Journal of Advanced Computer Science & Technology, 4 (1) (2015) 148-155
www.sciencepubco.com/index.php/JACST

doi: 10.14419/jacst.v4i1.4353
Research Paper

New method for rapid diagnosis of Hepatitis disease based on reduction feature and machine learning

Nahid Khorashadizade *, Hassan Rezaei

Department of Computer Science, Faculty of Mathematics, University of Sistan and Baluchestan, Iran
**Corresponding author E-mail: nahid.kh@pgs.usb.ac.ir*

Abstract

Hepatitis disease is caused by liver injury. Rapid diagnosis of this disease prevents its development and suffering to cirrhosis of the liver. Data mining is a new branch of science that helps physicians for proper decision making. In data mining using reduction feature and machine learning algorithms are useful for reducing the complexity of the problem and method of disease diagnosis, respectively. In this study, a new algorithm is proposed for hepatitis diagnosis according to Principal Component Analysis (PCA) and Error Minimized Extreme Learning Machine (EMELM). The algorithm includes two stages; in reduction feature phase, missing records were deleted and hepatitis dataset was normalized in [0,1] range. Thereafter, analysis of the principal component was applied for reduction feature. In classification phase, the reduced dataset is classified using EMELM. For evaluation of the algorithm, hepatitis disease dataset from UCI Machine Learning Repository (University of California) was selected. The features of this dataset reduced from 19 to 6 using PCA and the accuracy of the reduced dataset was obtained using EMELM. The results revealed that the proposed hybrid intelligent diagnosis system reached the higher classification accuracy and shorter time compared with other methods.

Keywords: *Classification; Dimensionality reduction; Error minimized extreme learning machine*; *Hepatitis disease; Principal component analysis.*

1. Introduction

Hepatitis affects the life of millions of people all over the world every year and leading to the disturbance of their life and in some cased their death. Therefore, it is known as one of the most significant and dangerous diseases. Because of limited causative agents and consequently late diagnosis of the disease, it is one of the major problems of health sectors in the world. Hepatitis may be caused by some assigns, including bacteria and digestion toxins, hereditary diseases, drug side effects and alcoholic drinks consumption; but viruses are the major factors [3]. As the virus penetrates into the body, it reaches the liver and grows and reproduces inside the organ. Virus entrance is followed by inflammation and damage and prevents the functionality of the liver.

Data mining is used extensively for rapid diagnosis of hepatitis disease. A large number of researchers have suggested smart diagnosis methods for diseases, including hepatitis using data mining. Huang and Wang [9] proposed an approach based on a genetic algorithm for feature selection and designing the sorting parameters of support vector machine. Dogantekin et al [7] introduced an automatic diagnosis system of linear discriminating analysis and adaptive network based on fuzzy inference system for hepatitis diagnosis which is a combination of feature extraction and classification. Calisir and Dogantekin [3] used PCA method for feature reduction of hepatitis dataset and thus decreased the features of this disease to 10.

In the present study a hybrid method of PCA-EMELM are proposed to obtain maximum accuracy with a minimum time of hepatitis diagnosis. The structure of the paper is as follows: in section 2 the concepts of PCA and Single hidden Layer feed Forward Network (SLFNs) Learning methods will be introduced. In section 3 the proposed hybrid method is explained. Section 4 presents experimental results of the proposed method on UCI hepatitis dataset and with comparison with other methods and section 5 is the conclusion.

