British Journal of Mathematics & Computer Science
9(5): 394-406, 2015, Article no.BJMCS.2015.211
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Filter Based Feature Selection for Automatic Detection of Erythemato-squamous Diseases

A. H. El-Baz[1*]

[1]*Department of Mathematics, Faculty of Science, Damietta University, New Damietta, Egypt.*

Article Information

DOI: 10.9734/BJMCS/2015/17618
Editor(s):
(1) Dariusz Jacek Jakóbczak, Chair of Computer Science and Management in this department, Technical University of Koszalin, Poland.
(2) Tian-Xiao He, Department of Mathematics and Computer Science,Illinois Wesleyan University, USA.
Reviewers:
(1) Anonymous, Al-Zaytoonha University, Jordan.
(2) Anonymous, Fuzhou University, China.
(3) Anonymous, Rzeszów University of Technology, Poland.
(4) Anonymous, Universiti Sains Malaysia, Malaysia.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=1146&id=6&aid=9623

Original Research Article

Received: 20 March 2015
Accepted: 30 April 2015
Published: 06 June 2015

Abstract

This paper presents an automatic diagnosis model of erythemato-squamous diseases. The proposed model consists of two stages. In the first stage, two filter based feature selection methods, namely rough set using Johnson's algorithm and ranked features for feature selection of erythemato-squamous diseases are employed to select the optimal feature subset from the original feature set for dimensionality reduction in order to further improve the diagnostic accuracy. Next, for the sake of comparison, the diagnoses decisions are made by four different classification algorithms: k-nearest neighbors, Naive Bayesian classifier, linear discriminant analysis and decision tree. Experimental results show that the accuracies of the four base classifiers using ranked features outperformed those using rough set with Johnson's algorithm and the base classifiers without using feature selection. Using erythemato-squamous diseases dataset taken from UCI (University of California at Irvine) machine learning database. The accuracies of these four classifiers using ranked features on test sets (50% of the dataset) are 97.21, 98.32, 96.09, and 98.32, respectively. Therefore, we can conclude that the ranked features method is very promising in detection of erythemato-squamous diseases compared to the rough set using Johnson's algorithm and also compared favorably with previously reported results. This tool enables doctors to differentiate six types of erythemato-squamous diseases using clinical and histopathological parameters obtained from a patient.

Keywords: Dermatology; erythemato-squamous diseases; feature selection; ranked feature; rough set; decision tree; Naive Bayesian; KNN; LDA.

**Corresponding author: ali_elbaz@yahoo.com, elbaz@du.edu.eg;*

1 Introduction

The process of differentiating between two or more patients that share similar signs or symptoms of erythemato-squamous diseases is considered one of the challenge problems in medical diagnosis. Despite very little differences, all Erythemato-squamous diseases share the clinical features of erythema and scaling. In erythemato-squamous diseases, there are six diseases: Psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis and pityriasis rubra pilaris. Usually a biopsy is necessary for the diagnosis of these diseases, but unfortunately these diseases share many histopathological features as well. Another difficulty for the differential diagnosis is that one disease may show the features of another disease at the beginning stage and may have the characteristic features at the following stages. Patients were first evaluated clinically with 12 features. Afterwards, skin samples were taken for the evaluation of 22 histopathological features. The values of the histopathological features are determined by an analysis of the samples under a microscope [1].

Filter, Wrapper and embedded methods are considered the main three categories of feature selection. Filter methods measuring some properties in the dataset and provide an index for each feature. In Wrapper and embedded methods, a classifier is used to evaluate the importance of each feature. In other word, Wrapper employs a performance measure like classification accuracy of the classifier to guide a search process in the feature space and embedded methods use the internal parameters of the classifier to assess features.

Feature ranking is a topic of interest for many researchers. In the feature ranking, a ranked list of features is produced and one can select the top ranked features, where the number of features to select can be analytically or experimentally determined or set by the user. Many feature selection algorithms use feature ranking as a principal or auxiliary step because of its simplicity, scalability, and good empirical success. Furthermore, a ranked list of feature might be interesting by itself, as for instance in the microarray analysis, where the ranked list of features is used by biologists to find correlations among top ranked features and some diseases [2].

In this study, we give an automatic differential diagnosis of erythemato-squamous diseases which is considered one of the critical problems in dermatology. The technique is based on two steps. First, two different filter feature selection techniques for reducing the dimensionality of feature space of erythemato-squamous dataset namely rough set and feature ranking techniques are employed. Next, to compare their efficiency in this dataset, four different base classifiers are used namely k-nearest neighbors (KNN), Naive Bayesian, linear discriminant analysis (LDA) and decision tree.

