Expert Systems with Applications 36 (2009) 12500–12505

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# A new feature selection method based on association rules for diagnosis of erythemato-squamous diseases

Murat Karabatak [a,*], M. Cevdet Ince [b]

[a] Firat University, Department of Electronics and Computer Science, 23119 Elazig, Turkey
[b] Firat University, Department of Electric–Electronics Engineering, 23119 Elazig, Turkey

ARTICLE INFO

*Keywords:*
Association rules
Neural network
Erythemato-squamous
Feature selection

ABSTRACT

In this paper, a new feature selection method based on Association Rules (AR) and Neural Network (NN) is presented for the diagnosis of erythemato-squamous diseases. AR is used for reducing the dimension of erythemato-squamous diseases dataset and NN is used for efficient classification. The proposed AR+NN system performance is compared with that of other feature selection algorithms+NN. The dimension of input feature space is reduced from thirty four to twenty four by using AR. In test stage, 3-fold cross validation method is applied to the erythemato-squamous diseases dataset to evaluate the proposed system performances. The correct classification rate of proposed system is 98.61%. This research demonstrated that the AR can be used for reducing the dimension of feature space and proposed AR+NN model can be used to obtain fast automatic diagnostic systems for other diseases.



## 1. Introduction

Data classification process using knowledge obtained from known historical data has been one of the most intensively studied subjects in statistics, decision science and computer science. It has been applied in problems of medicine, social science management and engineering. Variable problems such as disease diagnosis, image recognition, and credit evaluation using classification techniques (Michie, Spiegelhalter, & Tayor, 1994). In medical and other domains, linear programming approaches were efficient and effective methods (Bennett & Mangasarian, 1992; Freed & Glover, 1981; Grinold, 1972; Smith, 1968). Recently, intelligent methods such as NN and support vector machines have been intensively used for classification tasks (Ryua, Chandrasekaranb, & Jacobc, 2007).

One of the application areas of analyzing database and pattern recognition is automated diagnostic systems. The aims of these studies are to assist doctors in making diagnostic decision. Thanks to modern facilities, very large databases can be collect in medicine. These databases need special techniques for analyzing, processing and effective use of them. Data mining and knowledge discovery in database are an approach to find relationships buried in data (Choua, Leeb, Shaoc, & Chenb, 2004). The methodologies consist of data visualization, machine learning and statistical techniques and these can be summarized as classification, prediction and clustering etc. (Curt, 1995).

In this study, an AR+NN method was proposed to be used in erythemato-squamous diseases diagnosis problem. This method consists of two stages. In the first stage, the input feature vector dimension is reduced by using AR. AR is used as feature selection method in this stage. This provides elimination of unnecessary data. In the second stage, neural network is used these inputs and classifies the erythemato-squamous diseases data.

The rest of the paper is organized as follows. In Section 2, we present the related works. In Section 3, we detailed feature selection and feature selection algorithms. In Section 4, we describe the erythemato-squamous diseases dataset. Our used algorithm and methods are given in details in Section 5 and in Section 6. We give the experimental results to show the effectiveness of our method in Section 7. Finally, we conclude this paper in Section 8.

## 2. Related works

There are many techniques and algorithms to predict and classify erythemato-squamous diseases. In Demiroz, Govenir, and Ilter (1998), a new classification algorithm called VHI5 (Voting Feature Intervals) was developed and applied to problem of differential diagnosis of erythemato-squamous diseases. In Govenir and Emeksiz (2000), an expert system for differential diagnosis erythemato-squamous diseases was presented incorporating decisions made by three classification algorithm: nearest neighbor classifier, näıve Bayesian classifier and voting feature intervals-5. In Ubeyli and Guler (2005), adaptive neuro-fuzzy inference system was presented for detection of erythemato-squamous diseases and obtained 95.5% classification accuracy. In Nanni (2006), LSVM, RS, B1_5, B1_10,

* Corresponding author. Tel.: +90 424 2370000x4292; fax: +90 424 2367064.
*E-mail addresses:* mkarabatak@firat.edu.tr (M. Karabatak), mcince@firat.edu.tr (M.C. Ince).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.04.073

B1_15, B2_5, B2_10 and B2_15 algorithms were used and obtained 97.22, 97.22, 97.5, 98.1, 97.22, 97.5, 97.8 and 98.3% classification accuracy, respectively. In Polat and Güneş (2006), a novel method was presented for differential diagnosis of erythemato-squamous disease. That method was based on fuzzy weighted pre-processing, k-NN (nearest neighbor) based weighted pre-processing and decision tree classifier. In this study, we used AR for feature selection and neural network for classify for the diagnosis of this disease and we obtained 98.6% classification accuracy.