2. Basic concepts

2.1. Principal component analysis algorithm

If the dataset has no variance in one dimension or there is a dimension with smaller variance relative to other dimensions, elimination of such dimension will not lead to losing significant information. But if dataset has a substantial variance in all dimensions, none of these dimensions cannot be removed. Because, we lose some of the information. Such data points can be imaged along the maximum variance dataset and consequently decrease the dimension without losing information.
PCA algorithm is the most common method of feature extraction [3]. This method is founded on linear algebra that reduces the dimension using principal components determination. Principal components are the eigenvectors of the covariance matrix of the data. The maximum variance of the dataset is along eigenvectors corresponding to the largest eigenvalue. In this method, data is mapped from a complex and multi-dimensional space into a simple space with the same dimensions and the considerable amount of data is stored at the first dimension. Altogether, PCA method is the orthogonal linear conversion in vector space with ordered dimensions in such a way that the first principal component has the maximum possible variance [13]. Therefore, the order of dataset components is remained according to the highest impact on the variance. The feature reduction using the PCA method is described in the following.
For a given p-dimensional data set X, m principal axes $T_1, T_2, \dots, T_m$ where $1 \leq m \leq p$, are orthonormal axes with maximum variance in projected space that can be given by the m leading eigenvectors of the sample covariance matrix Equation (1).

$$S = (1/N)\sum_{i=1}^{N}(x_i - \mu)(x_i - \mu)^T \quad (1)$$

In Equation (1) $x_i \in X$, μ is the mean and N is the number of samples. The ith maximum eigenvalue of S, λ_i, is calculated using the following equation:
$ST_i = \lambda_i T_i, \quad i = 1, \dots, m$

Reduced data set y of Equation (2) is obtained [13].

$$y = [y_1, y_2, \dots, y_m] = T^T X \quad (2)$$

Those vectors with minimum dimensions are significant if they can produce separability more or equal than previous vectors. If less significant vectors are removed, the number of dimensions can be reduced using abovementioned equation. These dimensions should be removed which are independent, orthogonal and with maximum impact on separability.

2.2. Single hidden layer feed forward networks (SLFNs)

Feed forward neural networks are ideal classifiers for non-linear mapping. SLFNs are feed forward neural networks with only one hidden layer. Hong et al [14] suggested a new learning algorithm for SLFNs called Extreme Learning Machine (ELM) that reduced required time of network learning to a great extent. In this algorithm, the nodes of hidden layer are selected at random and output weights are determined analytically through generalized inverse operation on the hidden layer output matrices. This algorithm is discussed briefly in following [6].
ELM is a learning method for SLFN which selects the parameters of hidden layer at random. Standard SLFN with m hidden neuron is described mathematically in Equation (3):

$$f_m(x) = \sum_{i=1}^{m} \beta_i G(a_i, b_i, x), \qquad a_i, x \in R^n \quad (3)$$

b_i is the ith Hidden neuron bias and $a_i = [a_{i1}, a_{i2}, \dots, a_{in}]$ is the weight vector showing that the ith hidden neuron is connected to input neurons. $\beta_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ip})$ is the weight vector of the ith hidden neuron connecting it to the output neurons. $G(a_i, b_i, x)$ is the output of ith hidden neuron with x as the input that is defined for additive hidden node with activating function of $g(x): R \rightarrow R$ (e. g. Sigmoid, sinuous etc) like below:
$G(a_i, b_i, x) = g(a_i . x + b_i), \qquad b_i \in R$

In $a_i . x$, "." is the internal multiplier.
For a hidden node, Radial Basis Function (RBF) with activating function $g(x): R \rightarrow R$ (like Gaussian) $G(a_i, b_i, x)$, is defined as below:
$G(a_i, b_i, x) = g(||x - a_i||/b_i)$

2.2.1. Extreme learning machine (ELM)

If there are N samples of (X_i, Y_i) where $X_i = [X_{i1}, X_{i2}, \dots, X_{in}]^T \in R^n$ are the inputs of the network and $Y_i = [Y_{i1}, Y_{i2}, \dots, Y_{ip}]^T \in R^P$ are the outputs, the ELM algorithm is summarized in three steps [11]:

1) Random production of hidden layer parameters (a_i, b_i);
2) Calculation of output matrix of hidden layer, H;

$$H(a_1, \dots, a_m, b_1, \dots, b_m, X_1, \dots, X_N) = \begin{bmatrix} G(a_1, b_1, X_1) & \cdots & G(a_m, b_m, X_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, X_N) & \cdots & G(a_m, b_m, X_N) \end{bmatrix}$$

3) Calculation of output weight vector $\hat{\beta}$ using the $\hat{\beta} = H^+Y$;

In this network, training error should be the minimum for better generalization. The problem is briefly solved with minimum squares error $H\beta = Y$ where:

$$\beta = \begin{bmatrix} \beta_1^T \\ . \\ . \\ \beta_m^T \end{bmatrix}_{m \times p}, Y = \begin{bmatrix} Y_1^T \\ . \\ . \\ Y_N^T \end{bmatrix}_{N \times p}$$

Solving the norm of minimum squares is reduced to $\hat{\beta} = H^+Y$ where H^+ is the Moore-Penrose generalized inverse of H matrix and is calculated as $H^+ = (H^TH)^{-1}H^T$ [6].

