Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2019) 27: 1290 – 1307
© TÜBİTAK
doi:10.3906/elk-1808-85

An improved tree model based on ensemble feature selection for classification

Chandralekha MOHAN[*], **Shenbagavadivu NAGARAJAN**
Department of Computer Applications, University College of Engineering, Bharathidasan Institute
of Technology Campus, Anna University, Tiruchirappalli, Tamil Nadu, India

Received: 09.08.2018 • **Accepted/Published Online:** 30.12.2018 • **Final Version:** 22.03.2019

Abstract: Researchers train and build specific models to classify the presence and absence of a disease and the accuracy of such classification models is continuously improved. The process of building a model and training depends on the medical data utilized. Various machine learning techniques and tools are used to handle different data with respect to disease types and their clinical conditions. Classification is the most widely used technique to classify disease and the accuracy of the classifier largely depends on the attributes. The choice of the attribute largely affects the diagnosis and performance of the classifier. Due to growing large volumes of medical data across different clinical conditions, the need for choosing relevant attributes and features still lacks method to handle datasets that target specific diseases. This study uses an ensemble-based feature selection using random trees and wrapper method to improve the classification. The proposed ensemble learning classification method derives a subset using the wrapper method, bagging, and random trees. The proposed method removes the irrelevant features and selects the optimal features for classification through probability weighting criteria. The improved algorithm has the ability to distinguish the relevant features from irrelevant features and improve the classification performance. The proposed feature selection method is evaluated using SVM, RF, and NB evaluators and the performances are compared against the FSNBb, FSSVMb, GASVMb, GANBb, and GARFb methods. The proposed method achieves mean classification accuracy of 92% and outperforms the other ensemble methods.

Key words: Machine learning, classification, wrapper method, bagging, feature selection, SVM, RF, NB, FSNBb, FSSVMb, GASVMb, GANBb, GARFb

1. Introduction

The health care industry comprises records in an electronic format, which often includes prescriptions, consultation notes, images, clinical data, drugs, insurance, and other administrative records. Big data offers the potential to diagnose and discover patterns and trends, enabling health care providers to improve patient care, cost management, and decision-making. Using data mining techniques, health care providers can analyze patient data to detect a disease, predict the future likelihood, find hidden information and patterns that can effectively benefit patients, minimize costs, and offer better health management.

With compliance to regulatory requirements, patient care and record keeping drive hospitals to digitize their data and store vast amounts of records. It is estimated that about 200 exabytes of medical data in the health care industry exist and will continue to increase in the future. Health care organizations and specialists have understood the importance of the utilization of massive health records to offer better health management, clinical decisions, and observations of diseases with respect to geographical boundaries. With a massive amount

[*]Correspondence: rdbmchandralekha@gmail.com

of data constantly growing, the health care industry often faces problems with respect to appropriate health care applications and tools due to the diversity of the data types available.

The medical industry uses machine learning and data mining techniques for applications such as drug discovery, disease classification and diagnosis, disease outbreaks, decision-making, personalized health care, and many more. However, these machine learning applications suffer from large dimensions of data, data handling, computation time, choice of prediction methods, and classification accuracy. The performance of any classifier depends on the attributes selected and its relevancy either through supervised or unsupervised methods. To improve the classifier's performance, researchers primarily use preprocessing steps, which involves handling missing values, imputations, data outliers, etc. [1].

Preprocessing can be classified into three main groups, which are data cleaning, data aggregation, and data transformation. In addition to preprocessing, researchers included selecting the best and most relevant features for prediction and classification. The authors of [2] concluded that feature selection can be effective in removing features that can be noise in the classification model. Optimal features can help a classifier to reduce the complexity of building and training the model, thus improving the classification accuracy [3]. Feature selection methods are of two main groups that are distinguished based on the classifier. In the two main types of feature selection, filter methods are free from the classifier while embedded and wrapper methods [4] depend on the classifier to select the best features with respect to the accuracy of the classifier. There also exists a third type of method, which combines different algorithms and methods into a new approach called the hybrid approach [5].

Features, on the other hand, describe the vulnerability to certain disease conditions. Features are the symptoms that are associated with the disease, and removing or ignoring a feature can affect the overall treatment process. Utilizing every feature to make decisions and waiting for such features cannot help patients in emergency care, however. In order to provide a quick diagnosis and treatment with minimal information, disease diagnosis tools and decision support systems are best suited for medical experts. Also, instead of treating patients effectively, predicting and diagnosing diseases early can save lives, time, and costs. In this study, the process of selecting the best features using a wrapper and random trees in which the internal evaluating feature's importance is motivated by the random forest method [6] is proposed. The subset generated by the wrapper algorithm has immense probability of overfitting. In order to reduce the overfitting, random trees are constructed for the subset generated by the wrapper algorithm, and through an ensemble model of random trees, the best features are selected to improve the classification performance. The proposed method has the ability to distinguish the irrelevant and relevant features by building the best trees from the wrapper subset. The performance of the classification is compared against different models using 15 different datasets. This paper is organized into 5 sections. Following the introduction, Section 2 presents a literature review, Section 3 discusses the proposed method, Section 4 discusses the experimental study, and Section 5 presents the results and conclusion.

2. Literature review

Feature selection is performed through three distinct approaches: filter, wrapper, and embedded approaches. Combining these approaches, researchers used different algorithms to build subsets or to remove the redundant features in disease diagnosis. Chandralekha et al. [7] studied the performance of various algorithms to predict cardiovascular diseases and emphasized that machine learning algorithms play a major role in disease diagnosis and in identifying hidden patterns across different disease conditions, through which the disease conditions can be predicted earlier, reducing treatment time and cost. With respect to specific decision goals and data types, the feature selection methods and procedures vary largely; however, the goal of feature selection is to find the relevant features and subsets that are more relevant in defining the target classes. Compared to traditional feature selection methods, an ensemble method has more power to generalize and improve accuracy [8].

Hoque et al. [9] proposed an ensemble feature selection method using mutual information based on feature-feature and feature-class for creating relevant features and unneeded features. The subset is arrived at by combining 5 different methods. ReliefF, gain ratio, info gain, chi-square, and symmetric uncertainty were combined to extract the feature subsets for feature-feature and feature-class entropy. The proposed feature selection method outperformed individual models for high-dimensional datasets.

Settouti et al. [10] proposed semisupervised ensemble learning using an optimized co-forest and permutations from a random forest. The study involved using unlabeled samples to improve the classification performance. The proposed method uses two approaches; the first approach selects the best subset and the second approach measures the feature importance. The proposed method is compared with bagging score, constrained Laplacian score, spectral analysis, and SEFR methods using UCI and ASU datasets. The proposed method outperforms other methods by using unlabeled samples.