## 3. Feature selection

Feature selection plays an important role in classifying systems such as neural networks. The aim of feature selection is to segregate the irrelevant and redundant attributes from a data set, thus the dimension of the data set will be reduced. Therefore, the complexity is reduced and the performance can be increased. The feature selection problem is investigated in many different research areas. One of the most popular methods for dealing with this problem is the principle component analysis (PCA) method (Chatterjee, Roychowdhury, & Chong, 1998; Karabatak, İnce, & Avcı, 2008). This method transforms the existing attributes into new ones considered to be crucial in classification. However from the viewpoint of maintaining data, this method is not desirable, as it needs to process all the data when new data is added.

Linear discriminant analysis (LDA) is another well-known feature selection method for dimensionality reduction and classification that projects high-dimensional data onto a low dimensional space where the data achieves maximum class separability (Duda, 2001).

Generally, both PCA and LDA methods reduce the $d$ dimensional input space to $d'$ where $d' < d$. The main drawback of this method is that it is not immune from distortion under transformation. Simply scaling some of the attributes can cause serious changes to the results. Recently, the feature selection problem has been dealt with intensely and some solutions have been proposed.

Forward feature selection (FFS) and backward feature selection (BFS) methods are commonly used in the literacy. These methods add or extract the features to the new feature vector. For each step the new feature vector is classified and the new feature vector is determined according to the best performance (Duda, Hart, & Stork, 2001; Genç et al., 2007; Hyvärinen, 1999; Hyvarinen & Oja, 1999; Kitler, 1978).

The implementation of these methods is sometimes enough hard when the dimension of the data set is considerably high. To eliminate the disadvantages of both FFS and BFS methods, a new method called plus-l takeaway-r is developed, which is the combination of the FFS and BFS methods (Stearns, 1976). This new method carried out the feature selection by running the FFS method for only $l$time and BFS method $r$ times. Method ends the procedure when the feature number $d$ reaches $d'$ feature number.

Pudil et al. proposed a feature selection algorithm which is similar to plus-l takeaway-r algorithm (Pudil, Ferri, Novovicova, & Kittler, 1994). The FFS and BFS steps in Pudil's algorithm are controlled dynamically. That is, while applying BFS step if the performance of the obtained subsets are worse, the BFS step is ignored and again FFS step is considered. This process repeated until obtaining the desired subset.

Individual feature selection (IFS) method reveals the effects of the each feature of a feature set to the performance of a classifier. However, it is thought that the method cannot produce successful results, it is seen that the method yields successful results for several database (Narendra & Fukunaga, 1977; Yu & Yuan, 1993). Moreover, branch and bound feature selection algorithms are based on genetic algorithms (Siedlecki & Sklansky, 1989; Vriesenga, 1995).

## 4. Erythemato-squamous disease dataset

The differential diagnosis of erythemato-squamous diseases is a difficult problem in dermatology. They all share the clinical features of erythema and scalling, with very little differences. erythemato-squamous diseases have six groups. These groups are psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis and pityriasis rubra pilaris. These diseases are frequently seen in the outpatient dermatology departments (Demiroz et al., 1998; Govenir & Emeksiz, 2000).

In this study, the erythemato-squamous diseases dataset was used and analyzed. This dataset was taken from UCI (University of California, Irvine) machine learning repository. There are 366 records in this database and each record has 34 attributes. Each record contains 12 clinical features and 22 histopathological features. These attributes are detailed in Table 1. Moreover, "age" feature was removed because in some record the "age" feature was missed (see Table 2).

