

# A Robust Multi-Class Feature Selection Strategy Based on Rotation Forest Ensemble Algorithm for Diagnosis of Erythemato-Squamous Diseases

Akin Ozcift · Arif Gulten

Received: 16 May 2010 / Accepted: 5 July 2010 / Published online: 13 July 2010  
© Springer Science+Business Media, LLC 2010

**Abstract** In biomedical studies, accuracy of classification algorithms used in disease diagnosis systems is certainly an important task and the accuracy of system is strictly related to extraction of discriminatory features from data. In this paper, we propose a new multi-class feature selection method based on Rotation Forest meta-learner algorithm. The feature selection performance of this newly proposed ensemble approach is tested on Erythemato-Squamous diseases dataset. The discrimination ability of selected features is evaluated by the use of several machine learning algorithms. In order to evaluate the performance of Rotation Forest Ensemble Feature Selection approach quantitatively, we also used various and widely utilized ensemble algorithms to compare effectiveness of resultant features. The new multi-class or ensemble feature selection algorithm exhibited promising results in eliminating redundant attributes. The Rotation Forest selection based features demonstrated accuracies between 98% and 99% in various classifiers and this is a quite high performance for Erythemato-Squamous Diseases diagnosis.

**Keywords** Feature selection · Wrapper subset selection · Ensemble learning · Disease diagnosis · Boosting · Bagging · Rotation forest · Random forest · Classification performance · Erythemato-squamous diseases

## Introduction

Disease diagnosis by the use of classification algorithms is a widely applied phenomenon in biomedical studies. Consistent accuracy of classification algorithms used in disease diagnosing expert systems is certainly an important task to consider. The high volume of data produced by biomedical technologies requires some helpful classification approaches to support the analysis of this data [1]. On the other hand, some kinds of diseases, such as cancer, require developing a number of classification schemes to substitute invasive approaches for the benefit of patients [2].

Machine learning supports medical diagnosis in variety of applications to identify disease [3]. A wide range of diseases are attempted to be diagnosed by the use of some kind of computer software. There are numerous diseases reported in the literature that utilize machine-learning approaches. Particularly, some of the disease diagnosis studies that make use of an *ensemble algorithm based feature selection* strategy are cancer [4], acute abdominal pain [5], breast cancer, heart disease, and diabetes [6].

In this study, as a case disease to evaluate our feature selection scheme, we will utilize erythemato-squamous diseases. The diseases in this group are *psoriasis*, *seboreic dermatitis*, *lichen planus*, *pityriasis rosea*, *chronic dermatitis* and *pityriasis rubra pilaris*. Since, at first sight they all share the clinical features of erythema and scaling with very few differences, the differential diagnosis of these diseases is a difficult problem in dermatology. Another difficulty for differential diagnosis is that a disease may show the histopathological features of another disease at the beginning stage and may have the characteristic features at the following stages [7]. The

---

A. Ozcift (✉) · A. Gulten  
Electrical-Electronics Department, Firat University,  
Engineering Faculty,  
Elazig, Turkey  
e-mail: akinozcift@hotmail.com

**Table 1** Machine learning algorithms used in this study and their abbreviations

No	Algorithm	Abbreviation
1	Bayesian Network	BNET
2	Naive Bayes	NB
3	Multi Layer Perceptron	MLP
4	Simple Logistic	SL
5	Support Vector Machine	SVM
6	K*, Instance Based Learner	KSTAR
7	Decision Tree Based Rule Learner	PART
8	Functional Tree Classifier	FT

disease is diagnosed by the evaluation of 34 features. The summary and the properties of features of dermatology dataset are given in Table 3.

High dimensional data, in general, requires the extraction of most descriptive or discriminative features to be selected and hence the dimension of dataset is reduced. In this context, dimension reduction plays an important role in diagnosing systems to remove irrelevant features from a data set. Dimension reduction procedure is useful to decrease dataset complexity with the possible advantage of increased classification performance. In a relatively high dimensional data set, i.e. erythematous-squamous diseases, we used an ensemble algorithm based feature selection approach to determine *best discriminative features* out of 34 attributes. This feature selection strategy is intended to increase the accuracy in diagnosis of dermatology diseases.