Filali et al. [11] proposed unsupervised feature selection based on an ensemble method, which combines bagging and random subspace on k-means variants. The proposed algorithm, RFS-KM, outperforms RFS-FKM, RFS-IRP-K, and RFS-IRP-Kmeans for different datasets. The proposed algorithm uses clustering ensembles using out-of-bag importance from the traditional RF method. The method targets the clustering quality that matches the known classes. The clustering accuracy was given by NMI rate, which corresponds to the number of true variables in the dataset.

Chaudhary et al. [12] proposed an improved random forest classifier for multiclass problems. Using an attribute evaluator, feature selector, and gain-ratio, the performance of the random forest classifier is improved. The random sampling method is applied to address class balance. Using symmetrical uncertainty and correlation-based feature selection improved the random forest classifier's performance. Compared with the traditional random forest method, the improved random forest showed an improved classification accuracy rate of 97.80%.

Sakri et al. [13] proposed a feature selection technique to predict breast cancer recurrence. The feature selection method uses particle swarm optimization embedded with three different classifiers, namely KNN, NB, and fast decision tree. Out of 34 features, the proposed method selects the best four features, and on the testing set, the classification performance was improved for all three classifiers chosen. The performance of KNN is improved from 70% to 81%, NB performance improved from 76% to 80%, and fast decision tree improved from 66% to 75% while using PSO-based feature selection. Also, NB performed better when compared to KNN and fast decision tree.

Cai et al. [14] proposed a new feature evaluation for multilabel classification using a neighborhood relationship-preserving score. The proposed study uses linear estimates of similarities of the neighbors and, based on the similarities, measures the importance of each feature present in the dataset through greedy steps and ranking methods. A study conducted on the six datasets shows that the proposed feature selection method outperforms other multilabel feature selection methods such as LP + RF, PMU, MDMR, and FIMF.

Brankovici et al. [15] proposed a new classification method by combining feature selection and classification. The method uses expansion of attributes and a selection method through refining the probability distribution and combining the rating of different model groups and using distance correlation filtering to reduce the features. The features that are independent of the model results are removed. The proposed RFSC outperforms other methods on six datasets.

Park et al. [16] proposed a hybrid feature selection method to diagnose hypertension using SU and a Bayesian network. Using a wrapper approach, the selection of features is done through a backward search method and correlation between features, a subset is built, and the importance of the features in the subset is measured using the Bayesian network. The proposed feature selection outperforms other methods like IG, gain ratio, and ReliefF with higher accuracy of 92%.

Agre et al. [17] proposed a new feature weighting scheme based on a user-defined threshold on variability of the features. The proposed method targets improvement of ReliefF to a more robust feature selection strategy through aggregating with principal components. Initially the weights from PCA and ReliefF are calculated and are combined with PCA through the threshold. The weights are transferred into explained variability and the features are selected. A study conducted on twelve different datasets shows that the proposed method can be used for feature reduction and feature selection, which improves the classification accuracy.

Osanaiye et al. [18] proposed an ensemble-based multifilter feature section method for intrusion detection. The multifilter feature selection method utilizes four different filter methods and the features are selected based on the results of four filtering methods. The study uses information gain, gain ratio, chi-square, and relief for filtering the features. The selected filtering method ranks the features and, based on majority vote, a subset of features are selected with a threshold. CFS, gradual feature removal, and linear correlation-based methods are compared against the proposed method using different classifiers. The proposed ensemble method achieved an accuracy of 99% with 13 features on the NSL-KDD dataset.

Shunmugapriya et al. [19] proposed a new method to select features using the properties of an ant colony and artificial bee colony, as both algorithms are known for their metaheuristic searches, and using their unique optimization capabilities, the feature selection process is improved. The result of the new AC-ABC algorithm outperforms other state-of-the-art methods, which improved the classification accuracy. A study conducted on 13 different datasets proved that the new model is efficient in reducing the features with low computational complexities.

Koutanaei et al. [20] investigated different feature selection algorithms and an ensemble of classifiers to achieve high performance results. The study employed different feature selection methods such as GA, IG, PCA, and relief with SVM as a base classifier. The study concluded that PCA and ANN-Adaboost are best for the feature selection method to improve accuracy and are recommendeded for classifying credit scoring problems.

Sasikala et al. [21] proposed a feature selection method using extraction, subset selection, feature reranking, and classification to effectively select best features. The MMFS algorithm involves PCA to extract important features, CFS to create subsets, and SU to rerank features by removing the biased features and normalization. Later the selected features are classified using SVM, MLP, DT, and NB methods. The proposed feature selection method improved the accuracy of all the classifiers selected.

Ebrahimpour et al. [22] proposed a novel method based on MRMR for selecting best features. The proposed algorithm is based on the CFS algorithm that filters the features using ensembles of ranking methods. The developed MRMR method is compared against CFS, FCBF, and INT and the results of the study confirm that the ensemble model is capable of selecting quality subsets from high-dimensional datasets and yields better numbers of features with high accuracy and sensitivity.

Tseng et al. [23] proposed a new method of detecting risk factors to detect ovarian cancer recurrence. The method ranks the features using five algorithms, namely SVM, C5.0, ELM, MARS, and RF. The overall importance of the features is measured through ensemble learning and the selected risk factors are applied to the classifiers. The selected risk factors improve the classification accuracy of all the algorithms. C5.0 was found to perform better than the other algorithms when applied to a real-world dataset.

Kamkar et al. [24] proposed a supervised feature selection method on tree structures. The tree-lasso-based method outperformed other selection methods like T-test, IG, ReliefF, and Lasso. Also, the classification based on tree-lasso was superior to LR, NB, SVM, DT, and RF. The study concluded that the proposed method can be used to model potential risk factors.

Lu et al. [25] developed a new algorithm incorporating mutual information maximization and an adaptive genetic algorithm to select the informative gene expression and remove the redundant data. When applied on a real-world dataset, the proposed method performs well against MIM, ReliefF, and SFS. The classification of using the filtered gene sequence was robust when compared against BP, SVM, ELM, and RELM. The method is capable of reducing the gene data and also maintains high classification accuracy.

Liu et al. [26] proposed a feature selection method deriving from the maximum nearest neighbor, an information theory-based approach to extract the features that are highly informative. The proposed algorithm is compared against four different feature selection methods and classifiers such as CART, LSVM, and KNN. The results show that the proposed feature selection method achieves better results than other feature selection methods.