**Table 1**
Erythemato-squamous diseases data description of attributes.

| Attribute number | Attribute description | Values of attributes |
|---|---|---|
| *Clinical attributes* | | |
| 1 | Erythema | 0–3 |
| 2 | Scaling | 0–3 |
| 3 | Definite borders | 0–3 |
| 4 | Itching | 0–3 |
| 5 | Koebner phenomenon | 0–3 |
| 6 | Polygonal papules | 0–3 |
| 7 | Follicular papules | 0–3 |
| 8 | Oral mucosal involvement | 0–3 |
| 9 | Knee and elbow involvement | 0–3 |
| 10 | Scalp involvement | 0–3 |
| 11 | Family history | 0–1 |
| 34 | Age | Linear |
| *Histopathological attributes* | | |
| 12 | Melanin incontinence | 0–3 |
| 13 | Eosinophils in the infiltrate | 0–3 |
| 14 | PNL infiltrate | 0–3 |
| 15 | Fibrosis of the papillary dermis | 0–3 |
| 16 | Exocytosis | 0–3 |
| 17 | Acanthosis | 0–3 |
| 18 | Hyperkeratosis | 0–3 |
| 19 | Parakeratosis | 0–3 |
| 20 | Clubbing of the rete ridges | 0–3 |
| 21 | Elongation of the rete ridges | 0–3 |
| 22 | Thinning of the suprapapilary epidermis | 0–3 |
| 23 | Spongiform pustule | 0–3 |
| 24 | Munro microabcess | 0–3 |
| 25 | Focal hypergranulosis | 0–3 |
| 26 | Disappearance of the granular layer | 0–3 |
| 27 | Vacuolisation and damage of basal layer | 0–3 |
| 28 | Spongiosis | 0–3 |
| 29 | Saw-tooth appearance of retes | 0–3 |
| 30 | Follicular horn plug | 0–3 |
| 31 | Perifollicular parakeratosis | 0–3 |
| 32 | Inflammatory monoluclear inflitrate | 0–3 |
| 33 | Band-like infiltrate | 0–3 |

**Table 2**
Erythemato-squamous diseases class distribution.

| Class code | Class | Number of instances |
|---|---|---|
| 1 | Psoriasis | 112 |
| 2 | Seboreic dermatitis | 61 |
| 3 | Lichen planus | 72 |
| 4 | Pityriasis rosea | 49 |
| 5 | Cronic dermatitis | 52 |
| 6 | Pityriasis rubra pilaris | 20 |
| Total observation (N) | | 366 |

## 5. Preliminaries

### 5.1. Association rules

In order to see how AR can be used in erythemato-squamous diseases data with NN, first of all it is needed to define AR. AR find interesting associations and/or relationships among large set of data items. AR show attributes value conditions that occur frequently together in a given dataset. They allow capturing all possible rules that explain the presence of some attributes according to the presence of other attributes. A typical and widely-used example of association rules mining is Market Basket Analysis (Agrawal et al., 1993).

Let $I = (i_1, i_2, \ldots, i_m)$ be a set of literals, called items. Let D be a database of transaction, where each transaction T is a set of items such that $T \subseteq I$. For a given itemset $X \subseteq I$ and a given transaction T, we say that T contains X if and only if $X \subseteq I$. The support count of an itemset X is defined to be $sup_x$= the number of transactions in D that contain X. we say that an itemset X is large, with respect to a support threshold of s%, if $sup_x \geqslant |D|$ x s%, where |D| is the number of transactions in the database D. An association rules is an implication of the form "$X \Rightarrow Y$", where $X \subseteq I, Y \subseteq I$ and $X \cap Y = \emptyset$. The AR "$X \Rightarrow Y$" is said to hold in database D with confidence c% if no less than c% of the transactions in D that contain X also contain Y. The rule $X \Rightarrow Y$ has support s% in D if $sup_{X \cup Y} = |D| \times s\%$ (Agrawal et al., 1993). Thus, AR aims at discovering the patterns of co-occurrence of attributes in a database. For instance, an association rule in a supermarket basket data may be in 10% of transactions, 85% of people buying milk also buy yoghurt in that transaction. AR may be useful in many applications such as business applications, market basket analysis, store layout and promotion on the items, telecommunication alarm correlation, university course enrollment, texture and image processing (Karabatak, Şengür, İnce, & ve Türkoğlu, 2006).