The paper is organized as follows: first ensemble learning and particularly Rotation Forest (RF) ensemble algorithm are discussed briefly. Feature extraction techniques with particular attention to ensemble feature selection (EFS) will be explained as second. The classification metrics for evaluation of experiments and the experiments conducted will be the third and fourth sections respectively. The last section will include discussion of results.

### Ensemble learning and classifier ensembles

In machine learning, ensemble approaches use multiple weak classifiers to accomplish more accurate results compared to a single strong classifier alone. Ensemble algorithms combine multiple hypotheses from search space to get a better *ensemble solution* [8].

Ensemble classification algorithms generally obtain better prediction accuracies with diversity of the classifiers forming the ensemble. Decision-tree algorithms are particularly preferable to provide ensembles with diversity, [9,

10]. In our study, we used J48 decision tree algorithm (a java variant of C4.5 algorithm) in all of our classifier ensembles.

For large datasets, ensemble classifiers partitions data into subsets to train each classifier with different subset. Hence a combination rule is used to obtain final classification result. If, the dataset contains relatively little data, then classifiers are trained with bootstrap samples of the data. A bootstrap sample is a random sample of the data drawn with replacement [11, 12].

In the literature, widely used ensemble approaches are bagging, boosting, adaptive boosting (Adaboost), stacked generalization and mixture of experts. In this study, we used Adaboost, Decorate, Bagging, Random Forest and Rotation Forest ensemble algorithms to select discriminatory features of erythematous-squamous data. We will introduce these multi-class approaches briefly and for detailed discussion the reader might refer to the nice survey [11] and discussion [13] by Polikar.

Bagging is a *bootstrap aggregating* algorithm whose diversity is obtained by using bootstrapped subsets of the training data. Each subset of data is used to train a base classifier of the same type. The diverse decisions from individual classifiers are then combined by taking a simple majority vote rule [13].

Adaboost is an adaptive algorithm for constructing a *strong* classifier as linear combination of *weak* classifiers. The algorithm is adaptive in the sense that it calls a weak classifier repeatedly and the weight of each incorrectly classified example is increased so that the new classifier focuses more on those examples. The resultant classification is the combination of predictions using majority vote rule [13].

*Random forest* is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the class's output by individual trees. The method combines bagging idea and random selection of features in order to construct a collection of decision trees with controlled variation [14].

*Decorate* uses an existing *strong* learner to build an effective diverse group of classifiers. The diversity of

**Table 2** Ensemble learning algorithms used in this study and their abbreviations

No	Ensemble Algorithms	Abbreviation
1	Rotation Forest	ROTF
2	Adaboost	ADAB
3	Bagging	BAG
4	Decorate	DEC
5	Random Forest	RANDF

**Table 3** Structure of erythemato-squamous diseases data and the features selected by ensemble algorithms

No	Attribute	Selected Features by Ensemble Algorithms				
		ROTF	ADAB	BAG	DAG	RANDF
Clinical attributes						
1	Erythema					
2	Scaling				X	
3	Definite borders				X	
4	Itching				X	
5	Koebner phenomenon	X	X	X	X	X
6	Polygonal papules	X				
7	Follicular papules	X			X	X
8	Oral mucosal involvement	X	X			
9	Knee and elbow involvement	X				
10	Scalp involvement		X			
11	Family history					
34	Age Linear					
Histopathological attributes						
12	Melanin incontinence					
13	Eosinophils in the infiltrate				X	
14	PNL infiltrate	X	X	X	X	X
15	Fibrosis of the papillary dermis	X	X	X	X	X
16	Exocytosis					
17	Acanthosis		X		X	X
18	Hyperkeratosis					
19	Parakeratosis					
20	Clubbing of the rete ridges				X	
21	Elongation of the rete ridges	X	X	X	X	X
22	Thinning of the suprapapillary epidermis		X		X	
23	Spongiform pustule					
24	Munro microabcess					
25	Focal hypergranulosis					X
26	Disappearance of the granular layer	X	X	X		X
27	Vacuolization and damage of basal layer		X		X	X
28	Spongiosis	X	X	X	X	X
29	Saw-tooth appearance of retes		X	X	X	
30	Follicular horn plug					
31	Perifollicular parakeratosis	X	X	X	X	X
32	Inflammatory mononuclear infiltrate					X
33	Band-like infiltrate	X				X

classifiers in Decorate is provided by adding different randomly constructed examples to the training set when building new members of ensemble [15].

