

An Automatic Diagnosis System for Hepatitis Diseases Based on Genetic Wavelet Kernel Extreme Learning Machine

Derya AVCI[†]

Abstract – Hepatitis is a major public health problem all around the world. This paper proposes an automatic disease diagnosis system for hepatitis based on Genetic Algorithm (GA) Wavelet Kernel (WK) Extreme Learning Machines (ELM). The classifier used in this paper is single layer neural network (SLNN) and it is trained by ELM learning method. The hepatitis disease datasets are obtained from UCI machine learning database. In Wavelet Kernel Extreme Learning Machine (WK-ELM) structure, there are three adjustable parameters of wavelet kernel. These parameters and the numbers of hidden neurons play a major role in the performance of ELM. Therefore, values of these parameters and numbers of hidden neurons should be tuned carefully based on the solved problem. In this study, the optimum values of these parameters and the numbers of hidden neurons of ELM were obtained by using Genetic Algorithm (GA). The performance of proposed GA-WK-ELM method is evaluated using stational methods such as classification accuracy, sensitivity and specificity analysis and ROC curves. The results of the proposed GA-WK-ELM method are compared with the results of the previous hepatitis disease studies using same database as well as different database. When previous studies are investigated, it is clearly seen that the high classification accuracies have been obtained in case of reducing the feature vector to low dimension. However, proposed GA-WK-ELM method gives satisfactory results without reducing the feature vector. The calculated highest classification accuracy of proposed GA-WK-ELM method is found as 96.642 %.

Keywords: Pattern recognition, Wavelet Kernel (WK) based Extreme Learning Machines (ELM), Genetic Algorithm (GA), Classification accuracy, Sensitivity and specificity analysis, ROC curves, Hepatitis.

1. Introduction

In medicine, the inflammation occurred in liver is called hepatitis. It can be caused by infections with viruses, bacteria, fungi, exposure to toxins such as alcohol and autoimmunity. The hepatitis damages to cells of liver and tenderness, swelling and inflammation in the liver are some symptoms of this disease. Commonly, the liver can handle significant amounts of damage, and the liver function is still effective. However, it will decline if the disease is not fully controlled at an early stage. The hepatitis can be acute or chronic and it is a common disease over the world. The different types of hepatitis are caused by different things, but they all produce inflammation of the liver [1]. Viral hepatitis refers to several common contagious diseases caused by viruses that attack the liver. The most important types of viral hepatitis are hepatitis-A, hepatitis-B, and hepatitis-C [2]. New forms of viral hepatitis such as D, E and G are also discovered. Hepatitis-A mainly infects by fecal contaminated substances taken by mouth. As a result of poor hygienic conditions, this results in epidemics based

on water or food, especially in developing countries [1]. In infection of hepatitis A, any indication may be observed. Therefore, the patients could not become aware of the hepatitis-A. Hepatitis-B virus could not infect easily such as hepatitis-A. But, it is commonly transmitted to another person by transfusion. To decrease infection risk of hepatitis-B virus, transfusion should be strictly controlled. Other infection ways of hepatitis-B virus are drugs, sexual intercourse and birth.

Hepatitis-C virus is also responsible from 80% of hepatitis taking place after transfusion and it is the most severe type. It assaults the liver and caused swelling and redness in liver [3]. Non-viral forms of hepatitis can be caused by toxic agents (drugs or chemicals), alcohol, or autoimmune processes. Another form of hepatitis is toxic hepatitis [1].

For inexperienced practitioners, one of the most important problems is to detect hepatitis virus [2]. The most common problems in the field of automatic diagnostic are selection of necessary parameter set for right diagnostics and forming of powerful algorithm which doesn't require long time to run [4]. In addition, the doctors couldn't correctly diagnosis diseases due to fatigue, inexperience and thoughtfulness. In diagnosis process, it can be useful to use an auxiliary tool which minimizes human based wrong

[†] Corresponding Author: Dept. of Electrical and Electronic Engineering, Engineering Faculty, Firat University, Elazig, Turkey. (derya2344@hotmail.com)

Received: January 20, 2014; Accepted: November 4, 2014

diagnosis and helps inexperienced practitioners. In literature, different automatic classification systems have been proposed for diagnosis of hepatitis disease.