#### 5.1.1. Apriori algorithm

The Apriori algorithm is a state of the art algorithm most of the association rule algorithms are somewhat variations of this algorithm (Agrawal et al., 1993). The Apriori algorithm works iteratively. It first finds the set of large 1-itemsets, and then set of 2-itemsets, and so on. The number of scan over the transaction database is as many as the length of the maximal item set. Apriori is based on the following fact: the simple but powerful observation leads to the generation of a smaller candidate set using the set of large item sets found in the previous iteration. The Apriori algorithm presented in Agrawal and Srikant (1994) is given as follows;

```
Apriori()
L 1={large 1-itemsets}
k=2
while L k-1 ≠φ do
begin
  Ck = apriori_gen (Lk-1 )
  for all transactions t in D do
  begin
    Ct = subset(Ck, t)
    for all candidatec ∈ Ctdo
    c.count = c.count+1
    end
    Lk = {c ∈ Ck|c.count ⩾ minsup}
    k=k+1
end
```

Apriori first scans the transaction databases D in order to count the support of each item *i* in *I*, and determines the set of large 1-itemsets. Then, iteration is performed for each of the computation of the set of 2-itemsets, 3-itemsets, and so on. The $k^{th}$ iteration consists of two steps (Rushing, Ranganath, Hinke, & Graves, 2002);

- Generate the candidate set $C_k$ from the set of large (k-1)-itemsets, $L_{k-1}$.
- Scan the database in order to compute the support of each candidate itemset in $C_k$

The candidate generation algorithm is given as follows;

```
Apriori_gen(Lk-1)
Ck = φ
for all itemsets X ∈ Lk-1 and Y ∈ Lk-1 do
  if X1 = Y1 ∧ ... ∧ Xk-2 = Yk-2 ∧ Xk-1 < Yk-1 then begin
  C = X1X2 ... Xk-1Yk-1
  add C to Ck
end
delete candidate itemsets in Ck whose any subset is
  not in LK-1
```

The candidate generation procedure computes the set of potentially large k-itemsets from the set of large $(k-1)$-itemsets. A new candidate k-itemset is generated from two large $(k-1)$-itemsets if their first $(k-2)$ items are the same. The candidate set $C_k$ is a superset of the large k-itemsets. The candidate set is guaranteed to include all possible large k-itemsets because of the fact that all subsets of a large itemset are also large. Since all large itemsets in $L_{k-1}$ are checked for contribution to candidate itemset, the candidate set $C_k$ is certainly a superset of large k-itemsets. After the candidates are generated, their counts must be computed in order to determine which of them are large. This counting step is really important for the efficiency of the algorithm, because the set of the candidate itemsets may be possibly large. Apriori handles this problem by employing a hash tree for storing the candidate. The candidate generation algorithm is used to find the candidate itemsets contained in a transaction using this hash tree structure. For each transaction T in the transaction database D, the candidates contained in T are found using the hash tree, and then their counts are incremented. After examining all transaction in D, the ones that are large are inserted into $L_k$ (Karabatak et al., 2006).

### 5.2. Neural networks

Neural Networks (NNs) are biologically inspired and mimic the human brain. They are occurring neurons. These neurons are connected each other with connection links. These links have weights. They multiplied with transmitted signal in network. The output of each neuron is determined by using an activation function such as sigmoid and step. Usually nonlinear activation functions are used. NNs are trained by experience, when applied an unknown input to the network it can generalize from past experiences and product a new result (Hanbay, Turkoglu, & Demir, 2008; Bishop, 1996; Haykin, 1994). The output of the neuron net is determined by Eq. (1). A simple artificial neuron model is shown in Fig. 1.

$$y(t+1) = a\left(\sum_{j=1}^{m} w_{ij}x_j(t) - \theta_i\right) \text{ and } f_i \triangleq \text{net}_i = \sum_{j=1}^{m} w_{ij}x_j - \theta_i \qquad (1)$$

where, X = (X1, X2... Xm ) represents the m input applied to the neuron, Wi represent the weights for input Xi, $\theta$i is a bias value, a(·) is activation function. NNs models have been used for pattern matching, nonlinear system modeling, communications, electrical