*Rotation Forest* is relatively a new ensemble strategy based on principal component analysis (PCA). The algorithm provides diversity by the use of PCA that constructs feature axis rotation for each base classifier [16]. In traditional Rotation Forest algorithm, decision trees are chosen for rotation task, because of their sensitivity to rotation of the feature axes [17].

**Ensemble feature selection**

There are mainly two approaches to reduce the dimensions of datasets: i) Feature Extraction (FE) and ii) Feature Selection (FS). Feature extraction techniques mainly consist of data transformation techniques to model high feature space with a lower dimension. This data transformation may be linear as in the case of Principal Component Analysis or it is a non-linear conversion with the utilization of a feed-forward Neural Networks [18]. PCA particularly

**Table 4** Root mean squared error scores of classifiers trained with features selected by ensemble algorithms

Ensemble Algorithm	Root Mean Square Error							
	BNET	NB	MLP	SL	SVM	KSTAR	PART	FT
Rotation Forest	0.0651	0.0797	0.0836	0.0822	0.3107	0.0816	0.0991	0.0760
Adaboost	0.0745	0.0733	0.0777	0.0968	0.3108	0.0811	0.0999	0.0851
Bagging	0.0864	0.0813	0.0880	0.0906	0.3117	0.1067	0.0953	0.0861
Decorate	0.0882	0.0915	0.1080	0.0991	0.3109	0.0951	0.1356	0.1026
Random Forest	0.0714	0.0817	0.0860	0.0957	0.3117	0.0893	0.1096	0.0851
All Features Used	0.0658	0.0832	0.0822	0.0915	0.3111	0.1169	0.1145	0.0845

involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components [19].

The second set of Feature Reduction methods comprises techniques to select optimal feature subset that maximizes the classification performance. More formally, for a given inducer algorithm (base classifier)  $I$ , for dataset  $D$  with instance space  $\{X_1, X_2, \dots, X_n\}$ , the aim is to find an optimal feature space  $X_{opt}$  that maximizes accuracy of inducer algorithm,  $I$ . FS techniques attempts to achieve this goal by removing redundant features from original dataset with the use of mainly two variable selection approaches; feature rankers that only consider intrinsic properties of the data, wrapper methods that embed the model hypothesis search within the feature subset search. For further detail of feature extraction techniques, the survey of Saeys et.al [20] is a good source to be referred.

In particular, a Wrapper Feature Selection (WFS) algorithm consists of a base classifier and a search approach such as Genetic Algorithms, Simulated Annealing, Tabu Search, and Best First Search as a wrapper around the classifier. These search approaches are usually equipped with some selection strategy such as forward selection or backward elimination. The classifier is introduced with set of features to select most discriminative features from whole feature space. The selection criterion is in general to select a subspace of relevant features that maximizes the accuracy of base classifier [21, 22].

Ensemble learning strategy, as mentioned, is expected to increase the accuracy of a classification problem by combining output of the base classifiers. Many applications in the literature support this anticipation with promising results. Hence, it is rational to expect an ensemble learning strategy to select optimal features that will increase accuracy of a classification problem. In the literature most FS techniques depending on some type of ensemble algorithms use Adaboost ensemble feature selection. Some case studies utilizing Adaboost ensemble feature selection are given in [23–25].