Vivekanandan et al. [27] proposed a new feature selection method based on an evolutionary search algorithm. The modified differential evolution algorithm (DE) is compared against integrated fuzzy AHP and a feedforward network. The traditional DE and modified DE algorithms are compared using heart disease. The results demonstrated that the modified DE (83%) outperforms traditional DE with respect to feature selection and processing time against the BPO + Rough Set and NB (79%), BPO + Rough Set + SVM (75%), and firefly + Rough Set + ANN (81%) in classifying heart disease with selected features.

Bellal et al. [28] proposed a new method using unlabeled data in a guided feature selection method called supervised ensemble learning guided feature ranking (SEFR). The proposed SEFR was compared against RF, FW + SemiFS, and sSelect method using datasets taken from the UCI repository. The proposed method achieved ranking of permuted features correctly so that the classification performance was highly improved.

Hong et al. [29] proposed a new feature selection algorithm using clustering ensembles and population-based incremental learning. The different clusters were transformed into a similarity matrix and the subsets of feature were selected by combining the subset similarity scores. Using the PBIL search method, the subsets of features were selected using K-means as the base cluster. Compared with traditional K-means, the proposed CEFS method outperforms with 93% accuracy.

Yang et al. [30] proposed a novel method to address the problem of solving high-dimensional data and to select the best features. The algorithm is a wrapper method built with random forest-based feature importance and SVM. Using forward selection and backward selection, the random forest is trained to compute the feature scores. The selected features are classified using SVM as a base classifier. The proposed method achieved higher accuracy on datasets taken from the UCI repository.

Maldonado et al. [31] proposed a wrapper-based feature selection method using a backward selection method. The features are selected based on the error coefficient and compared against four different SVM variants of HO-SVM, RFE-SVM, SVM-L, and SVM-poly. HO-SVM outperformed the other methods and the study highly recommended the proposed work for high-dimensional problems.

Zheng et al. [32] proposed a new method to extract tumor features for detecting breast cancer. The hybrid feature selection approach combines k-means and the SVM algorithm. The k-means algorithm builds a new feature based on the tumor characteristics, i.e. benign and malignant. The extracted feature is used in the classification process using SVM. The ensemble model produced an accuracy of 97% and extracted six features with the tumor dataset.

Zhou et al. [33] proposed a cost-based feature selection while constructing decision trees. The proposed model selects the features with respect to low cost. The cost-sensitive feature ranking method is evaluated on 10 different datasets and compared against SVM and RF. The proposed method outperforms with a total low cost of 42 compared to other methods.

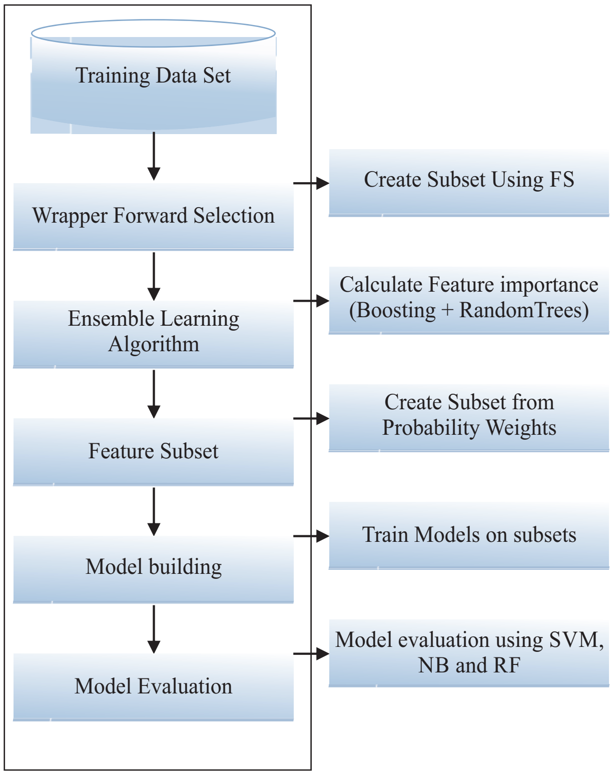
Training Data Set
Wrapper Forward Selection
Ensemble Learning Algorithm
Feature Subset
Model building
Model Evaluation
Create Subset Using FS
Calculate Feature importance (Boosting + RandomTrees)
Create Subset from Probability Weights
Train Models on subsets
Model evaluation using SVM, NB and RF

Figure 1. Proposed method.

3. Proposed method

Feature selection is a process of creating a subset of features aimed to improve classification results and performance. Feature selection is used for a variety of reasons, such as data reduction, reducing time complexity, and improving performance and accuracy. Feature selection often refers to selecting the best and optimal features related to the target classes. The main objective of this study is to find the best relevant features for a given dataset while irrelevant features are eliminated to improve the classification accuracy. The general scheme of the proposed model is shown in Figure 1. The names and the symbols used are given in Table 1. Irrelevant features often increase the error rates and, to reduce the classification error rate, features that are highly informative are selected for classification. For X dataset with Y features, FS creates a Y-dimensional space of S^{Y} in which a subspace i from $\mathrm{S}^{\mathrm{Y}i}$ comprises the most relevant features that explain the target class. The common

Table 1. List of symbols.

Symbol	Description
$X=(x_1,x_2,...,x_d)$	X is a vector of features
y	y is the target class
E	Represents the ensemble of decision tree classifiers
P_k	Hyperparameters of a classifier
$X_{sub} = ((X_1,y_1)...,(X_n, y_n))$	Subset vector of features
$CR(X,C_1) = \frac{v(X,C_1)}{RK}$	Class relationship score of a feature
$C^i=mvote\{C^{Rk}\}k_1$	Majority vote of random trees
R_k	Random trees
O_k	Subset with k features
b^+	A feature that contributes to classifier performance
C	Stopping criterion
S^y	Y in feature subspace

feature selection includes the filter, embedded, and wrapper methods. Wrapper methods optimize the learning iteratively through a greedy search and use algorithms for rating criteria. The rating criteria ensure that the selections are in tune with the algorithms' performances. The search strategy forms a permutation of different features Y in the feature space S^Y and stopping criterion C is used to stop producing subsets. The nearest neighbor is used as the base classifier for selecting the best feature subset based on the higher performance. The resulting subset is used to build random trees in order to further avoid high correlation between the variables and achieve generality.