The rest of this paper is organized as follows. Section 2 provides the problem of erythemato-squamous diseases. Section 3 gives a literature review on automatic detection of erythemato-squamous diseases. Section 4 provides a brief description on feature selection where rough set and ranked features are introduced as a tool for reducing the number of features as a priori algorithms. Section 5 presents a general overview about k-nearest neighbors, Naive Bayesian, Linear discriminant analysis, and decision tree. In Section 6, experimental results are given. Discussion of the obtained results is given in Section 7. Conclusions are demonstrated in Section 8.

2 Erythemato-squamous Diseases Dataset

In this study, the UCI erythemato-squamous diseases dataset was used and analyzed [3]. There are 366 samples in this data set and each sample has 34 features which contain 12 clinical and 22 histopathological features. These features are detailed in Table 1. The family history feature has the value '1' if any of these diseases has been observed in the family and '0' otherwise. The age feature simply represents the age of the patient which has been missed in eight samples, so these samples were removed in our experiments. Every other feature (clinical and histopathological) was given a degree in the range of '0' to '3'. Here, '0' indicates that the feature was not present, '3' indicates the largest amount possible, and '1', '2' indicate the relative intermediate values.

Table 1. The UCI erythemato-squamous diseases dataset

The erythemato-squamous diseases (Number of patients)	Features	
	Clinical	**Histopathological**
Psoriasis (111) Seboreic dermatitis (60) Lichen planus (71) Pityriasis rosea (48) Cronic dermatitis (48) Pityriasis rubra pilaris (20)	Feature 1: Erythema Feature 2: Scaling Feature 3: Definite borders Feature 4: Itching Feature 5: Koebner phenomenon Feature 6: Polygonal papules Feature 7: Follicular papules Feature 8: Oral mucosal involvement Feature 9: Knee and elbow involvement Feature 10: Scalp involvement Feature 11: Family history, (0 or 1) Feature 34: Age (linear)	Feature 12: Melanin incontinence Feature 13: Eosinophils in the infiltrate Feature 14: PNL infiltrate Feature 15: Fibrosis of the papillary dermis Feature 16: Exocytosis Feature 17: Acanthosis Feature 18: Hyperkeratosis Feature 19: Parakeratosis Feature 20: Clubbing of the rete ridges Feature 21: Elongation of the rete ridges Feature 22: Thinning of the suprapapillary epidermis Feature 23: Spongiform pustule Feature 24: Munro microabcess Feature 25: Focal hypergranulosis Feature 26: Disappearance of the granular layer Feature 27: Vacuolisation and damage of basal layer Feature 28: Spongiosis Feature 29: Saw-tooth appearance of retes Feature 30: Follicular horn plug Feature 31: Perifollicular parakeratosis Feature 32: Inflammatory monoluclear infiltrate Feature 33: Band-like infiltrate