In this study, we used rotation forest ensemble decision tree algorithm wrapped with best first search strategy. The wrapper uses forward selection to select optimum subset out of 34 of erythemato-squamous disease variables. As to the best of our knowledge, Rotation Forest algorithm is first used in such feature selection scheme. In order to compare the success of obtained results, we made use of various ensemble algorithms, i.e. Adaboost, Bagging, Random Forest and Decorate. While this wrapper based ensemble feature selection is performed, a ten-fold cross validation technique is used to estimate the accuracy of the ensemble learning scheme for the selected features. All of the algorithms and the experiments of this study are implemented in Weka software, an open-source machine-learning environment built on top of Java platform. Leading advantages of Weka are good community support, cross-platform portability and support to be extendible by using Java language. The environment offers most of the

**Table 5** Kappa statistics values of classifiers trained with features selected by ensemble algorithms

Ensemble Algorithm	Kappa Statistics Error							
	BNET	NB	MLP	SL	SVM	KSTAR	PART	FT
Rotation Forest	0.9863	0.9761	0.9760	0.9829	0.9760	0.9726	0.9589	0.9761
Adaboost	0.9795	0.9726	0.9726	0.9726	0.9760	0.9795	0.9624	0.9727
Bagging	0.9692	0.9726	0.9692	0.9692	0.9486	0.9555	0.9658	0.9726
Decorate	0.9555	0.9590	0.9453	0.9590	0.9692	0.9486	0.9212	0.9487
Random Forest	0.9829	0.9761	0.9658	0.9658	0.9589	0.9658	0.9453	0.9692
All Features Used	0.9761	0.9693	0.9692	0.9658	0.9624	0.9317	0.9488	0.9692

**Table 6** Accuracy of classifiers corresponding to selected features by ensemble algorithms

Ensemble Algorithm	Accuracy (%)							
	BNET	NB	MLP	SL	SVM	KSTAR	PART	FT
Rotation Forest	98.91	98.08	98.08	98.64	98.08	97.81	96.72	98.08
Adaboost	98.36	97.81	97.81	97.81	98.08	98.36	96.99	97.81
Bagging	97.54	97.81	97.54	97.54	95.90	96.44	97.26	97.81
Decorate	96.45	96.72	95.62	96.72	97.54	95.90	93.71	95.90
Random Forest	98.63	98.08	97.26	97.26	96.72	97.26	95.62	97.54
All Features Used	98.08	97.54	97.54	97.26	96.99	94.53	95.90	97.54

important machine-learning algorithms that can be used by researchers efficiently.

**Error analysis concepts and classification performance indexes**

The performance of each selected feature group must be measured with the use of some metrics to evaluate the feature discrimination ability of the algorithms. In simpler terms, the quality of features will be measured with some widely used classifiers and their corresponding performances. The performance of classifiers, and hence the quality of selected features, will be evaluated using two groups of metrics. One group of metrics will measure the quality of classification by means of error indexes such as Root Mean Square Error (RMSE) and Kappa Statistic Error (KSE). The quality of classification will be measured by the use of Accuracy (ACC) and Area under receiver operating characteristic (AUC).

*Root Mean Square Error (RMSE)* In statistical learning theory, mean square error of a classifier (estimator) is one of the measures to quantify the difference between estimation and the true value of the quantity being estimated. MSE is a risk function, corresponding to the expected value of the squared error loss. MSE measures the average of the total error amount by which the classifier

differs from the quantity to be estimated. The MSE is incorporates both the variance of the classifier and its bias. The Root Mean Square Error (RMSE), i.e. square root of MSE, is a measure of the differences between values predicted by a model and the actual values. RMSE is a good metric for precision and it is one of the good single measures for predictive power of a classifier. RMSE can take values between one and zero and it is assumed to be better as its value tends to be zero [26]. For a classification problem, RMSE is calculated with Eq. 1.

$$RMSE = \sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n}} \tag{1}$$

In Eq. 1, predicted value is defined by *p* and actual value is defined by *a*.