Algorithm 1

```
Input: $X=(x_1,x_2,...,x_d)$ (the whole dimensional space with $S^Y$)
Output: $O_k = \{b_i | i=1, 2,..., k; b_i \in X\}$ where $k = (0, 1, 2,..., S^Y)$
$O_k$ returns a subset with k features, Where $k < S^Y$
Initialize selection: $O_k$ = empty, k = 0 (k is the feature size)
Step 1:
$b^+=$ arg max I( $b_k$+ b), where $b \in X-O_k$
$O_k$ +1= $O_k$ + $b^+$
k = k + 1
Go to Step 1
$b^+$ is the feature that adds to the performance of the classifier and added to $O_k$
repeat until C is reached
Go to 1
```

The most powerful learning ensemble tree model is based on classification trees that build trees through 2 random steps. The first step takes bootstrap samples and grows trees on each single feature. The leaf node consists of features and its corresponding members in the intermediate nodes. The second step adds a new tree to the bottom of the existing trees. Each tree is a kind of class preference to the target class. Trees that have the most class preferences are classified into the target class, respectively.

The ensemble classifier is built using random classification trees, through which individual trees are grown for each variable by combining multiple bootstrapped samples Bt from the training set. Let X= $(x_1,x_2,...,x_d)$ be a vector of features. Each feature x_1 is a random variable in the input vector X. Let y be the classification of disease (y = 0 indicates absence and y = 1 indicates presence). Now features from X predict y through ensembles of the classifiers E such that E = $(E_1(X),...,E_k(X))$. Every $E_1(X)$ is a decision tree and we denote the hyperparameters of a decision tree for the classifier $E_1(X)$ as $P_k=(P_{k1},...,P_{kn})$. The decision tree classifier is given by $E_k(X) = E(X|P_k)$. Each tree based on hyperparameters P_k casts votes to class y in the input vector X, the class with highest vote gains.

P_k determines the subset of trees X_{sub} based on the hyperparameters, and then the vector of the subset is denoted by $X_{sub} = ((X_1,y_1)...,(X_n, y_n))$. The probability of class C_1 for the ensemble $E_k(X)$ is given by the occurrence of C_1 for the classifier $E_k(1\leq k \leq K)$. The class relationship score CR is computed through votes and the number of trees in the forest. It can be represented as

$$\mathrm{CR}(\mathrm{X},\mathrm{C}_1) = \frac{v(X,C_1)}{RK}.$$

The feature selection score Sr can be given by Sr = $\Sigma z(yi, C(X_i)) \times CR(X,C_1))$, where z = (0 for i $\neq$ j and 1 otherwise) and C(X) is the class label assigned by the classifier $E(X|P_1)$. The majority vote of a tree among a number of trees in Rk is given by

$$\mathrm{C}^{\mathrm{i}} = \mathrm{mvote}\{\mathrm{C}^{\mathrm{Rk}}\}\mathrm{k}_1.$$

The next step is to calculate the feature importance, which lies in the difference between the samples in out-of-bag and original samples. The difference between the samples is calculated using mean error values of the permuted and the original samples. The mean error is calculated using Gini impurity. The Gini index of the tree with node n is calculated using Gini(n) = 1- $\Sigma_Y^2\, (fYi^2)$, where *fi* is the frequency of class Y in the node n. The greater decrease in error reflects the feature importance. The difference between the original and the permutated sample mean errors reveals the association and the purpose of using mean error is to remove the weak associations. The mean error is given by

$$\mathrm{E}^{\mathrm{oob}} = \frac{1}{N_{\mathrm{oob}}} \Sigma_{i=1}^{n} (Y, Y'^{oob}).$$

The proposed feature ranking strategy employees the feature importance score from the above method and a new random forest is constructed using the feature importance score as a weighting criterion for sampling of features. The probability Pi of the features, which is proportional to weight wi, is used for selecting the best split from the feature sampling. The weighting criterion wi is given by the following formula:

$$wi = \frac{1}{Pi}.$$

By using the variable importance scores for feature sampling, the proposed model improves the classification accuracy by eliminating noisy features in the dataset.

Algorithm 2

```
Input: X=($x_1$,$x_2$,...,$x_d$) (the whole dimensional space with $S^Y$),
$R_k$ = {$b_i$ | i=1, 2,..., k; $b_i \in$ X} where k = (0, 1, 2,..., $S^Y$)
K = Tree Numbers
Output: Random Trees $R_k$
For k to 1 $\longrightarrow$ k do
Build a bagged subset of samples from $R_k$. Select randomly X features.
For X to 1 $\longrightarrow$ $S^Y$ do
Calculate decrease in the leaf node impurity.
    Calculate the feature importance score
Resample using feature weights
Build random forest and split trees based on weighting criteria
```

The features that are filtered through forward selection have much reduced variations achieved through minimum nearest neighbor k, while features that are grown into trees have much variance. In order to reduce the variance and avoid overfitting, the training data can be split into different bootstrapped samples. By averaging the different samples, the variance in the trees can be reduced, also avoiding overfitting. When using different bootstrapped samples to build trees, the samples generated at the random split would eventually have two different variables at least. As a result, the correlation between the trees decreases and each variable become independent. The problem of variance usually arises for small numbers of training instances and this can be bootstrapped to avoid higher variances. Also, when the training instances increase, the variance naturally decreases, and through bootstrapped samples, the problem of variance can be reduced by averaging the bootstrapped samples. The averaging reduces the variance and improves the classification performance. To test the performance of the proposed ensemble learning method ESFS, different methods such as FSNBb, FSSVMb, GASVMb, GANBb, GASVMb, and GARFb are compared using 15 different datasets.

4. Experiments

4.1. Dataset

The proposed feature selection method is applied on fifteen different datasets taken from the UCI repository. The feature numbers corresponding to datasets are given in Table 2. To test the performance of the model, different datasets with varying numbers of rows, columns, feature dimensions, feature types, and class levels were chosen for this study. The Mushroom dataset has large data instances with 8124, Audiology has a high number of features with 226, the Zoo dataset has the smallest number of data instances with 101, and the Diabetes dataset has the smallest number of features with 9. Features with different data types, sizes, scales, and numbers are useful to capture the performance of the model across all dimensions and investigate the behavior of the model in terms of data size, feature numbers, and different scales. Benchmark datasets and datasets that are popular among the data mining community are chosen for this study. The fifteen different datasets chosen for the study are widely preferred by the research community for studies related to dimensionality reduction, subset selection, feature selection, feature extraction, and other data mining problems.

Table 2. Datasets used in the study.