3 Literature Review on Automatic Detection of Erythemato-squamous Diseases

There are several machine learning methods reported in literature for the automatic diagnosis of erythemato-squamous diseases using a benchmark dataset mentioned above. Güvenir et al. [1] developed a new classification algorithm, called VFI (for Voting Feature Intervals) and they applied it to differential diagnosis of erythemato-squamous diseases. Classification in the VFI algorithm is based on a real-valued voting. Each feature equally participates in the voting process and the class that receives the maximum amount of votes is declared to be the predicted class. The VFI algorithm achieved 96.2% accuracy using 10-fold cross-validation. Güvenir and Emeksiz [4] presented an expert system incorporating decisions made by three classification algorithms: k-nearest neighbors classifier, Naive Bayesian classifier and voting feature intervals. This system stores the patient records in a database for further reference. Übeyli and Güler [5] proposed a technique based on adaptive neuro-fuzzy inference system, and they obtained 95.5% for correct classification accuracy. Luukka and Leppälampi [6] proposed an approach based on fuzzy similarity classifier and the correct classification was 97.02%. The methods based on fuzzy weighted pre-processing, KNN based weighted pre-processing, and decision tree classifier were proposed by Polat and Gunes [7], and their classification accuracy reached to 88.18%, 97.57%, and 99.00%, respectively. Nanni [8] obtained 97.22%, 97.22%, 97.5%, 98.1%, 97.22%, 97.5%, 97.8%, and 98.3% using LSVM, RS, B1_5, B1_10, B1_15, B2_5, B2_10, and B2_15 algorithms. Luukka [9] presented similarity classifier using similarity measure derived from Yu's norm in classification of medical data sets, and the classification accuracy of diagnosis of erythemato-squamous diseases was 97.8%. Übeyli [10] obtained 98.32% classification accuracy on the differential diagnosis of erythemato-squamous diseases, using multiclass support vector machines with the error correcting output codes (ECOC). Polat and Güneş [11] obtained 96.71% classification correct rate on diagnosis of erythemato-squamous diseases using a novel hybrid intelligence method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problem. Übeyli [12] obtained classification accuracy 97.77% using combined neural networks (CNN) model to guide model selection for diagnosis of erythemato-squamous diseases. Liu et al. [13] obtained 96.72%, 92.18%, 95.08%, and 92.20% using feature selection algorithm with dynamic mutual information, which was estimated using four typical classifiers named Naive Bayes, 1-Nearest neighbor, C4.5 and PIPPER. Karabatak and Ince [14] proposed a new feature selection method based on Association Rules (AR) and Neural Network (NN) for diagnosis of erythemato-squamous diseases, and their correct classification rate was 98.61%, and the dimension of feature space was reduced from 34 to 24 by using AR. Xie and Wang [15] obtained 98.61% correct classification rate by using improved F-score and Sequential Forward Search for feature selection and SVM for classification. A diagnosis model based on particle swarm optimization (PSO), support vector machines (SVMs) and association rules (ARs) was developed by Abdi and Giveki [16] to diagnose erythemato-squamous diseases and the obtained result was 98.91%. Inbarani et al. [17] proposed new supervised feature selection methods based on hybridization of Particle Swarm Optimization (PSO), PSO based Relative Reduct (PSO-RR) and PSO based Quick Reduct (PSO-QR). The best results obtained are 94.86% and 98.84% for PSO-QR when Naive Bayes and KStar are used as classifiers, respectively and 95.89% and 98.56% for PSO-RR when also Naive Bayes and KStar are used as classifiers, respectively. Ravichandran et al. [18] presents a novel approach based on fuzzy extreme learning machine (FELM). By combining fuzzy logic and ELM, more accurate results with increased performance were obtained. The total classification accuracy of the FELM model was 93% where 310 records were used as training data and 56 other records used as testing data.

4 Feature Selection

Feature selection methods play a vital role in different artificial intelligence disciplines. A subset of features may produce better predictive models than the entire feature set. This is because learning algorithms may be adversely affected by the presence of irrelevant and/or redundant features. Besides improving classification accuracy, feature selection significantly reduces the computational time necessary to induce the models, leading to simpler and faster classifiers for classifying new instances; facilitates data visualization and data understanding; and reduces the measurement and storage requirements. In the following two subsections, a brief description of two feature selection techniques namely the rough set and the rank features is given.

4.1 Rough Set

Rough Set (RS) theory is an intelligent mathematical tool proposed by Pawlak in 1982 to deal with uncertainty and incompleteness [19]. Over the past few years, RST has become a topic of great interest to researchers and has been applied to many domains. It is based on the concept of an upper and a lower approximation of a set, the approximation space and models of sets. The main advantage of RS theory is that it does not need any preliminary or additional information about data: like probability density function in statistics or basic probability assignment in Dempster - Shafer theory and membership grade in fuzzy set theory. One of the major applications of RS theory is the attribute reduction that is possible to find a minimal subset. The reduction of attributes is achieved by comparing equivalence relations generated by sets of attributes. Using the dependency degree as a measure, attributes are removed and reduced set provides the same dependency degree as the original. This section recalls some essential definitions from RST that are used for feature selection. Detailed description and formal definitions of the theory can be found in [20-21].

4.1.1 Indiscernibility relation

The indiscernibility relation is considered the mathematical basis of rough sets theory. The B-indiscernibility relation is denoted by IND(B), is defined as:

$$\mathrm{IND(B)} = \{(x,x') \in U^2 \mid \forall a \in B,\ a(x)=a(x')\}$$

where a(x) denotes the value of attribute of object x. If $(x,x') \in \mathrm{IND(B)}$, x and x' are said to be indiscernible with respect to B. The equivalence classes of the B-indiscernibility relation are denoted by $[x]_B$.