*Kappa Statistic Error (KSE)* Kappa error or Cohen’s Kappa Statistics value is used to measure the agreement between predicted and observed categorizations of a dataset, while correcting the agreement that occurs by chance [27]. In classification algorithm performance comparisons, just using the percentage of misses as the single meter for accuracy can give misleading results. The cost of error must also be taken into account, while making such assessments. Kappa error, in this aspect, is a good measure to inspect classifications that may be due to chance. In general, Kappa error takes values between (−1,1). As the Kappa value calculated for classifiers approaches to ‘1’, then the

**Table 7** AUC value of classifiers corresponding the selected features by ensemble algorithms

Ensemble Algorithm	AUC							
	BNET	NB	MLP	SL	SVM	KSTAR	PART	FT
Rotation Forest	0.999	0.981	0.981	0.986	0.981	0.978	0.967	0.981
Adaboost	0.984	0.978	0.978	0.978	0.981	0.984	0.970	0.978
Bagging	0.975	0.978	0.975	0.975	0.959	0.964	0.973	0.978
Decorate	0.965	0.967	0.957	0.967	0.976	0.959	0.937	0.959
Random Forest	0.986	0.981	0.973	0.973	0.967	0.979	0.956	0.975
All Features Used	0.981	0.975	0.976	0.973	0.970	0.946	0.959	0.975

**Table 8** The comparison of accuracies of algorithms with and without feature selection

Algorithms	Accuracy (%)		
	without FS	with FS	Difference
BNET	98.08	98.91	0.83
NB	97.54	98.08	0.54
MLP	97.54	98.08	0.54
SL	97.26	98.64	1.38
SMO	96.99	98.08	1.09
KSTAR	94.53	98.36	3.83
PART	95.90	97.26	1.36
FT	97.54	98.08	0.54
AVG			1.26%

performance of the classifier is assumed to be more realistic rather than by chance. Therefore, in the performance analysis of classifiers, Kappa error is a recommended metric to consider for evaluation purposes and it is calculated by Eq. 2 [28].

$$K = \frac{p_0 - p_c}{1 - p_c} \tag{2}$$

In Eq. 2,  $p_0$  demonstrates total agreement probability and  $p_c$  agreement probability due to chance.

A binary classification problem deals with two class predictions. The goal of such problems is to map data samples into one of the groups, i.e. benign or malignant, with possible maximum estimation accuracy. For such a two-class problem, the outcomes are labeled as positive ( $p$ ) or negative ( $n$ ). The possible outcomes with respect to this classification scheme is frequently defined in statistical learning as *true positive* (TP), *false positive* (FP), *true negative* (TN) and *false negative* (FN). These four outcomes are connected to each other with a table that is

frequently called as *confusion matrix*. For a binary classification scheme, confusion matrix is used to derive most of the well known performance metrics such as sensitivity, specificity, accuracy, positive prediction value, F-measure, AUC and ROC curve. These metrics are evaluated using the confusion matrix outcomes, i.e.  $TP$ ,  $FP$ ,  $TN$  and  $FN$  predictive values [27].

By the use of these concepts, we may define  $ACC$  and  $AUC$  as follows:

*Accuracy (ACC):*  $ACC$  is a widely used metric to determine class discrimination ability of classifiers, and it is calculated using Eq. 3.

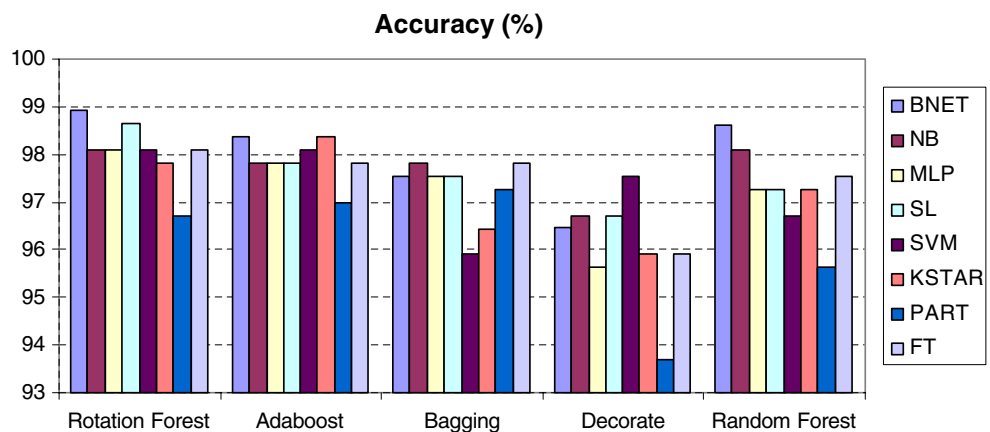
$$ACC = (TP + TN)/(P + N) \tag{3}$$

This is one of primary metrics in evaluating classifier performances and it is defined as the percentage of test samples that are correctly classified by the algorithm. The inspection of  $ACC$  values is easy for an experimental study.