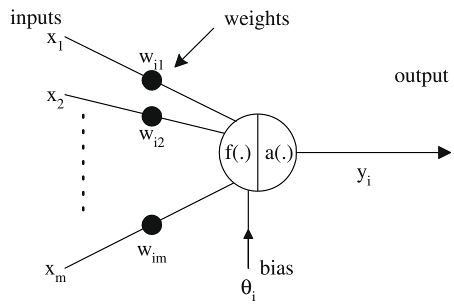


**Fig. 1.** Artificial neuron model.

and electronics industry, energy production, chemical industry, medical applications, data mining and control because of their parallel processing capabilities. When designing a NN model a number of considerations must be taken into account. First of all the suitable structure of the NN model must be chosen, after this the activation function, the number of layers and the number of units in each layer must be chosen. Generally desired model consists of a number of layers. The most general model assumes complete interconnections between all units. These connections can be bidirectional or unidirectional.

## 6. Applications

Feature extraction is the key for pattern recognition so that it is arguably the most important component of designing the intelligent system based on pattern recognition since even the best classifier will perform poorly if the features are not chosen well. A feature extractor should reduce the pattern vector (i.e., the original waveform) to a lower dimension, which contains most of the useful information from the original vector (Sengur, 2008). Fig. 2 shows the proposed automatic detection system block diagram. It consists of two parts: (a) feature extraction and reduction with AR (b) classification with NN.

***AR Layer:*** AR is a method to find the associations and/or relationships among items in large databases. So, we can use it to detect relations among inputs of any system and later eliminate some unnecessary inputs. We propose two different techniques to eliminate inputs. These are named as AR1 and AR2, respectively (Karabatak & Ince, in press).

***AR1:*** The AR1 technique uses all input parameters and their all records to find relations among the input parameters. If we find rules that have enough support value and high confidence value, we can eliminate some inputs thanks to these rules. In the AR form ($X \Rightarrow Y$), Y itemset also depend on X itemset. Thus, we can eliminate all items in Y itemset. So, these are not necessary to be used in NN inputs.

***AR2:*** Especially, we can use AR2 with classification problems. AR2 uses all input parameters but not all their records. We find only large itemsets for every class. All items in these large itemsets are most important items to classification. Thus, we can only use these items to classify all data. If an item of large itemset of any class is large in other classes and it has different value, this item must be used as NN inputs.

In this study, we used AR2 but we did not use AR1 to reduce the number of NN inputs for erythemato-squamous diseases detection problem. Because, we could not find AR1 form ($X \Rightarrow Y$) which has enough support and confidence value. So we only used AR2.

Erythemato-squamous diseases has six classes. These are psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis and pityriasis rubra pilaris classes. Using AR2, we found large itemsets of all classes given as follows;

Large Itemset of psoriasis class:
6-7-8-12-13-15-25-27-28-29-30-31-33
Large Itemset of seboreic dermatitis class:
5-6-7-8-9-11-12-15-20-22-24-25-26-27-29-30-31-33
Large Itemset of lichen planus class:
7-9-10-11-14-15-20-21-22-23-24-30-31
Large Itemset of pityriasis rosea class:
6-7-8-9-10-11-12-13-15-20-21-22-23-24-25-27-29-30-31-33
Large Itemset of cronic dermatitis class:
5-6-8-9-10-11-12-13-14-20-22-23-24-25-26-27-29-30-31-33
Large Itemset of pityriasis rubra pilaris class:
5-6-8-12-13-15-20-21-22-23-24-25-26-27-29-33

According to this large itemsets, we can say that 6th, 7th, 8th, 12th, 13th, 15th, 25th, 27th, 28th, 29th, 30th, 31th and 33th input parameters already can define psoriasis class and etc. These parameters are the most important parameters for erythemato-squamous diseases detection problem. So, we only used these inputs in NN.