4.1.2 Set approximation

There are two concepts related to set approximation which are the lower and upper approximations of a set [21]. Let S=(U,A) be an information system and let $B \subseteq A$ and $X \subseteq U$. The set X can be approximated using the information contained in B by constructing the B-lower approximation of X and the B-upper approximation of X, denoted $\underline{B}X$ and $\overline{B}X$ respectively, where

$$\underline{B}(X) = \{x \in U \mid [x]_B \subseteq X\},$$

$$\overline{B}(X) = \{x \in U \mid [x]_B \cap X \neq \varnothing\}$$

The objects in $\underline{B}X$ can be with certainty classified as members of X on the basis of knowledge in B, while the objects in $\overline{B}X$ can be only classified as possible members of X on the basis of knowledge in B. The set

$$BN_B(X) = \overline{B}(X) - \underline{B}(X)$$

is called the B-boundary region of X, and thus consists of those objects that we cannot decisively classify into X on the basis of knowledge in B. The set $U - \overline{B}X$ is called the B-outside region of X and consists of those objects which can be with certainty classified as do not belonging to X on the basis of knowledge in B. A set is said to be rough if its boundary region is non-empty, otherwise the set is crisp.

Rough set can characterized numerically by $\alpha_B(X) = \left|\underline{B}(X)\right| / \left|\overline{B}(X)\right|$ which called the accuracy of approximation [21]. Obviously $0 \le \alpha_B(X) \le 1.$ If $\alpha_B(X) = 1,$ X is crisp with respect to B, and otherwise, if $\alpha_B(X) < 1,$ X is rough with respect to B.

4.1.3 Reduct and core

A reduct is a minimal set of attributes from A (the whole attributes set) that provided that the object classification is the same as with the full set of attributes. Given C and D$\subseteq$A, a reduct is a minimal set of attributes such that IND(C) = IND(D). Let RED(A) denote all reducts of A. The intersection of all reducts of A is referred to as a core of A, i.e., CORE(A) = $\cap$ RED(A), the core is common to all reducts.

4.1.4 Dependency of attributes

Dependency between attributes is defined as follow: a set of attributes *D* depends totally on a set of attribute *C*, if all values of attribute from *D* are uniquely determined by values of attributes from *C*. In other words, *D* depends totally on *C*, if there exists a functional dependency between values of *D* and *C*. Formally dependency can be defined as follow: Let D and C be subsets of A. We say that D depends on C in degree k ($0 \le k \le 1$), if

$$k = \gamma(C, D) = \frac{|\mathrm{POS}_C(D)|}{|U|},$$

where $\mathrm{POS}_C(D) = \bigcup_{X \in U/D} \underline{C}(X)$,

called a positive region of the partition U/D with respect to C, is the set of all elements of U that can be uniquely classified to blocks of the partition U/D, by means of C. Obviously

$$\gamma(C, D) = \sum_{X \in U/D} \frac{|\underline{C}(X)|}{|U|}$$

If k=1 we say that D depends totally on C, and if K<1, we say that D depends partially (in a degree k) on C.

4.2 Rank Features

A feature ranking produces an ordered list of features where this list is ordered by decreasing importance. Based on this ranking we can select a subset of the top k ranked features [22].

In this paper, ranked feature is used to obtain the 16 ranked discriminant features of each disease (class). Table 2 gives the discriminant features of each disease. Then, the final set of reduced attributes of all diseases is obtained by the union of these six features subsets. The final feature vector contains 26 features out of 34 features. The deleted features are features number 1, 2, 11, 13, 17, 18, 32 and 34. Fig. 1 shows the feature's number versus it's repetition in different 6 classes.