*Area under the receiver operating characteristic (AUC):*  $AUC$  is a widely used metric in classification studies with relevant acceptance and it gives a good summary about performance of a classifier.  $AUC$  value is calculated from the area under the ROC curve. ROC curves are usually plotted using *true positives rate* versus *false positives rate*, as the discrimination threshold of classification algorithm is varied. In this aspect, Since a ROC curve compares the classifiers' performance across the entire range of class distributions and error costs; an  $AUC$  value is accepted to be a good measure of comparative performances of classification algorithms [29] and it is calculated with Eq. 4.

$$AUC = \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \tag{4}$$

**Fig. 1** Comparison of ensemble feature selection algorithms depending on accuracies of classifiers



**Table 9** A summary of measures of ensemble feature selection algorithm performance depending on classifiers

Ensemble Algorithm	Best value count out of four metrics							
	BNET	NB	MLP	SL	SVM	KSTAR	PART	FT
Rotation Forest	4	3	3	4	4	–	1	4
Adaboost	–	1	1	–	3	4	–	–
Bagging	–	–	–	–	–	–	3	–
Decorate	–	–	–	–	–	–	–	–
Random Forest	–	3	–	–	–	–	–	–

**Machine learning algorithms used in this study**

In this study, to evaluate the discriminative performances of features selected by ensemble algorithms, we will make use of eight classifiers. The name of the algorithms and their abbreviations are given in Table 1. For the sake of convenience, we will use abbreviation of classifiers where needed in this study.

In Table 1, for classifiers KSTAR and PART, we used original names as they are given in Weka environment. The details of classifiers will not be explained in this study for the sake of convenience. However, a good review for the most of the supervised machine learning algorithms such as [27] and [30] might be referred for detail. The ensemble algorithms and their abbreviations are also provided in Table 2.

**Experiments**

In this section, we will present the experiments carried out and their corresponding results. For convenience, both the structure of erythematous-squamous diseases data and the selected features are presented in Table 3.

With these selected features in Table 3, we executed eight algorithms to observe discriminative value of attributes of each multi-class feature selection approach. Additionally, we tested the performances of the algorithms with using *all 34 features* to evaluate the quantitative effect of feature selection algorithms. We also collected four groups of metrics to evaluate feature selection performances of multi-class algorithms. The error measures, i.e. Root Mean Square Error (RMSE) and Kappa Statistics Error (KSE), are given in Tables 4 and 5 respectively.

The evaluation of Table 4 is such that as the RMSE value approaches to zero, it is more reliable to think the accuracy of algorithm to be more confident. With this criterion in mind, it is clearly seen that Rotation Forest selected features has better error metrics for five out of eight classifiers.

For KSE as a second error metric, Table 5 gives valuable information about the confidence of the classifier performances. For values in Table 5; if Kappa Statistic value for a

classifier approaches to *one*, its performance is more valuable and less prone to be by chance. With this consideration, the classification reliability of Rotation Forest based features is seen to be superior and hence more confident for six out of eight measures.

In order to evaluate the effectiveness of selected features two more obvious metrics, i.e. Accuracy and AUC value, are calculated for each classifier. The accuracy of the classifiers is presented in Table 6 and the AUC values are given in Table 7 respectively.

The accuracies in Table 6 shows that Rotation Forest algorithm is clearly successful in six out of eight classifiers with superior (or at least the same) accuracy values. Hence, these results verify efficiency of Rotation Forest feature selection strategy.