Different methods in literature such as Multilayer Neural Network (MLNN) with backpropagation (BP) and Radial Basis Function (RBF) with OLS in [2], MLNN with Levenberg Marquardt (LM) in [5], combined diagnosis system based genetic algorithm (GA) and artificial neural network (ANN) in [6], Linear Discriminant Analysis (LDA) and Adaptive Network based on Fuzzy Inference System (ANFIS) in [7], the application of ANN for forecasting the incidence of hepatitis A by using ARIMA model in [8], (PCA-ANN) in [3], a hybrid diagnosis system using feature selection (FS) with fuzzy weighted pre-processing and artificial immune recognition in [9], PCA-LSSVM are presented for automatic diagnosis of hepatitis disease in [10].

NNs have been widely used in pattern recognition and regression. The NNs are commonly trained by back-propagation based on a gradient-based learning rule [11]. Up to now, the gradient-based learning methods have been widely applied for learning of NNs [12-13]. However, it has several shortcomings such as difficult setting of learning parameters, slow convergence, training failures due to local minima, and repetitive learning to improve performance of NNs. Also, it is clear that gradient descent-based learning methods are generally very slow [14].

In Ref [6], the GA is used for selection of feature subset for the input of ANN. Proposed method is not suitable for real-time implementation. Besides, the feature vector is randomly reduced to a lower dimension in Refs. [3, 7, 9, 10]. ANFIS structure might not a good performance if a huge amount of data exists.

Recently, a new learning algorithm called Extreme Learning Machine (ELM) which randomly selected all the hidden nodes parameters of generalized Single-hidden Layer Feedforward Networks (SLFNs) and analytically determines the output weights of SLFNs is proposed in [15-21]. Although output weights are analytically calculated, there is no rule in determination of number of hidden neurons and type of the kernel function. To obtain a good classification performance of ELM, these parameters should be determined properly. In here, the hepatitis diseases database is used [22-25].

This paper proposes an automatic hepatitis diagnosis system for hepatitis using pattern recognition based on Genetic Algorithm (GA) – Wavelet Kernel (WK) based Extreme Learning Machines (ELM). The classifier used in this paper is single layer neural network (SLNN) and it is trained by ELM learning method. In Wavelet Kernel Extreme Learning Machine (WK-ELM) structure, there are three adjustable parameters of wavelet kernel. These parameters and the numbers of hidden neurons play a major role in the performance of ELM. Therefore, values of these parameters and numbers of hidden neurons should be tuned carefully based on the solved problem. In this

study, the optimum values of these parameters and the numbers of hidden neurons of ELM were obtained by using Genetic Algorithm (GA). The hepatitis disease datasets are obtained from UCI machine learning database. The performance of proposed GA-WK-ELM method is evaluated through statical methods such as classification accuracy, sensitivity and specificity analysis and ROC curves. In here, the numbers of hidden neurons of ELM and parameters of wavelet kernel function are optimized by GA. In GA structure, a individual composes of total of 20 bits. These are:

- The first four bits (1st, 2nd, 3rd and 4th bits) of each of these individuals represent the w parameter values (between 1 and 16) of the wavelet kernel functions.
- The second four bits (5th, 6th, 7th and 8th bits) of each of these individuals represent the x parameter values (between 1 and 16) of the wavelet kernel functions,
- The third four bits (9th, 10th, 11th and 12th bits) of each of these individuals represent the y parameter values (between 1 and 16) of the wavelet kernel functions,
- Rest of 20 bits represents number of hidden neurons (between 5 and 132).