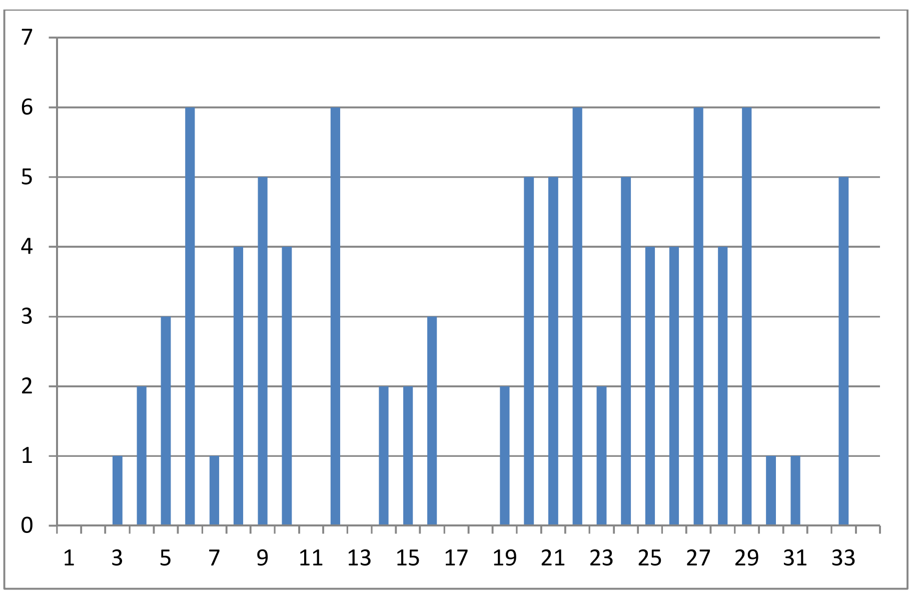

Fig. 1. The feature's number versus it's repetition in different 6 classes

Table 2. The discriminant features of each disease

Type of disease (class name)	Ranked selected set of attributes
Psoriasis	20, 22, 21, 28, 16, 10, 9, 19, 24, 3, 26, 29, 6, 33, 12, 27
Seboreic dermatitis	28, 20, 22, 5, 26, 21, 9, 24, 27, 16, 29, 6, 12, 25, 8, 33
Lichen planus	33, 27, 29, 6, 12, 25, 8, 21,14, 20, 22, 16, 9, 10, 4, 23
Pityriasis rosea	21, 9, 20, 22, 10, 28, 33, 27, 6, 12, 25, 8, 23, 29, 24, 4
Cronic dermatitis	15, 5, 14, 20, 10, 9, 22, 26, 28, 24, 27, 29, 6, 12, 25, 33
Pityriasis rubra pilaris	7, 31, 5, 22, 26, 21, 24, 30, 27,29, 6, 12, 8, 15, 33,19

5 Base Classifiers

To compare the performance of the rough set and the ranked features as filter based features selection methods. Four different classifiers are used, namely k-nearest neighbors, Naive Bayesian, decision tree, linear discreminant analysis, Brief description of each one of them is given below:

5.1 K-nearest Neighbors Classifier (KNN)

The KNN method is one of the most popular nonparametric methods [23] used for classification of new objects. KNN consists of a supervised learning algorithm, which instantly classifies the results of a query instance based on the majority of the KNN category. Classes are determined based on the minimum distance from the query instance to the training samples.

On the other hand, consider the case of *m* classes $\{C_i\}_{i=1}^{m}$ and a set of *N* sample objects $\{y_i\}_{i=1}^{N}$ whose classification is a priori known. The nearest neighbor technique classifiers an incoming object x in the pattern class of its nearest neighbor in the set $\{y_i\}_{i=1}^{N}$, i.e. if $\|x - y_j\| = \min_{1 \le i \le N} \|x - y_i\|$ then $x \in C_j$. This technique can be modified by considering the *k* nearest neighbors to x and using a majority-rule type classifier.

Major advantages of the KNN method are its simplicity and ease of implementation. KNN is not negatively affected by large training data [23].

5.2 Naive Bayesian Classifier Using Multinomial Distribution

Naive Bayesian classifier (NBC) is an algorithm that approaches the classification problem using conditional probabilities of the features [24]. The probability of the instance belonging to a single class is calculated by using the prior probabilities of classes and the feature values for an instance. NBC assumes that features are independent. In NBC, each feature participates in the classification by assigning probability values for each class, and the final probability of a class is the product of each single feature probabilities; and for an *n* dimensional domain, the probability of the instance belonging to a class $\mathrm{P(e|C_i)}$ can be computed as

$$\mathrm{P(e|C_i)} = \prod_{f=1}^{n} \mathrm{P(e_f|C_i)}$$

NBC estimates the conditional probability density function $\mathrm{P(e|C_i)}$ for a given feature value e_f for the *fth* feature using the frequency of observed instances around e_f. In our experiments, $\mathrm{P(e|C_i)}$ is computed by assuming Multinomial distribution.

5.3 Decision Tree Classifier

Decision tree (DT) classifier is a technique to solve a classification problem. DT can be used to create decision rules inferred from the training data and then these rules can be used to predict the value of a target variable. DT uses a learning classification algorithm to best fits the relationship between the feature set and class label of the input data. i.e, the goal of the DT learning classification algorithm is to build predictive model that accurately predict the class labels of previously unknown patterns. For more detail see [25].