While AUC values of Table 7 are examined, it is clearly seen that performance of Rotation Forest ensemble selection is also supported with AUC values. The algorithm has got better values in six out of eight classifiers. AUC values, as an indirect measure of feature selection quality, indicate efficiency of Rotation Forest based selection strategy in this scheme.

Here, the effect of feature selection might be questioned to observe the quality of feature selection algorithms. In order to make this comparison, we run eight classifiers without feature selection (all 34 features used) and we measured corresponding accuracy of classifiers. The accuracies of algorithms with all features used are given in Table 8. The highest accuracy of each classifier is also presented in the same table in order to make comparison easier.

The comparison in Table 8 shows the efficiency of feature selection strategy presented in this study. An average of 1.26% increase in accuracy shows the efficiency of RF based feature selection strategy.

To demonstrate the accuracy based performance of the classifiers depending on feature selection strategies visually, we produced Fig. 1 using Table 6.

**Discussion**

In order to evaluate the success of Rotation Forest based feature selection compared to other ensemble algorithms,

we make a comparison using Table 9. In this table, for each ensemble algorithm, we discuss how many metrics support the success of the feature selection. For instance, as we examine Table 9 for algorithm BNET, we note that the classification algorithm is either superior or the same in four metrics while considering Rotation Forest feature selection. In other words, Rotation Forest based features have better classification performances in BNET algorithm in four metrics compared with other ensemble algorithms.

Furthermore Table 9 shows that Rotation Forest based features, except in KSTAR algorithm, has at least one performance index either superior to other ensemble algorithms or equal. For instance Bagging and Random Forest has significantly good values in one classifier. However this is seen to be a local success while whole table is inspect. The only ensemble algorithm that is comparable to Rotation Forest is Adaboost. While Table 9 is examined, it is seen that Adaboost is successful in KSTAR and SVM and a little in NB and MLP classification algorithms with respect to Rotation Forest. However, Rotation Forest algorithm has 23 superior metrics in eight classifiers compared to nine good metrics of Adaboost algorithm. This statistics in Table 9 proves the quality of features selected with Rotation Forest.

While Table 3 is examined for the selected features, it is seen that some features, i.e. *Koebner phenomenon*, *PNL infiltrate*, *Fibrosis of the papillary dermis*, *Elongation of the rete ridges*, *Spongiosis*, *Perifollicular parakeratosis*, is selected with five algorithms. It is rational to think that these are the most important features of erythematous-squamous diseases. We examined the algorithms with these features to see the corresponding performances; however the results demonstrated that these features are not enough to represent the dataset's class information.

Furthermore, ensemble feature selection technique is computationally expensive while compared to classical statistical feature filters or rankers. To evaluate worthiness of this cost, we made chi-square ranking of erythematous-squamous attributes and selected top 12 features (the same number of features with Rotational Forest algorithm) to see the classifier accuracies of these features. With the same classifiers run on the selected *chisquare* features, we observed best accuracy of 86% with Bayes Network classifier. Hence, it is concluded that the computational cost of ensemble feature selection might be tolerable for the sake of such noticeable accuracy.

Lastly, we will give an accuracy comparison with the results seen in the literature for classification of erythematous-squamous diseases. In a recent article [31], the accuracies of classification of erythematous-squamous diseases related studies are given as to be 95.5%, 98.3% and 98.6%. Referring to this study, it is possible to state rotation forest ensemble algorithm based *features*, with accuracies of BNET 98.91% and SL 98.64%, are quite effective.

This experimental study as a whole proves the efficiency of Rotation Forest based feature selection scheme to be successful. The method with this success is promising and it is a good candidate to select discriminatory features in classification problems.

