosis of heart disease (angiographic disease status)

Features of missidor dataset:

- Features
- The binary result of quality assessment. 0 = bad quality 1 =sufficient quality
- The binary result of pre-screening, where 1 indicates severe retinal
- Abnormality and 0 its lack

- The results of MA detection. Each feature value stand for the number of MAs found at the confidence levels alpha = 0.5, . . . , 1,respectively.
- Contain the same information as 2-7) for exudates. However, as exudates are represented by a set of points rather than the number of pixels constructing the lesions, these features are normalized by dividing the number of lesions with the diameter of the ROI to compensate different image sizes
- The Euclidean distance of the center of the macula and the center of the optic disc to provide important information
- The diameter of the optic disc
- The binary result of the AM/FM-based classification
- Class label. 1 = contains signs of DR (Accumulative label for the Messidor classes 1, 2, 3), 0 = no signs of DR

Features of wdbc dataset

- Features
- ID number
- Diagnosis (M = malignant, B = benign)

3-32

- Ten real-valued features are computed for each cell nucleus:
- Radius (mean of distances from center to points on the perimeter)
- Texture (standard deviation of gray-scale values)
- Perimeter
- Area
- Smoothness (local variation in radius lengths)
- Compactness (perimeter^2 / area - 1.0)
- Concavity (severity of concave portions of the contour)
- Concave points (number of concave portions of the contour)
- Symmetry
- Fractal dimension (“coastline approximation” - 1)

CONCLUSION

In this study, combinations of processes are proposed in three phases that tries to find the optimal subset of features.In the first phase, the samples are filtered by Resample filter technique. In the second phase, the method tries to find the optimal feature space by applying classifier subset evaluator with Sequential forward floating selection algorithm to remove the irrelevant features. In the third phase, the proposed method is tested on different sizes datasets and resulted new feature subsets that are used in more than one classifier to evaluate their measurements in (precision, recall, f-measure, ROC area and time to build model).

The proposed method is illustrated on different purposes datasets as well as on medical datasets and in most cases the experiments gave good results in reducing the feature space and enhancing the classifiers performance.

REFERENCES

Bermejo, P., L. de la Ossa, J.A. Gamez and J.M. Puerta, 2012. Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking. Knowledge-Based Syst., 25: 35-44.

Bidgoli, A.M. and M.N. Parsa, 2012. A hybrid feature selection by resampling, chi-squared and consistency evaluation techniques. World Acad. Sci. Eng. Technol., 68: 276-285.

Chandrashekar, G. and F. Sahin, 2014. A survey on feature selection methods. Comput. Electr. Eng., 40: 16-28.

Liu, H., H. Motoda and L. Yu, 2004. A selective sampling approach to active feature selection. Artif. Intell., 159: 49-74.

Naseriparsa, M., A.M. Bidgoli and T. Varaee, 2014. A hybrid feature selection method to improve performance of a group of classification algorithms. Comput. Sci., 69: 28-36.

Patil, T.R. and S.S. Sherekar, 2013. Performance analysis of naive bayes and J48 classification algorithm for data classification. Int. J. Comput. Sci. Appl., 6: 256-261.

Peng, Y., Z. Wu and J. Jiang, 2010. A novel feature selection approach for biomedical data classification. J. Biomed. Inf., 43: 15-23.

Sasikala, S., A.S.A. Balamurugan and S. Geetha, 2014. Multi filtration feature selection (MFFS) to improve discriminatory ability in clinical data set. Appl. Comput. Inf., 12: 117-127.

Srimani, P.K. and M.S. Koti, 2013. Medical diagnosis using ensemble classifiers-a novel machine-learning approach. J. Adv. Comput., 1: 9-27.

Vanaja, S. and K.R. Kumar, 2014. Analysis of feature selection algorithms on classification: A survey. Int. J. Comput. Appl., Vol. 96,

Way, T.W., B. Sahiner, L.M. Hadjiiski and H.P. Chan, 2010. Effect of finite sample size on feature selection and classification: A simulation study. Med. Phys., 37: 907-920.

Win, M.T.M. and K.T. Khaing, 2014. Performance evaluation of different classifiers for detection of attacks in unauthorized accesses. Int. J. Sci. Eng. Technol. Res., 3: 2890-2894.

Xie, J. and C. Wang, 2011. Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. Ex. Syst. Appl., 38: 5809-5815.