S. no.	Dataset	Instances	No. of features
1	Mushroom	8124	22
2	Thyroid	7200	22
3	Diabetes	768	9
4	Liver	584	11
5	Breast Cancer	569	32
6	Heart (SA)	462	10
7	CKD	400	24
8	Dermatology	366	34
9	Ionosphere	351	34
10	Tumor Data	339	17
11	Heart (Cleveland)	303	14
12	Heart (Statlog)	270	13
13	Audiology	226	69
14	Lymphography	148	18
15	Zoo	101	17

4.2. Evaluation metrics

Using a confusion matrix, the performance of the classifiers is visualized through true positive, true negative, false positive, and false negative ratios. True positive (TP) gives the instances that are correct and positive predictions, false positive (FP) gives the instances of incorrect positive predictions, true negative (TN) gives the instances of correct negative predictions, and false negative (FN) gives the instances of incorrect negative predictions. The confusion matrix is shown in Table 3. Accuracy is given by the following formula:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)},$$

whereas sensitivity, specificity, precision, and F-score is given by:

$$\text{Sensitivity} = \frac{(TP)}{(TP+FN)},$$

$$\text{Specificity} = \frac{(TN)}{(TN+FP)},$$

$$\text{Precision} = \frac{(TP)}{(TP+FP)},$$

$$\text{F-score} = \frac{(2*TP)}{(2*TP+FP+FN)}.$$

The proposed feature selection algorithm and classifiers are developed using Python 3.6. To compare the performance of the proposed method, ensemble feature selection methods such as FSNBb, FSSVMb, GASVMb, GANBb, and GARFb are used. The accuracy of the classification on the training set of the selected datasets

Table 3. Confusion matrix.

	Prediction		
Actual		Positive	Negative
	Positive	TP	FP
	Negative	FN	TN

is compared for evaluating the proposed method. The details of the feature selection methods and their abbreviations are listed in Table 4.

Table 4. Abbreviations and methods used. *: Proposed method.

S. no.	Algorithm	Abbrevation
1	SFS+Bagging+NB	FSNBb
2	SFS+Bagging+SVM	FSSVMb
3	GA+Bagging+NB	GANBb
4	GA+Bagging+SVM	GASVMb
5	GA+Bagging+RF	GARFb
6	SFS+Bagging+RF	ELFS*

5. Results and discussion

According to Table 5, the results of the proposed classification method using a wrapper and ensemble of algorithms on all features shows that the proposed algorithm has achieved the highest accuracy performance in Diabetes (96%), CKD (95%), Ionosphere (90%), Tumor Data (81%), Heart (Cleveland) (91%), and Heart (Statlog) (92%). On the other hand, the genetic algorithm with random forest and bagging (GARFb) shows the highest performance for five datasets, namely Mushroom (86%), Thyroid (96%), Liver (97%), Heart (SA) (95%), and Heart (Statlog) (90%). The genetic algorithm and SVM and bagging (GASVMb) achieved higher performance in Lymphography (88%) and Zoo (87%). Wrapper and SVM and bagging (FSSVMb) achieved high performance on Breast Cancer (96%) and Dermatology (81%). Wrapper naive Bayes (FSNBb) and genetic algorithm naive Bayes (GANBb) do not show better performance comparatively with the other ensemble algorithms.

The experimental results show that random forest with genetic algorithm and bagging performed well against all other ensemble learning algorithms on five datasets with mean accuracy of 88% (Mushroom, Thyroid, Liver, Heart (SA), Heart (Statlog)), while the proposed method achieved a mean accuracy of 87% on six datasets (Diabetes, CKD, Ionosphere, Tumor Data, Heart (Cleveland), Heart (Statlog)). Genetic algorithm-based ensembles suffer greatly with high processing time due to convergence while wrapper-based ensembles have lower processing time. The accuracy of the ensemble learning algorithm using relevant features for classification shows an improved performance. The proposed algorithm shows an improved accuracy of 87% on Mushroom while selecting twelve features. On the Thyroid dataset the accuracy improved to 10% using six relevant features. On the Liver dataset, the accuracy improved to 94% with seven features. For the Breast Cancer dataset, the accuracy improved to 96% with 11 features. On Heart (SA), with 6 features, the accuracy improved to 93%. Similarly, the accuracy on the Dermatology dataset improved to 91% for fifteen features, while on the Tumor

Table 5. Accuracy of ensemble algorithms on datasets using all features (best results in bold).

S. no.	Dataset	Features	FSNBb	FSSVMb	GASVMb	GANBb	GARFb	ELFS
1	Mushroom	22	0.66	0.79	0.83	0.70	**0.86**	0.81
2	Thyroid	22	0.73	0.84	0.95	0.71	**0.96**	0.83
3	Diabetes	9	0.79	0.76	0.94	0.91	0.95	**0.96**
4	Liver	11	0.80	0.85	0.93	0.74	**0.97**	0.91
5	Breast Cancer	32	0.84	**0.96**	0.91	0.73	0.81	0.86
6	Heart (SA)	10	0.79	0.82	0.94	0.71	**0.95**	0.89
7	CKD	24	0.59	0.78	0.86	0.79	0.79	**0.95**
8	Dermatology	34	0.75	**0.81**	0.79	0.52	0.83	0.80
9	Ionosphere	34	0.60	0.77	0.81	0.70	0.88	**0.90**
10	Tumor Data	17	0.57	0.67	0.73	0.76	0.78	**0.81**
11	Heart (Cleveland)	14	0.68	0.79	0.85	0.76	0.89	**0.91**
12	Heart (Statlog)	13	0.74	0.81	0.87	0.69	0.91	**0.92**
13	Audiology	69	0.63	0.74	0.73	0.82	**0.90**	0.88
14	Lymphography	18	0.75	0.73	**0.88**	0.86	0.82	0.79
15	Zoo	17	0.67	0.78	**0.87**	0.73	0.86	0.86
	Average		0.71	0.79	0.86	0.74	**0.88**	**0.87**

dataset the accuracy improved to 94% using nine features. On Heart (Statlog) the accuracy improved to 96% using six features and on Lymphography the accuracy improved to 90% using eight features. On the Zoo dataset, the accuracy improved to 97% using nine features. However, for CKD, Ionosphere, Heart (Cleveland), and Audiology the performance did not improve when selecting relevant features. Comparatively, the mean accuracy of different ensemble algorithms shows FSNBb with 74%, FSSVMb with 80%, GASVMb with 86%, GANBb with 80%, GARFb with 91%, and ELFS with 92%. The proposed method achieved the highest mean accuracy of 92%, which is a result of selecting the relevant features and removing the noisy features. The proposed method evaluates the features using the search method and ensemble model for selecting the best features.

The existing method [34] uses an ensemble method with decision trees, and the main drawback with decision trees is that the learning is biased with respect to the training set and when training data differ they produce errors. Also, DTs suffer from poor generalization and they are robust to lower the impurity only through a single tree, which cannot converge to different features present in a single tree while splitting. The tree split is based on the single tree, which does not cover different features and may contain high variance, whereas in the random trees, trees are built for every feature and RF builds additional training sets so that generalization can be easily achieved, whereas generalization reduces the variance and improves the class prediction. The existing work uses the wrapper and evolutionary search and ensembles with DT and naive Bayes, while in our proposed work we modified the random forest to select the features based on the probability weighting method to resample the features. The performance of the proposed method achieved a higher accuracy rate (Figure 2) on several different datasets (Thyroid, Diabetes, Liver, Heart (SA), Tumor, Heart (Cleveland), Heart (Statlog), Lymphography, Zoo) against wrapper evaluators FSNBb, FSSVMb, GANBb, FASVMb, and GARFb (Table 6).