***NN Layer:*** Multi-layer perceptron (MLP): the intelligent classification is realized in this layer by using features, which are obtained from AR layer. The training parameters and the structure of the MLP used in this study are shown in Table 3. These were selected for the best performance, after several experiments. Fig. 3 shows the AR2+NN training performance.

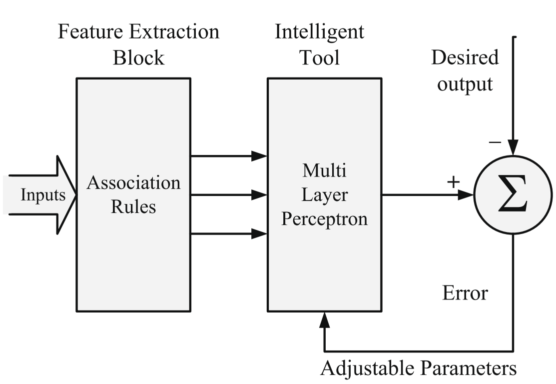


**Fig. 2.** The block diagram of the hybrid system.

**Table 3**
MLP architecture and training parameters.

| | |
|---|---|
| *Architecture* | |
| The number of layers | 3 |
| The number of neuron on the layers | Input:24<br>Hidden:35<br>Output: 1 |
| The initial weights and biases | Random |
| Activation Functions | Tangent-sigmoid<br>Tangent-sigmoid<br>Linear |
| *Training parameters* | |
| Learning rule | Levenberg–Marquardt<br>Back-propagation |
| Sum-squared error | 0.05 |

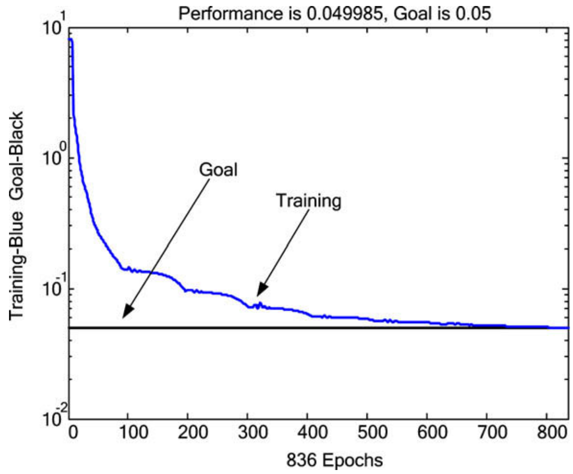


**Fig. 3.** Neural Network training performance.

## 7. Modeling results

This study was performed using erythemato-squamous dataset with 24 attributes and 366 records. In test stage, 3-fold cross validation method (X1, X2, X3) was applied and average values were calculated. The performance comparison and correct classification rates are tabulated in Table 4.

As shown in Table 4, the best classification performance was obtained with AR2+NN with 24 inputs and its average correct classification rate is 98.61%. The average correct classification rate of NN with 33 inputs is 97.7%. So, we can use AR2+NN for best classification performance.

As shown in Table 5, the best feature selection algorithm is AR2 method for classification of erythemato-squamous diseases.

**Table 4**
Performance comparison for erythemato-squamous diseases detection using NN, and AR2+NN.

| The classifier | Training data | Test data | Correct classification rate (%) |
|---|---|---|---|
| NN (33, 35, 1) | X1+X2 | X3 | 97.90 |
| | X1+X3 | X2 | 98.14 |
| | X2+X3 | X1 | 97.06 |
| | Average | | *97.70* |
| AR2 + NN (24, 26, 1) | X1+X2 | X3 | 99.68 |
| | X1+X3 | X2 | 98.43 |
| | X2+X3 | X1 | 97.73 |
| | Average | | *98.61* |

**Table 5**
Performance comparison for erythemato-squamous diseases detection using NN, AR2+NN and other feature selection algorithms+NN.