H matrix, as the output of the hidden layer, is determined by optimization of gradient descent in classical SLFNs. In these matrices input weights a_i, weights of hidden layer β_i and bias parameters b_i are set continuously. As a consequence, learning time increases. In the ELM learning algorithm, obtaining the parameters of hidden layer is not needed [6], therefore, they are implemented easily with high speed and good modification ability.

The ELM learning algorithm needs analytical calculation of output weights of SLFNs. Although ELM is simply applicable in machine learning, there are still two issues to be discussed [8]:

- The number of hidden nodes of SLFNs is the only factor that is determined by the user. Users usually have to select the number of nodes on hidden layer by try and error method and this is a challenging subject.
- Whether the computation complexity of ELM can be further reduced, especially if given a large number of training data and if a large number of hidden nodes required

Therefore, one of the problems in neural network researches is determination of automatic architecture of the network. A simple and efficient approach was proposed by Feng et al [8] based on automatic growth of hidden nodes and Rapid increase of learning for output weight. This approach is known as EMELM which adds random hidden nodes one by one or group by group with different sizes. In addition, during network growth the complexity of calculations is reduced considerably by updating the output weights increasingly.

2.2.2. Error minimized extreme learning machine (EMELM) algorithm

$\{(X_i, Y_i)\}_{i=1}^N$ is the representative of the training dataset and the number of maximum nodes of hidden layer is shown as L_{max}, small positive number is L_0 and expected learning error is shown with ϵ. This algorithm can be divided into two phases [8]:

Phase I:

1) A small group of hidden layers $\{(a_i, b_i)\}_{i=1}^{L_0}$ is considered for SLFN; L_0 is a small positive number determined by the user;
2) Calculation of output matrix of hidden layer H_1;

$$H_1 = \begin{bmatrix} G(a_1, b_1, X_1) & \cdots & G(a_{L_0}, b_{L_0}, X_1) \\ \cdots & \cdots & \cdots \\ G(a_1, b_1, X_N) & \cdots & G(a_{L_0}, b_{L_0}, X_N) \end{bmatrix}_{N \times L_0} \quad (4)$$

3) Calculation of corresponding output error;

$$E(H_1) = \|H_1 H_1^+ Y - Y\| \quad (5)$$

Phase II: Let $k = 0$.

While $L_k < L_{max}$, $E(H_k) > \epsilon$:

1) $k = k + 1$;

2) δL_{k-1}hidden node is added to SLFN at random and the new number of total hidden nodes will be $L_k = L_{k-1} + \delta L_{k-1}$. The corresponding output matrix of hidden layer $H_{k+1} = [H_k, \delta H_k]$ which is:

$$\delta H_k = \begin{bmatrix} G(a_{L_{k-1}+1}, b_{L_{k-1}+1}, X_1) & \cdots & G(a_{L_k}, b_{L_k}, X_1) \\ \cdots & \cdots & \cdots \\ G(a_{L_{k-1}+1}, b_{L_{k-1}+1}, X_N) & \cdots & G(a_{L_k}, b_{L_k}, X_N) \end{bmatrix}_{N \times \delta L_{k-1}} \quad (6)$$

3) Output weights of β will be updated with Equation (7) like below:

$$\begin{aligned} D_k &= ((I - H_k H_k^+)\delta H_k)^+ \\ U_k &= H_k^+ (I - \delta H_k D_k) \\ \beta_{k+1} &= H_{k+1}^+ Y = [U_k D_k]^T Y \end{aligned} \quad (7)$$

It is not necessary to keep constant the number of new hidden nodes at any phase. Equation (8) shows this issue.

$$\delta L_k \neq \delta L_{k+1} \quad (8)$$

3. Proposed hybrid model

The proposed algorithm in this research is called PCA-EMELM and includes two phases: (phase1) the feature reduction by PCA, (phase2) the classification by EMELM. First of all, records with missing values of data set hepatitis are removed and then the data set is normalized in [0,1] range. Applying PCA method, the features of data set are reduced from 19 to 6 and this reduced input data is classified with EMELM. The related flowchart is shown in Fig.1.