The proposed method has the ability to select relevant features and also to improve the classification

Table 6. Accuracy of ensemble algorithms using feature selection (best results in bold).

S. no.	Dataset	Features	FSNBb	FSSVMb	GASVMb	GANBb	GARFb	ELFS
1	Mushroom	12	0.86	0.89	0.93	0.90	**0.96**	0.87
2	Thyroid	6	0.83	0.82	0.85	0.82	0.90	**0.93**
3	Diabetes	5	0.79	0.76	0.94	0.91	0.95	**0.96**
4	Liver	7	0.81	0.83	0.89	0.84	0.93	**0.94**
5	Breast Cancer	11	0.78	**0.86**	0.89	0.83	0.91	**0.96**
6	Heart (SA)	6	0.75	0.72	0.88	0.81	0.89	**0.93**
7	CKD	14	0.69	0.75	0.76	0.74	0.89	**0.85**
8	Dermatology	15	0.78	0.80	0.83	0.82	**0.93**	0.91
9	Ionosphere	13	0.65	0.81	**0.86**	0.71	0.78	0.85
10	Tumor Data	9	0.59	0.77	0.84	0.67	0.87	**0.94**
11	Heart (Cleveland)	7	0.68	0.79	0.85	0.76	0.89	**0.91**
12	Heart (Statlog)	6	0.79	0.80	0.86	0.73	0.95	**0.96**
13	Audiology	19	0.57	0.64	0.79	0.77	**0.94**	0.89
14	Lymphography	8	0.65	0.78	0.84	0.81	0.87	**0.90**
15	Zoo	9	0.95	0.91	0.96	0.93	0.96	**0.97**
	Average		0.74	0.80	0.86	0.80	**0.91**	**0.92**

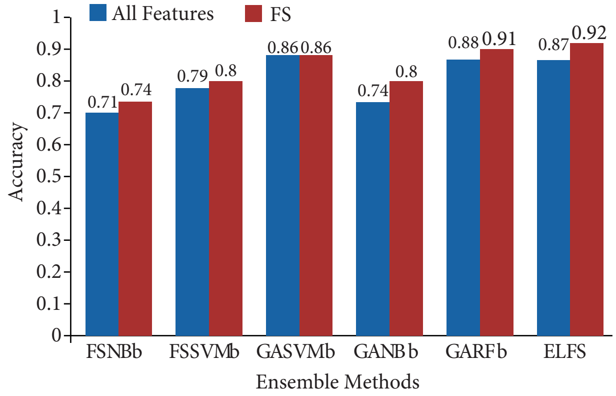

Figure 2. Mean accuracy of ensemble models.

performance compared to the existing ensemble methods. The precision scores (Table 7) for all features show that GARFb achieved the highest mean score of 87% compared to FSSVMb (0.84%) and GASVMb (0.85%), and the proposed method, ELFS, achieved 83%. The precision score indicates the ability of the classifier to correctly classify positive instances as positive. Also, for selected features (Table 8), the precision score of the proposed model improves to 94%, which indicates the ability of the model to correctly identify positive cases as positive. The ability to detect positive instances as positive reduces the misclassification error, thereby decreasing false positives, and improves the overall accuracy of the model. The precision of 94% shows that the model is able to correctly predict positive cases as positive 94 times out of every 100 times. The F-scores for all features (Table 9) show GASVMb with the highest score of 88%, while ELFS achieved 85%. When applying the selected features to the models, the proposed model, ELFS, achieved the highest F-score of 93% (Table 10). The precision and F-score values show more evidence that the classifier model gains better classification ability

Table 7. Precision of ensemble algorithms using all features.

S. no.	Dataset	Features	FSNBb	FSSVMb	GASVMb	GANBb	GARFb	ELFS
1	Mushroom	22	0.64	0.76	0.81	0.65	0.85	0.75
2	Thyroid	22	0.74	0.84	0.95	0.96	0.96	0.83
3	Diabetes	9	0.86	0.94	0.96	0.94	0.98	0.98
4	Liver	11	0.89	0.96	0.97	0.77	0.99	0.93
5	Breast Cancer	32	0.85	0.99	0.99	0.85	0.81	0.85
6	Heart (SA)	10	0.91	0.89	0.96	0.64	0.96	0.91
7	CKD	24	0.90	0.90	0.93	0.87	0.70	0.93
8	Dermatology	34	0.76	0.87	0.88	0.78	0.80	0.63
9	Ionosphere	34	0.53	0.78	0.87	0.75	0.89	0.92
10	Tumor Data	17	0.67	0.68	0.76	0.78	0.82	0.24
11	Heart (Cleveland)	14	0.65	0.82	0.92	0.77	0.96	0.97
12	Heart (Statlog)	13	0.80	0.81	0.91	0.64	0.93	0.93
13	Audiology	69	0.60	0.73	0.72	0.82	0.92	0.94
14	Lymphography	18	0.80	0.78	0.91	0.91	0.80	0.82
15	Zoo	17	0.76	0.87	0.16	0.63	0.73	0.83
	Average		0.76	0.84	0.85	0.78	**0.87**	0.83

Table 8. Precision of ensemble algorithms using feature selection.

S. no.	Dataset	Features	FSNBb	FSSVMb	GASVMb	GANBb	GARFb	ELFS
1	Mushroom	12	0.95	0.97	0.97	0.97	0.97	0.93
2	Thyroid	6	0.83	0.82	0.85	0.82	0.91	0.94
3	Diabetes	5	0.80	0.73	0.72	0.96	0.92	0.97
4	Liver	7	0.87	0.88	0.92	0.87	0.95	0.97
5	Breast Cancer	11	0.93	0.95	0.90	0.90	0.96	0.97
6	Heart (SA)	6	0.69	0.68	0.91	0.86	0.91	0.95
7	CKD	14	0.80	0.89	0.90	0.90	0.89	0.88
8	Dermatology	15	0.66	0.72	0.70	0.68	0.89	0.88
9	Ionosphere	13	0.90	0.90	0.88	0.76	0.85	0.89
10	Tumor Data	9	0.69	0.80	0.85	0.65	0.88	0.96
11	Heart (Cleveland)	7	0.65	0.79	0.91	0.70	0.86	0.90
12	Heart (Statlog)	6	0.84	0.80	0.90	0.77	0.95	0.97
13	Audiology	19	0.50	0.61	0.79	0.79	0.97	0.93
14	Lymphography	8	0.63	0.86	0.85	0.79	0.91	0.91
15	Zoo	9	0.91	0.91	0.96	0.89	0.96	0.98
	Average		0.78	0.82	0.87	0.82	0.92	**0.94**

through feature selection, which results in a subset of features that are relevant to the classes and removal of irrelevant features as redundant. Table 11 shows the model that achieved better performance through feature selection on each dataset and the proposed model shows better results on 10 datasets.