5.4 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a classification method originally developed by R. A. Fisher. It is simple, mathematically robust and often produces models whose accuracy is as good as more complex methods. LDA maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability**.** For more detail see [24].

6 Experimental Results

To evaluate the effectiveness of two features selection methods on the diagnosis of erythemato-squamous diseases, we conducted three experiments. In these experiments the data set is divided into two disjoint subsets, namely training set 50% and test set 50%. First, for the sake of comparison, the four base classifiers, KNN, Naive Bayesian, LDA and decision tree, are used in the whole dataset without features selection. The accuracies of KNN, Naive Bayesian, LDA and decision tree are 84.36, 96.65, 96.09 and 97, respectively. In the second experiment, rough set using Johnson's algorithm is used as a features selection method and the reduced feature vector is used as input to these four base classifier. Their accuracies on test set are 58.1, 69.27, 75.42 and 73.74, respectively. Thirdly, a ranked features selection is applied to the raw data set and the final reduced feature vector is obtained by union of the discriminant features of each disease. The accuracies of these four classifiers on test set are 97.21, 98.32, 96.09, and 98.32 respectively. Table 3 and Fig. 2 summarize the results obtained. It is clear that, in the case of erythemato-squamous diseases, ranked features selection gives better results compared to the base classifiers without using features selection and rough set based features selection using Johnson's algorithm.

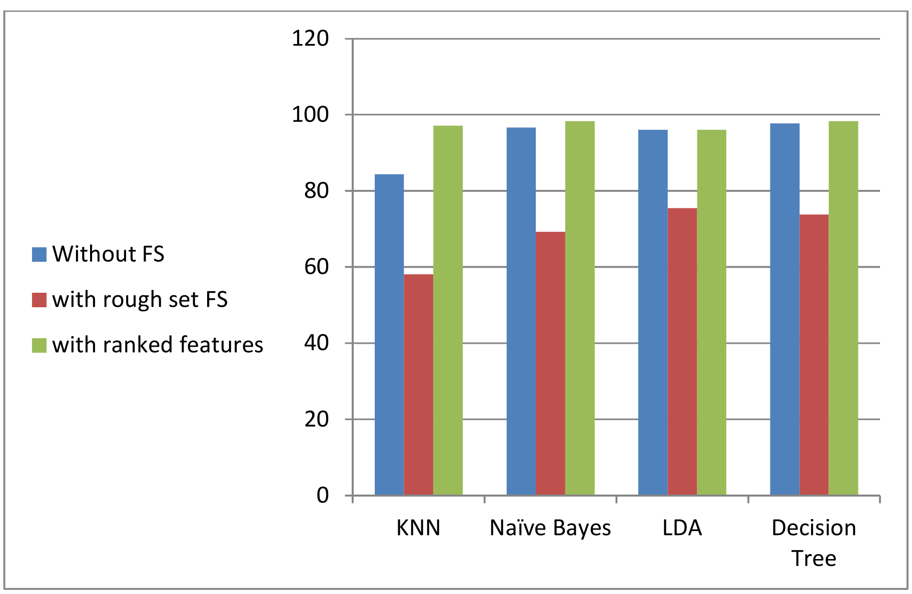

Fig. 2. The accuracies of different classifiers with and without feature selection (FS)

Table 4 gives the confusion matrices of the four classifiers using the ranked futures selection. In Table 5, the values of the statistical parameters of the classifiers using the ranked future selection method are given. We can conclude that Naive Bayesian and decision tree give superior performance when using ranked features. Also, from the confusion matrices we conclude that the Pityriasis rosea disease share some features with the Seboreic dermatitis disease. The number of misclassified patterns of Pityriasis rosea disease in the case of KNN, Naive Bayesian, decision tree and LDA are 5, 3, 2 and 7 respectively. Moreover, only one example from Pityriasis rubra pilaris disease is misclassified as Seboreic dermatitis disease when decision tree is used as base classifier.