## References

- Polat, K., and Gunes, S., An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digit. Signal Process.* 702–710, 2007, Elsevier.
- Ozcift, A., and Gulten, A., Assessing effects of preprocessing mass spectrometry data on classification performance. *Eur. J. Mass Spectrom.* 267–273, 2008.
- Kononenko, I., Machine learning for medical diagnosis: History, state of the art and perspective. *Artif. Intell. Med.* 23(1):89–109, 2001.
- Abeel, T., Helleputte, T., Van de Peer, Y., Dupont, P., and Saeys, Y., Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics (Oxford, England)* 26(3):392–398, 2010.
- Tsymbal, A., Pechenizkiy, M., and Cunningham, P., Diversity in search strategies for ensemble feature selection, Computer-Based Medical Systems, 2003. Proceedings. 16th IEEE Symposium, pp. 83–98, 2005.
- Abdel-Aal, R., Improved classification of medical data using abductive network committees trained on different feature subsets. *Comput. Methods Prog. Biomed.*, 141–153, 2005.
- Guvenir, H. A., Demiroz, G., and Ilter, N., Learning differential diagnosis of erythematous-squamous diseases using voting feature intervals. *Artif. Intell. Med.*, 147–165, 1998.
- Opitz, D., and Maclin, R., Popular ensemble methods: An empirical study. *J. Artif. Intell. Res.* 169–198, 1999.
- Kuncheva, L., and Whitaker, C., Measures of diversity in classifier ensembles. *Mach. Learn.* 181–207, 2003.
- Ho, T., Random decision forests. *Proceedings of the Third International Conference on Document Analysis and Recognition* 278–282, 1995.
- Polikar, R., Ensemble based systems in decision making. *IEEE Circuits Syst. Mag.* 21–45, 2006.
- Efron, B., Bootstrap methods: Another look at the jackknife. *Ann. Stat.* 1–26, 1979.
- Polikar, R., Ensemble learning. *Scholarpedia* 4(1):2776, 2009.
- Breiman, L., Random forests. *Mach. Learn.* 5–32, 2001.
- Melville, P., and Mooney, R., Constructing diverse classifier ensembles using artificial training examples, *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence* 505–510, 2003.
- Zhang, C., and Zhang, J., A variant of rotation forest for constructing ensemble classifiers. *Pattern Analysis Applications* 59–77, 2010.
- Rodriguez, J., and Kuncheva, L., Rotation forest: A new classifier ensemble method. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1619–1630, 2006.
- Martinez, A. M., and Zhu, M., Where are linear feature extraction methods applicable? *IEEE Transaction on Pattern Analysis and Machine Intelligence* 1934–1944, 2005.
- Fodor, I. K., A survey of dimension reduction techniques. Department of Energy by the University of California, 1–7, 2002
- Saeys, Y., A review of feature selection techniques in bioinformatics. *Bioinformatics Review* 2507–2517, 2007.



21. Kohavi, R., and John, G., Wrappers for feature subset selection. *Artif. Intell.* 273–324, 1997.
22. Guyon, I. (Editor), Feature extraction, foundations and applications. *Stud. Fuzziness Soft Comput.* 119–135, 2006.
23. Xiao, R., Li, W., Tian, Y., and Tang, X., Joint boosting feature selection for robust face recognition. *Computer Vision and Pattern Recognition. IEEE Computer Society Conference on* 2 1415–1422, 2006.
24. Chen, K., Chou, C., and Shih, S., Feature selection for iris recognition with AdaBoost. *Intelligent Information Hiding and Multimedia Signal Processing* 411–414, 2007.
25. Tian, B., Zheng, X., Zhang, R., and Zhao, Y., Feature selection based on AdaBoost in video surveillance system, *Proceedings of the 2009 Second international Conference on intelligent Computation Technology and Automation* 70–72, 2009.
26. Viswanathan, M., *Measurement error and research design.* Sage Publications, 44–60, 2005
27. Witten, I. H., and Ian, H., Data mining: practical machine learning tools and techniques. *Morgan Kaufmann Ser. Data Manage. Syst.* 153–168, 2005.
28. David, A., Comparison of classification accuracy using Cohen's Weighted Kappa. *Expert Syst. Appl.* 825–832, 2008.
29. Huang, J., and Ling, C., Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.* 299–310, 2005.
30. Kotsiantis, S. B., Supervised machine learning: A review of classification techniques. *Informatica* 249–268, 2007.
31. Karabatak, M., and Cevdet İnce, M., A new feature selection method based on association rules for diagnosis of erythematous diseases. *Expert Syst. Appl.* 12500–12505, 2009.