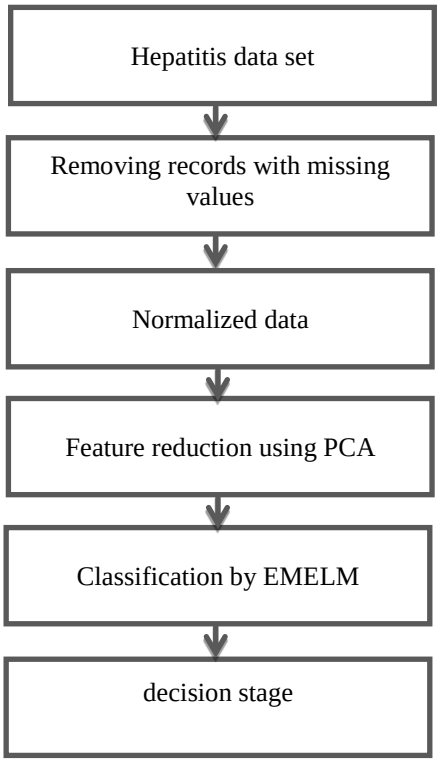

Fig. 1: Block diagram of proposed hybrid model.

3.1. Feature reduction phase

As mentioned above, records with missing values were removed and then data set was mapped in [0,1] range. The conversion function of min − max was used in this phase. Considering X as a representative of feature value, X_{min} and X_{max} as minimum and maximum values of a feature, the normalized feature values are calculated using Equation (9) [15]:

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (9)$$

In order to reduce the feature, eigenvectors are obtained for calculation of covariance matrix of data points. Then, m principal component with maximum variance are selected and conversion matrix T with m component is produced. Matrix T is applied on X and reduced data set of y is achieved.

3.2. Classification phase

In this phase, the reduced data set of previous phase is classified by SLFN with learning algorithm of EMELM. This network contains a hidden layer and an output. In this research, different activation functions are evaluated in order to optimize the network. The parameters used for the EMELM network are shown in Table 1.

Table 1: Training parameters of EMELM network

Number of layers	1
Output layer	1
Max neuron number of hidden layer	30
Expected error	0.27
Activation functions	Tangent Sigmoid, Sigmoid, Radial basis, Sine

For accuracy determination of this Hybrid approach, k-ford evaluation method was employed. This method, divides the reduced data set into k different parts. A number of k-1 parts were used as training data set and model is made accordingly. The other remained set (test data set) is used for evaluation. This procedure is repeated k times and each k part is used for evaluation only one time and specific accuracy is calculated for the model, each time [13].

3.3. Functionality evaluation methods

Accuracy criterion is the most important criterion for evaluation of algorithm. This criterion shows the portion (percentage) of accurately classified test data by classifier. The accuracy of classification is calculated using Equation (10) [11].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (10)$$

True Positive (TP) and True Negative (TN) shows the number of truly classified records. False Positive (FP) and False Negative (FN) shows inaccurate classifications. Sensitivity and specificity are two other indices for evaluation of a classification with formulations like Equations (11) and (12).

$$\text{Sensitivity} = \frac{TP}{TP+FN} \times 100\% \quad (11)$$

$$\text{Specificity} = \frac{TN}{FP+TN} \times 100\% \quad (12)$$

4. Analysis and experimental results

In present study, a hybrid system based on PCA and EMELM classifier was applied for hepatitis diagnosis. Hepatitis data set was gathered from UCI machine learning repository. This data set determines whether Patients suffering from hepatitis are alive or not. This data set contains 155 samples with 19 features. Decision-making feature includes two classes in which there are 32 (20.6%) dies and the rest of 123 samples (79.4%) are alive [11]. Approximately 48.3% of data set contains missing values. The features of this data set are shown in Table 2. The number of features reduced from 19 to 6 using PCA method and after elimination of records with missing values and normalization. Afterwards, the classification of reduced data set performed using EMELM algorithm.