Table 3. The accuracies of different classifiers with and without feature selection (FS)

	KNN	Naive Bayesian	LDA	Decision tree
Without features selection	84.3575	96.6480	96.0894	97.7654
Features selection by rough set using Johnson's algorithm	58.1006	69.2737	75.4190	73.7430
Features selection by ranked features	97.2067	98.3240	96.0894	98.3240

Table 4. Confusion matrices of the classifiers using the ranked futures selection

Classifier	Desired result	Output result					
		Psoriasis	Seboreic dermatitis	Lichen planus	Pityriasis rosea	Chronic dermatitis	Pityriasis rubra pilaris
KNN	Psoriasis	60	0	0	0	0	0
	Seboreic dermatitis	0	19	0	0	0	0
	Lichen planus	0	0	34	0	0	0
	Pityriasis rosea	0	5	0	25	0	0
	Chronic dermatitis	0	0	0	0	25	0
	Pityriasis rubra pilaris	0	0	0	0	0	11
Naive Bayesian	Psoriasis	60	0	0	0	0	0
	Seboreic dermatitis	0	19	0	0	0	0
	Lichen planus	0	0	34	0	0	0
	Pityriasis rosea	0	3	0	27	0	0
	Chronic dermatitis	0	0	0	0	25	0
	Pityriasis rubra pilaris	0	0	0	0	0	11
Decision tree	Psoriasis	60	0	0	0	0	0
	Seboreic dermatitis	0	19	0	0	0	0
	Lichen planus	0	0	34	0	0	0
	Pityriasis rosea	0	2	0	28	0	0

Table 4 continued…

	Chronic dermatitis	0	0	0	0	25	0
	Pityriasis rubra pilaris	0	1	0	0	0	10
LDA	Psoriasis	60	0	0	0	0	0
	Seboreic dermatitis	0	19	0	0	0	0
	Lichen planus	0	0	34	0	0	0
	Pityriasis rosea	0	7	0	23	0	0
	Chronic dermatitis	0	0	0	0	25	0
	Pityriasis rubra pilaris	0	0	0	0	0	11

Table 5. The values of the statistical parameters of the classifiers using the ranked futures selection

Classifier	Datasets	Statistical parameters (%)		
		Sensitivity	Specificity	Total classification accuracy
KNN	Psoriasis	100	100	97.2067
	Seboreic dermatitis	100	96.88	
	Lichen planus	100	100	
	Pityriasis rosea	83.33	100	
	Chronic dermatitis	100	100	
	Pityriasis rubra pilaris	100	100	
Naive Bayesian	Psoriasis	100	100	98.3240
	Seboreic dermatitis	100	98.12	
	Lichen planus	100	100	
	Pityriasis rosea	90	100	
	Chronic dermatitis	100	100	
	Pityriasis rubra pilaris	100	100	
Decision tree	Psoriasis	100	100	98.3240
	Seboreic dermatitis	100	98.75	
	Lichen planus	100	100	
	Pityriasis rosea	93.33	100	
	Chronic dermatitis	100	100	
	Pityriasis rubra pilaris	100	100	
LDA	Psoriasis	100	100	96.0894
	Seboreic dermatitis	100	95.63	
	Lichen planus	100	100	
	Pityriasis rosea	76.67	100	
	Chronic dermatitis	100	100	
	Pityriasis rubra pilaris	100	100	

7 Discussion

It is clear from the confusion matrices, Table 4, of the classifiers using the ranked futures selection that Pityriasis rosea disease has the same symptoms (features) of Seboreic dermatitis disease. So, the four classifiers often mistakenly classified some cases of Pityriasis rosea disease. Besides, in Table 5, the values of sensitivity and specifivity are 100% for the Psoriasis, Lichen planus, Chronic dermatitis and Pityriasis rubra pilaris diseases using the four classifiers. However, there are degradation in sensitivity and specifivity for both Seboreic dermatitis and Pityriasis rosea diseases which is another evidence that these two diseases share the same symptoms.

Moreover, the obtained results ensure that the using of ranked feature selection reduces the complexity of the classifiers' space and gives good results compared to the other machine learning methods mentioned in literature. On the other hand, ranked feature selection outperforms the feature selection using Johnson's algorithm.

For our study, we collected the data containing clinical and histopathological features of human from an open access data base namely, UCI (University of California at Irvine) machine learning database which is dedicated for academic research only. So, the issues of ethical matter (and the approval of an ethical committee related to this study) are not applicable here. The UCI must have taken care of those issues while preparing the data by performing the biological experiment and study through direct interaction with the concerned human. And we have acknowledged them here by citing the link (http://archive.ics.uci.edu/ml/).