Table 9. F-score of ensemble algorithms using all features.

S. no.	Dataset	Features	FSNBb	FSSVMb	GASVMb	GANBb	GARFb	ELFS
1	Mushroom	22	0.66	0.79	0.83	0.70	0.86	0.80
2	Thyroid	22	0.84	0.91	0.97	0.98	0.98	0.90
3	Diabetes	9	0.85	0.83	0.95	0.93	0.96	0.97
4	Liver	11	0.87	0.91	0.96	0.82	0.98	0.94
5	Breast Cancer	32	0.87	0.98	0.93	0.80	0.84	0.88
6	Heart (SA)	10	0.85	0.87	0.96	0.74	0.96	0.92
7	CKD	24	0.62	0.76	0.83	0.75	0.71	0.93
8	Dermatology	34	0.65	0.74	0.73	0.50	0.74	0.66
9	Ionosphere	34	0.63	0.81	0.86	0.76	0.91	0.92
10	Tumor Data	17	0.70	0.76	0.81	0.83	0.85	0.38
11	Heart (Cleveland)	14	0.69	0.81	0.87	0.78	0.91	0.93
12	Heart (Statlog)	13	0.77	0.83	0.89	0.69	0.92	0.93
13	Audiology	69	0.72	0.82	0.81	0.88	0.93	0.93
14	Lymphography	18	0.79	0.77	0.90	0.89	0.84	0.82
15	Zoo	17	0.67	0.78	0.86	0.67	0.70	0.83
	Average		0.75	0.82	**0.88**	0.78	0.87	0.85

Table 10. F-score of ensemble algorithms using feature selection.

S. no.	Dataset	Features	FSNBb	FSSVMb	GASVMb	GANBb	GARFb	ELFS
1	Mushroom	12	0.88	0.90	0.94	0.91	0.96	0.88
2	Thyroid	6	0.90	0.90	0.91	0.90	0.94	0.96
3	Diabetes	5	0.84	0.80	0.79	0.95	0.93	0.97
4	Liver	7	0.88	0.89	0.93	0.89	0.96	0.96
5	Breast Cancer	11	0.84	0.90	0.92	0.87	0.93	0.97
6	Heart (SA)	6	0.78	0.76	0.91	0.86	0.91	0.95
7	CKD	14	0.66	0.74	0.75	0.75	0.87	0.84
8	Dermatology	15	0.66	0.70	0.74	0.72	0.88	0.86
9	Ionosphere	13	0.77	0.86	0.89	0.77	0.83	0.89
10	Tumor Data	9	0.72	0.84	0.89	0.75	0.91	0.96
11	Heart (Cleveland)	7	0.69	0.80	0.87	0.76	0.90	0.92
12	Heart (Statlog)	6	0.82	0.82	0.88	0.76	0.96	0.97
13	Audiology	19	0.65	0.73	0.86	0.85	0.96	0.94
14	Lymphography	8	0.69	0.82	0.87	0.84	0.89	0.92
15	Zoo	9	0.94	0.90	0.96	0.92	0.96	0.97
	Average		0.78	0.82	0.87	0.83	0.92	**0.93**

6. Conclusion and future work

In this study, we proposed an ensemble learning classification method using bagging, wrapper, and random trees. The method (ELFS) derives a feature subset and uses probability weighting criteria to improve the classification. Using the probability weighting, the proposed method effectively distinguishes the relevant and irrelevant features and removes the noisy features. By removing noisy features, the classification performance is

Table 11. Performance and features selected on the datasets.

S.No	Dataset	No. of features	Features selected	Performances	Methods
1	Mushroom	22	12	0.96	**GARFb**
2	Thyroid	22	6	0.93	**ELFS**
3	Diabetes	9	5	0.96	**ELFS**
4	Liver	11	7	0.94	**ELFS**
5	Breast Cancer	32	11	0.96	**ELFS**
6	Heart (SA)	10	6	0.93	**ELFS**
7	CKD	24	14	0.89	**GARFb**
8	Dermatology	34	15	0.93	**GARFb**
9	Ionosphere	34	13	0.86	**GASVMb**
10	Tumor Data	17	9	0.94	**ELFS**
11	Heart (Cleveland)	14	7	0.91	**ELFS**
12	Heart (Statlog)	13	6	0.96	**ELFS**
13	Audiology	69	19	0.94	**GARFb**
14	Lymphography	18	8	0.90	**ELFS**
15	Zoo	17	9	0.97	**ELFS**

improved. The proposed method achieves classification mean accuracy of 92%, F-score of 93%, and precision of 94% compared to FSNBb, FSSVMb, GASVMb, GANBb, and GARFb methods. The F-score and precision value indicate that the model is capable of reducing misclassification errors and has the ability to classify instances correctly irrespective of the feature numbers. The experimental results show that the proposed model gains improvement in classification while applying feature selection and, compared to other methods, the proposed model is efficient to improve classification results and reduce computational time. In the future, the present work could be extended to evaluate the performance with respect to growing number of trees and to include more ensemble models using other search methods and feature selection methods that are widely used in data mining.

References

[1] Rodríguez D, Ruiz R, Cuadrado-Gallego J, Aguilar-Ruiz J. Detecting fault modules applying feature selection to classifiers. In: IEEE 2007 International Conference on Information Reuse and Integration; Las Vegas, NV, USA; 2007. pp. 667–672.

[2] Chandrashekar G, Sahin F. A survey on feature selection methods. Computers & Electrical Engineering 2014; 40 (1): 16-28. doi: 10.1016/j.compeleceng.2013.11.024

[3] Guyon I, Elisseeff A. An Introduction to Feature Extraction. In: Guyon I, Nikravesh M, Gunn S, Zadeh LA (editors). Feature Extraction. Studies in Fuzziness and Soft Computing. Berlin, Germany: Springer, 2006, pp. 1-25.