The most appropriate activation function is determined by classification accuracy according to learning function and network test. In hepatitis data set, sigmoid tangent is the most appropriate activating function for hidden layer. Implementation of this algorithm was performed on intel Core i3 CPU, 2.53GHz and 2GB RAM computer with Matlab 7.8.0.

In previous studies, single and combined methods have suggested for hepatitis diagnosis. First and foremost, the classifier of this research is compared with other classifier. In order to perform EMELM algorithm on hepatitis data set without feature reduction, the nodes of hidden layer are increased one-by-one to reach the condition of minimum expected error 0.27 or maximum number of neurons.

Table 2: Details of attributes in hepatitis database

Features	Domain value	Number missing values
AGE	10, 20, 30, 40, 50, 60, 70, 80	0
SEX	Male, female	0
STEROID	Yes, No	1
ANTIVIRALS	Yes, No	0
FATIGUE	Yes, No	1
MALAISE	Yes, No	1
ANOREXIA	Yes, No	1
LIVER BIG	Yes, No	10
LIVER FIRM	Yes, No	11
SPLEEN PALPABLE	Yes, No	5
SPIDERS	Yes, No	5
ASCITES	Yes, No	5
VARICES	Yes, No	5
BILIRUBIN	0.39, 0.80, 1.20, 2.00, 3.00, 4.00	6
ALK PHOSPHATE	33, 80, 120, 160, 200, 250	29
SGOT	13, 100, 200, 300, 400, 500	4
ALBUMIN	2.1, 3.0, 3.8, 4.5, 5.0, 6.0	16
PROTIME	10, 20, 30, 40, 50, 60,70, 80, 90	67
HISTOLOGY	Yes, No	0
CLASS	Die, Alive	0

The accuracy of this method and other classification methods are shown in Table 3. As it is seen in this Table, classification process with EMELM and without feature reduction shows better results in comparison with other methods; the same as ELM.

Table 3: Accuracy of classifications applying different methods on Hepatitis data set

Author	Model	Accuracy %
Pasi[12]	MLP	79.70
Ster and dobnikar[16]	AIS	82.00
Ster and dobnikar[16]	LDA	86.40
De Bock et al.[5]	GAM	89.20
Bascil and oztekin[1]	PNN	91.25
Bascil and temurtas[2]	MLP	91.87
Kaya and Uyar [11]	ELM	93.75
Classifier [8]	EMELM	93.75

The learning method of EMELM has less complexity than ELM; as in ELM while applying the changes in the architecture of the network, the weights of previous outputs are ignored and new output weight is calculated using entirely new hidden layer output matrix. But in EMELM, as the network grows in each stage, hidden layer output matrix corresponding to number of nodes added is calculated and output weights are increasingly updated. Therefore, EMELM reduces the complexity of calculations. The performance of hybrid methods of PCA-EMELM and PCA-ELM on hepatitis data set is shown in Fig.2. As it is seen, EMELM is an effective classifier and hybrid method of PCA-EMELM functions faster than PCA-ELM. Table 4 shows the comparison between proposed method and previous combined methods. The results confirm that PCA-EMELM has the highest accuracy classifications equal to 100% for 5-fold and 10-fold.