| Feature selection methods + classifier | Selected attributes | Correct classification rate (%) |
|---|---|---|
| Forward +YSA (24, 26, 1) | 2-5-6-7-8-9-10-11-13-14-15-20-21-22-23-24-25-26-27-28-29-30-31-33 | 97,8 |
| Backward+ YSA (24, 26, 1) | 2-3-4-5-7-8-9-13-14-16-17-19-21-22-24-25-26-27-28-29-30-31-32-33 | 96,9 |
| Plus-l, takeaway-r + YSA (24, 26, 1) | 2-5-6-7-8-9-10-11-12-14-15-20-21-22-23-24-25-26-27-28-29-30-31-33 | 97,5 |
| Individual + YSA (24, 26, 1 ) | 2-3-4-5-6-7-8-9-10-11-12-14-15-16-18-20-21-25-27-28-29-30-31-33 | 96,1 |
| Pudil +YSA (24, 26, 1) | 2-4-5-7-8-9-10-11-12-15-16-17-18-19-21-22-24-25-26-28-29-30-32-33 | 96,3 |
| Random + YSA (24, 26, 1) | 2-3-5-7-9-11-12-13-14-16-17-19-20-21-22-23-24-25-26-28-29-30-31-32 | 96,1 |
| *AR2 + YSA (24, 26, 1)* | *5-6-7-8-9-10-11-12-13-14-15-20-21-22-23-24-25-26-27-28-29-30-31-33* | *98,6* |
| Only NN (33, 35, 1) | 1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27- 28-29-30-31-32-33 | 97,7 |

## 8. Conclusions

In this study, a hybrid method for diagnosis of erythemato-squamous diseases based on Association Rules (AR) and Neural Network (NN) is presented (Karabatak & Ince, in press; Sengur, 2008). Feature extraction is the key for pattern recognition and classification. The best classifier will perform poorly if the features are not chosen well. A feature extractor should reduce the feature vector to a lower dimension, which contains most of the useful information from the original vector. So, AR is used for reducing the dimension of erythemato-squamous diseases dataset and NN is used for intelligent classification. The proposed AR+NN system performance is compared with NN model. The dimension of input feature space is reduced from 33 to 24 by using AR. In test stage, 3-fold cross validation method was applied to the erythemato-squamous diseases dataset to evaluate the proposed system performances. The correct classification rate of proposed system is 98.61% for 24 inputs. This research demonstrated that the AR can be used for reducing the dimension of feature vector and proposed AR+NN model can be used to obtain efficient automatic diagnostic systems for other diseases.

## References


Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of 20th international conference on very large databases* (pp. 487–499). Santiago, Chile.

Agrawal R., Imielinski T., & Swami A. (1993). Mining association rules between sets of items in large databases, In *Proceedings of ACM SIGMOD international conference on management of data*, Washington, DC.

Bennett, K. P., & Mangasarian, O. L. (1992). Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software, 1*, 23–34.

Bishop, C. M. (1996). *Neural networks for pattern recognition*. Oxford: Clarendon Press.

Chatterjee, C., Roychowdhury, V. P., & Chong, E. K. P. (1998). On relative convergence properties of principal component analysis algorithms. *IEEE Transactions on Neural Networks, cilt-9*(2), 319–329.

Choua, S.-M., Leeb, T.-S., Shaoc, Y. E., & Chenb, I.-F. (2004). Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications, 27*, 133–142.

Curt, H. (1995). The devil's in the detail: Techniques: Tools, and applications for database mining and knowledge discovery-part. *Intelligent Software Strategies*, 1–15.

Demiroz, G., Govenir, H. A., & Ilter, N. (1998). Learning differential diagnosis of eryhemato-squamous diseases using voting feature intervals. *Artificial Intelligence in Medicine, 13*, 147–165.

Duda, R. et al. (2001). *Pattern classification*. John Wiley. pp. s117–s124.

Duda, R., Hart, P., & Stork, D. (2001). *Pattern classification* (2nd ed.). New York: John Wiley and Sons.

Freed, E., & Glover, F. (1981). A linear programming approach to the discriminant problem. *Decision Sciences, 12*(1), 68–74.

Genç, H. M., Çataltepe Z., & Pearson, T. (2007). Yeni Bir Temel/Bağımsız Bileşen Analizi(TBA/BBA) Tabanlı uÜznitelik Seçme Yöntemi, IEEE 15. Sinyal İşleme ve İletişim Uygulamaları Kutultayı, Eskişehir.