8 Conclusion

Through this study we proposed an automatic diagnosis model of erythemato-squamous diseases using machine learning techniques. This model is straightforward using two-stage approach. First, two filter-based feature selection approaches namely rough set using Johnson's algorithm and features ranked, capable of searching for the optimal set of features and dimensionality reduction, are employed. Next, the diagnoses decisions are obtained by four different classification algorithms: k-nearest neighbors classifier, Naive Bayesian classifier, Linear discriminant analysis and decision tree. Comparison of the obtained results shows that the accuracies of the four base classifiers using ranked features outperformed those using rough set using Johnson's algorithm and the base classifiers without using feature selection. The accuracies of these four classifiers using ranked features on test sets (50% of the dataset) are 97.21, 98.32, 96.09, and 98.32, respectively. Considering the results, the ranked features method gives promising results for diagnosis of erythemato-squamous diseases and compared favorably with previously reported results.

Competing Interests

Author has declared that no competing interests exist.

References

[1] Güvenir HA, Demiröz G, Ilter N. Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals. Artif. Intell. Med. 1998;13:147–165.

[2] Pawlak Z, Skowron A. Rudiment of rough sets. Information Sciences. 2007;177(1):3-27.

[3] Blake CL, Merz CJ. UCI repository of machine learning databases; 1998.

[4] Güvenir HA, Emeksiz N. An expert system for the differential diagnosis of erythemato-squamous diseases. Expert Syst. Appl. 2000;18:43–49.

[5] Übeyli ED, Güler I. Automatic detection of erythematosquamous diseases using adaptive neuro-fuzzy inference systems. Computers in Biology and Medicine. 2005;35:421-433.

[6] Luukka P, Leppalampi T. Similarity classifier with generalized mean applied to medical data. Computers in Biology and Medicine. 2006;36:1026-1040.

[7] Polat K, Guneş S. The effect to diagnostic accuracy of decision tree classifier of fuzzy and k-NN based weighted pre-processing methods to diagnosis of erythemato-squamous diseases. Digital Signal Processing. 2006;16:922-930.

[8] Nanni L. An ensemble of classifiers for the diagnosis of erythematosquamous diseases. Neurocomputing. 2006;69:842-845.

[9] Luukka P. Similarity classifier using similarity measure derived from Yu's norms in classification of medical data sets. Computers in Biology and Medicine. 2007;37:1133-1140.

[10] Übeyli ED. Multiclass support vector machines for diagnosis of erythemato-squamous diseases. Expert Syst. Appl. 2008;35:1733-1740.

[11] Polat K, Guneş S. A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. Expert Syst. Appl. 2009;36:1587–159.

[12] Übeyli ED. Combined neural networks for diagnosis of erythematosquamous diseases. Expert Syst. Appl. 2009;36:5107-5112.

[13] Liu HW, Sun JG, Liu L, Zhang HJ. Feature selection with mutual information. Pattern Recognition. 2009;42:1330–1339.

[14] Karabatak M, Ince MC. A new feature selection method based on association rules for diagnosis of erythemato-squamous diseases. Expert Syst. Appl. 2009;36(10):12500-12505.

[15] Xie J, Wang C. Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. Expert Syst. Appl. 2011;38(5):5809-5815.

[16] Abdi MJ, Giveki D. Automatic detection of erythemato-squamous diseases using PSO–SVM based on association rules. Engineering Applications of Artificial Intelligence. 2013;26(1):603–608.

[17] Inbarani HH, Azar AH, Jothi G. Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis. Computer Methods and Programs in Biomedicine. 2014;113(1):175-185.

[18] Ravichandran KS, Narayanamurthy B, Ganapathy G, Ravalli S, Sindhura J. An efficient approach to an automatic detection of erythematosquamous diseases. Neural Comput. & Applic. 2014;25:105–114.

[19] Pawlak Z. Rough sets. International Journal of Parallel Programming. 1982;11(5):341–356.

[20] Pawlak Z. Why rough sets. In Proceedings of the Fifth IEEE International Conference on Fuzzy Systems, New Orleans, LA, USA. 1996;2:738–743.

[21] Pal SK, Skowron A. Rough fuzzy hybridization: A new trend in decision making. Springer-Verlag New York, Inc; 1999.

[22] Prati RC. Combining feature ranking algorithms through rank aggregation. The IEEE International Joint Conference on Neural Networks (IJCNN). 2012;1 – 8.

[23] Cover T, Hart P. Nearest neighbor pattern classification. IEEE Transactions on Information Theory. 1967;13:21–27.

[24] Duda RO, Hart PE, Stork DH. Pattern Classification (2nd ed.). Wiley Interscience; 2001.

[25] Quinlan JR. Induction of decision trees. Machine learning. 1986;1(1):81-106.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=1146&id=6&aid=9623