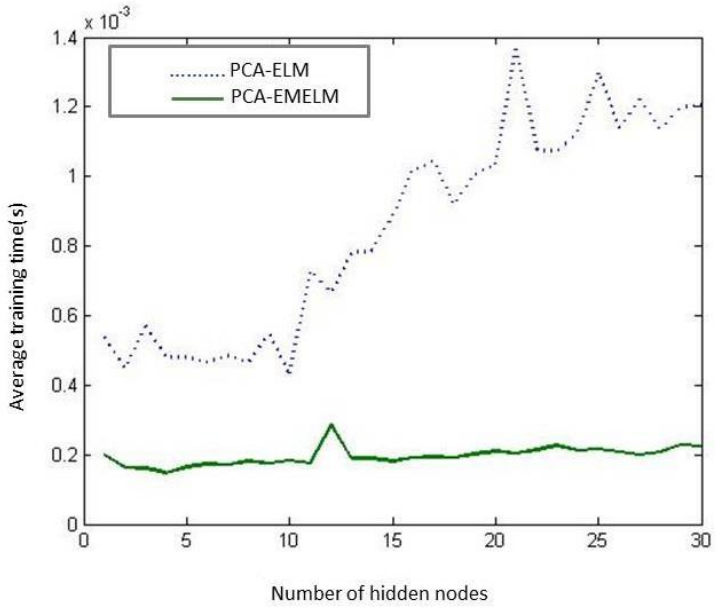

Fig. 2: Comparison between two combined methods with different learning processes for hepatitis data set

Table 4: the accuracy of proposed and previous combined methods on hepatitis data set

Author	**algorithm**	**Accuracy %**
Tan et al.[17]	GA-SVM	89.67
Polat and gunes [13]	PCA-AIRS	94.12
Dogantekin et al. [3]	PCA-LSSVM	95
Chen et al. [4]	LFDA-SVM	96.77
Proposed method(5-fold)	PCA-EMELM	100

Table 5 shows the results of two different feature reduction methods with the same classifier of EMELM. As it is seen, PCA feature reduction relative to genetic algorithm feature reduction has higher speed and accuracy. The average accuracy of hybrid algorithm PCA-EMELM with different folds is equal to 99.24%. Table 6 lists rates of sensitivity, specificity and accuracy with different folds. Therefore, PCA-EMELM has higher accuracy and takes less time compared to other methods.

Table 5: Comparison between the functionality of the method without feature reduction and combined methods

Hybrid method	Reduced number of features	Accuracy % (max)	Accuracy % (average)	Execution time(s)
GA-EMELM	8	2-fold= 87.65	89.56	8.2
		5-fold= 88.52		
		10-fold= 92.5		
PCA-EMELM	6	2-fold= 97.73	99.24	0.018
		5-fold= 100		
		10-fold= 100		

Table 6: Rates of sensitivity, specificity and accuracy for reduced set

Partition (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
2-fold	100	83.3	97.73
5-fold	100	100	100
10-fold	100	100	100

5. Conclusion

The proposed approach of PCA-EMELM is a combination of two processes: feature reduction and classification. Feature reduction decreases the complexity of the problem. Evolutionary algorithms are based on repetition; therefore feature reduction method by PCA performs faster than these algorithms. On the other hand, in this research proposed classification of EMELM with automatic determination of the number of hidden nodes and updating output weights

shows similar or better functionality as ELM, but with faster speed. The hybrid system of PCA-EMELM reduces the main features of hepatitis from 19 to 6 features and the average accuracy of this classification system is 99.24% for different folds. Therefore, The PCA-EMELM algorithm performs faster and more accurate decision-making approach for hepatitis diagnosis.

Acknowledgements

The authors are grateful to the reviewers for their useful comments which helped to improve the quality of this manuscript.

References

[1] Bascil, M.S., Oztekin, H,"A study on hepatitis disease diagnosis using probabilistic neural network", *Journal of Medical Systems*, Vol. 36, No. 3,(2012), pp.1603–1606, http://link.springer.com/article/10.1007/s10916-010-9621-x

[2] Bascil, M.S., Temurtas, F."A study on hepatitis disease diagnosis using multilayer neural network with levenberg marquardt training algorithm". *Journal of Medical Systems*, Vol. 35, No. 1, (2011), pp. 433–436, http://link.springer.com/article/10.1007/s10916-009-9378-2

[3] Çalişir, D., Dogantekin, E. "A new intelligent hepatitis diagnosis system: PCA–LSSVM". *Expert Systems with Applications*, Vol. 38, No. 8, (2011), pp.10705-10708, http://www.sciencedirect.com/science/article/pii/S0957417411000340