Govenir, H. A., & Emeksiz, N. (2000). An expert system for the differential diagnosis of erythemato-squamous diseases. *Expert System with Applications, 18*, 43–49.

Grinold, R. C. (1972). Mathematical programming methods of pattern classification. *Management Science, 19*(3), 272–289.

Hanbay, D., Turkoglu, İ., & Demir, Y. (2008). An expert system based on wavelet decomposition and neural network for modeling Chua's circuit, *Expert Systems with Applications*, *34*(4), 2278–2283.

Haykin, S. (1994). *Neural networks, a comprehensive foundation*. New York: Macmillan College Publishing Company Inc.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks, 10*(3), 626–634.

Hyvarinen, A., & Oja, E. (1999). Independent component analysis: Algorithms and applications. *Neural Networks, 13*, 411–430.

Karabatak, M., & Ince, M. C. (in press). An expert system for detection of breast cancer based on association rules and neural network, *Expert Systems with Applications* (corrected proof, available online 6 march 2008).

Karabatak, M., Şengür, A., İnce, M.C., & ve Türkoğlu, İ. (2006). Association rules for texture classification, IMS.
Karabatak, M., İnce, M. C., & Avcı, E. (2008). Göğüs kanseri için temel bileşen analizi yöntemi tabanlı uzman teşhis sistemi, IEEE 16. Sinyal İşleme ve İletişim Uygulamaları Kutultayı, Didim-Aydın.
Kitler, J. (1978). Feature set search algorithm, In C. H. Chen (Ed.), *Pattern recognition and signal processing* (pp. 41–60). Hollanda.
Michie, D., Spiegelhalter, D. J., & Tayor, C. C. (1994). *Machine learning, neural and statistical classification*. London: Ellis Horwood.
Nanni, L. (2006). An ensemble of classifiers for the diagnosis of erythemato-squamous diseases. *Neurocomputing, 69*, 842–845.
Narendra, P. M., & Fukunaga, K. (1977). A Branch and Bound algorithm for feature subset selection. *IEEE Transaction on Computers, cilt-26*(9), s917–s922.
Polat, K., & Güneş, S. (2006). 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, 16*(6), 922–930.
Pudil, P., Ferri, F. J., Novovicova, J., & Kittler, J. (1994). Floating search methods for feature selection with nonmonotonic criterion functions. In *IEEE 12th international conference on pattern recognition* (Vol. cilt-II, pp. s279–s283).
Rushing, J. A., Ranganath, H. S., Hinke, T. H., & Graves, S. J. (2002). Image segmentation using association rule features. *IEEE Transactions on Image Processing, 11*, 558–566.
Ryua, Y. U., Chandrasekaranb, R., & Jacobc, V. S. (2007). Breast cancer prediction using the isotonic separation technique. *European Journal of Operational Research, 181*, 842–854.
Sengur, Abdulkadir (2008). An expert system based on principal component analysis, artificial immune system and fuzzy k-NN for diagnosis of valvular heart diseases. *Computers in Biology and Medicine, 38*(3), 329–338.
Sengur, Abdulkadir (2008). An expert system based on linear discriminant analysis and adaptive neuro-fuzzy inference system to diagnosis heart valve diseases. *Expert Systems with Applications, 35*(1-2), 214–222.
Siedlecki, W., & Sklansky, J. (1989). Anote on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters, 10*, 335–347.
Smith, F. W. (1968). Pattern classifier design by linear programming. *IEEE Transactions on Computers, C-17*(4), 367–372.
Stearns, S. D. (1976). On selecting features for pattern classifiers. In *Third international conference on pattern recognition* (pp. s71–s75).
Ubeyli, E. D., & Guler, I. (2005). Automatic detection of erythemato-squamous diseases using adaptive neuro-fuzzy inference systems. *Computers in Biology and Medicine, 35*, 421–433.
Vriesenga, M. R. (1995). Genetic selection and neural modeling for designing pattern classifier. Doctora thesis, University of California-Irvine.
Yu, B., & Yuan, B. (1993). A more efficient branch and bound algorithm for feature subset selection. *Pattern Recognition, 26*(6), 883–889.