[4] Blum AL, Langley P. Selection of relevant features and examples in machine learning. Artificial Intelligence 1997; 97 (1-2): 245-271. doi: 10.1016/S0004-3702(97)00063-5

[5] Hsu HH, Hsieh CW, Lu MD. Hybrid feature selection by combining filters and wrappers. Expert Systems with Applications 2011; 38 (7): 8144-8150. doi: 10.1016/j.eswa.2010.12.156

[6] Breiman L. Random forests. Machine Learning 2001; 45 (1): 5–32. doi: 10.1023/A:1010933404324

[7] Chandralekha M, Shenbagavadivu N. Performance analysis of various machine learning techniques to predict cardiovascular disease: an empirical study. Applied Mathematics & Information Sciences 2018; 12 (1): 217-226. doi: 10.18576/amis/120121

[8] Guan Y, Myers CL, Hess DC, Barutcuoglu Z, Caudy AA et al. Predicting gene function in a hierarchical context with an ensemble of classifiers. Genome Biology 2008; 9 (1):S3. doi: 10.1186/gb-2008-9-s1-s3

[9] Hoque N, Singh M, Bhattacharyya DK. EFS-MI: An ensemble feature selection method for classification. Complex & Intelligent Systems 2018; 4 (2): 105-118. doi: 10.1007/s40747-017-0060-x

[10] Settouti N, Chikh MA, Barra V. A new feature selection approach based on ensemble methods in semi-supervised classification. Pattern Analysis and Applications 2017; 20 (3): 673-686. doi: 10.1007/s10044-015-0524-9

[11] Filali A, Jlassi C, Arous N. Dimensionality reduction with unsupervised ensemble learning using K-means variants. In: IEEE 2017 14th International Conference on Computer Graphics, Imaging and Visualization; Marrakesh, Morocco; 2017. pp. 93-98.

[12] Chaudhary A, Kolhe S, Kamal R. An improved random forest classifier for multi-class classification. Information Processing in Agriculture 2016; 3 (4): 215-222. doi: 10.1016/j.inpa.2016.08.002

[13] Sakri S, Rashid NA, Zain ZM. Particle swarm optimization feature selection for breast cancer recurrence prediction. IEEE Access 2018; 6: 29637-29647. doi: 10.1109/ACCESS.2018.2843443

[14] Cai Z, Zhu W. Feature selection for multi-label classification using neighborhood preservation. IEEE/CAA Journal of Automatica Sinica 2018; 5 (1): 320-330. doi:10.1109/JAS.2017.7510781

[15] Brankovic A, Falsone A, Prandini M, Piroddi L. Randomised algorithm for feature selection and classification. 2018; arXiv preprint. arXiv:1607.08400

[16] Park HW, Li D, Piao Y, Ryu KH. A hybrid feature selection method to classification and its application in hypertension diagnosis. In: ITBAM 2017 Information Technology in Bio- and Medical Informatics Conference; Lyon, France; 2017. pp. 11-19.

[17] Agre G, Dzhondzhorov A. A weighted feature selection method for instance-based classification. In: AIMSA 2016 Artificial Intelligence: Methodology, Systems, and Applications Conference; Varna, Bulgaria; 2016. pp. 14-25.

[18] Osanaiye O, Cai H, Choo KKR, Dehghantanha A, Xu Z et al. Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing. EURASIP Journal on Wireless Communications and Networking 2016; 2016 (1): 130. doi: 10.1186/s13638-016-0623-3

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

[20] Koutanaei FN, Sajedi H, Khanbabaei M. A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. Journal of Retailing and Consumer Services 2015; 27: 11-23. doi: 10.1016/j.jretconser.2015.07.003

[21] Sasikala S, Balamurugan SA, Geetha S. Multi filtration feature selection (MFFS) to improve discriminatory ability in clinical data set. Applied Computing and Informatics 2016; 12 (2): 117-127. doi: 10.1016/j.aci.2014.03.002

[22] Ebrahimpour MK, Eftekhari M. Ensemble of feature selection methods: a hesitant fuzzy sets approach. Applied Soft Computing 2017; 50: 300-312. doi: 10.1016/j.asoc.2016.11.021

[23] Tseng CJ, Lu CJ, Chang CC, Chen GD, Cheewakriangkrai C. Integration of data mining classification techniques and ensemble learning to identify risk factors and diagnose ovarian cancer recurrence. Artificial Intelligence in Medicine 2017; 78: 47-54. doi: 10.1016/j.artmed.2017.06.003

[24] Kamkar I, Gupta SK, Phung D, Venkatesh S. Stable feature selection for clinical prediction: exploiting ICD tree structure using tree-lasso. Journal of Biomedical Informatics 2015; 53: 277-290. doi: 10.1016/j.jbi.2014.11.013

[25] Lu H, Chen J, Yan K, Jin Q, Xue Y et al. A hybrid feature selection algorithm for gene expression data classification. Neurocomputing 2017; 256: 56-62. doi: 10.1016/j.neucom.2016.07.080

[26] Liu J, Lin Y, Lin M, Wu S, Zhang J. Feature selection based on quality of information. Neurocomputing 2017; 225: 11-22. doi: 10.1016/j.neucom.2016.11.001

[27] Vivekanandan T, Iyengar NCSN. Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. Computers in Biology and Medicine 2017; 90: 125-136. doi: 10.1016/j.compbiomed.2017.09.011

[28] Bellal F, Elghazel H, Aussem A. A semi-supervised feature ranking method with ensemble learning. Pattern Recognition Letters 2012; 33 (10): 1426-1433. doi: 10.1016/j.patrec.2012.03.001

[29] Hong Y, Kwong S, Chang Y, Ren Q. Unsupervised feature selection using clustering ensembles and population based incremental learning algorithm. Pattern Recognition 2008; 41 (9): 2742-2756. doi: 10.1016/j.patcog.2008.03.007

[30] Yang J, Yao D, Zhan X, Zhan X. Predicting disease risks using feature selection based on random forest and support vector machine. In: ISBRA 2014 Bioinformatics Research and Applications Conference; Zhangjiajie, China; 2014. pp. 1-11.

[31] Maldonado S, Weber R. A wrapper method for feature selection using support vector machines. Information Sciences 2009; 179 (13): 2208-2217. doi: 10.1016/j.ins.2009.02.014

[32] Zheng B, Yoon SW, Lam SS. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Systems with Applications 2014; 41 (4): 1476-1482. doi: 10.1016/j.eswa.2013.08.044

[33] Zhou Q, Zhou H, Li T. Cost-sensitive feature selection using random forest: selecting low-cost subsets of informative features. Knowledge-Based Systems 2016; 95: 1-11. doi: 10.1016/j.knosys.2015.11.010

[34] Panthong R, Srivihok A. Wrapper feature subset selection for dimension reduction based on ensemble learning algorithm. Procedia Computer Science 2015; 72: 162-169. doi: 10.1016/j.procs.2015.12.117