[4] Chen, H., Liu, D., Yang, B., Liu, J., Wang, G. "A new hybrid method based on local fisher discriminant analysis and support vector machines for hepatitis disease diagnosis". *Expert Systems with Applications*, Vol. 38, No. 9, (2011), pp.11796–11803, http://www.sciencedirect.com/science/article/pii/S095741741100474X

[5] De Bock, K. W., Coussement, K., Van den Poel, D. "Ensemble classification based on generalized additive models". *Computational Statistics & Data Analysis*, Vol. 54, No. 6, (2010), pp. 1535-1546, http://www.sciencedirect.com/science/article/pii/S0167947309004654

[6] Ding, S., Zhao, H., Zhang, Y., Xu, X., Nie, R. "Extreme learning machine: algorithm, theory and applications". *Artificial Intelligence Review*, (2013), pp.1-13, http://link.springer.com/article/10.1007/s10462-013-9405-z

[7] Dogantekin, E., Dogantekin, A., Avci, D. "Automatic hepatitis diagnosis system based on linear discriminant analysis and adaptive network based on fuzzy inference system". *Expert Systems with Applications*, Vol. 36, No. 8, (2009), pp.11282-11286, http://www.sciencedirect.com/science/article/pii/S0957417409003091

[8] Feng, G., Huang, G. B., Lin, Q., Gay, R. "Error minimized extreme learning machine with growth of hidden nodes and incremental learning". *Neural Networks, IEEE Transactions on*, Vol. 20, No. 8, (2009), pp. 1352-1357, http://dl.acm.org/citation.cfm?id=1657611

[9] Huang, C. L., Wang, C. J. "A GA-based feature selection and parameters optimization for support vector machines". *Expert Systems with applications*, Vol. 31, No. 2, (2006), pp. 231-240, http://www.sciencedirect.com/science/article/pii/S0957417405002083

[10] Jilani, T. A., Yasin, H., Yasin, M. M. "Pca-ann for classification of hepatitis c patients". *International Journal of Computer Applications*, Vol. 14, No. 7, (2011), pp. 0975–8887, http://www.ijcaonline.org/archives/volume14/number7/1899-2530

[11] Kaya, Y., & Uyar, M. "A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease". *Applied Soft Computing*, Vol. 13, No.8, (2013), pp. 3429-3438, http://www.sciencedirect.com/science/article/pii/S1568494613001130

[12] Pasi, L. "Similarity classifier using similarities based on modified probabilistic equivalence relations". *Knowledge-Based Systems, Vol.* 22, No. 1, (2009), pp. 57–62, http://www.sciencedirect.com/science/article/pii/S0950705108001317

[13] Polat, K., Güneş, S. "Prediction of hepatitis disease based on principal component analysis and artificial immune recognition system". *Applied Mathematics and computation*, Vol. 189, *No.* 2, (2007), pp. 1282-1291, http://www.sciencedirect.com/science/article/pii/S0096300306017140

[14] Rajesh, R., Prakash, J. S. "Extreme learning machines-a review and state-of-the- art". *International journal of wisdom based computing*, Vol. 1, No.1, (2011), pp. 35-49.

[15] Sartakhti, J. S., Zangooei, M. H., Mozafari, K. "Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA) ". *Computer methods and programs in biomedicine*, Vol. 108, No. 2, (2012), pp.570-579, http://www.sciencedirect.com/science/article/pii/S0169260711002082

[16] Ster, B., Dobnikar, A. "Neural networks in medical diagnosis: comparison with other methods". *Proceedings of the International Conference on Engineering Applications of Neural Networks*, Vol. 1, No. 1, (1996), pp. 427–430.

[17] Tan, K.C., E.J. Teoh, Yu, Q., Goh, K.C. "A hybrid evolutionary algorithm for attribute selection in data mining". *Expert Systems with Applications*, Vol. 36, No. 4, (2009), pp. 8616–8630, http://www.sciencedirect.com/science/article/pii/S095741740800729X