The 40 number of these individuals are randomly chosen for initial population. Thus, it is purposed to obtain the best possible performance from ELM classifier. The training and testing dataset for proposed method is obtained from UCI dataset. These dataset compose of 155 data. The randomly selected 100 of 155 data are used for training of classifier whereas remaining data is used for testing of classifier. For different kernel function and number of hidden neurons, the results of proposed method are given. Further, a comparison is performed with previous studies to show validity of proposed method. From results, the proposed method is quite powerful tool for automatic diagnosis of hepatitis and may work in real-time systems.

In ref. [27], Wang et al. claim that wavelet kernel functions can widen the kernel function selection range of Kernel ELM (KELM) and support its further development. The important reason of using wavelet kernel function is having adjustable w , x , and y parameters. Thus, the wavelet kernel function gains adaptive features for best classification and test results in different data types [28]. In Wavelet Kernel Extreme Learning Machine (WK-ELM) structure, there are three adjustable parameters of wavelet kernel. These parameters and the numbers of hidden neurons play a major role in the performance of ELM. Therefore, values of these parameters and numbers of hidden neurons should be tuned carefully based on the solved problem. In this study, the optimum values of these w , y , x parameters and the numbers of hidden neurons of ELM were obtained by using Genetic Algorithm (GA). According to the results of this study, the wavelet kernel ELM has a better performance than classical ELM for automatic diagnosis system of hepatitis disease. At the

same time, the wavelet kernel functions perform very well when the algorithm is compared with the classical kernel functions in ELM model. The comparing results are given in Table 4.

The paper is organized as follows. Section 2 presents pattern diagnosis concept. In section 3, Classic ELM and in section 4, Wavelet Kernel (WK) based Extreme Learning Machines (ELM) are briefly presented respectively. In section 5, the GA is explained. The description of hepatitis dataset used is given in Section 6. The application of GA-WK - ELM for hepatitis diagnosis is presented in Section 6. Finally, Section 7 provides conclusions of this study.

2. Pattern Recognition Concept

Pattern recognition could simply be defined as the process of categorizing the input signal. This process can be divided into two main groups as classification and regression. In the classification process, input signal is separated into two or more classes while pattern recognition system tries to find the appropriate mapping between the input and the output variables in the regression process. Generally, regression problems are harder to solve than classification problems due to several reasons [22]. Fig. 1 shows conventional pattern recognition concept for classification. As shown in Fig. 1, the pattern recognition concept composes of two stages. They include feature extraction and classification stages. The feature extraction stage is the most important part of pattern diagnosis. The appropriate features are extracted by a feature extractor. If the appropriate features are not selected, the classification performance will be poor even though using the best classifier. Therefore, feature extractor should reduce the dimension of pattern vector to a lower. The reduced feature vector should also cover useful information of the original vector. In last stage, reduced feature vector is given to inputs of classifier for the classification [20, 21].

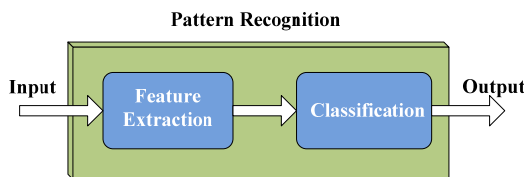


Fig. 1. The block diagram of the pattern recognition concept

3. Classic Extreme Learning Machine

The NNs have been widely used in pattern recognition and regression problems. Commonly, the learning of NN has been performed by using gradient-based learning algorithms. However, such methods has several drawbacks

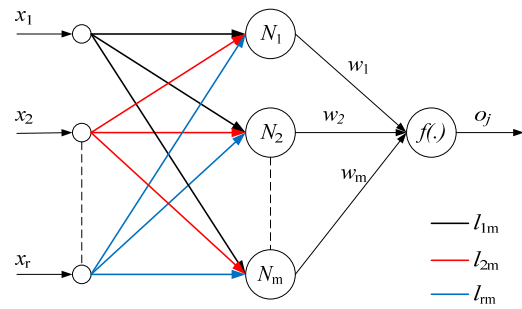


Fig. 2. Structure of SLFN

such as difficult setting of learning parameters, slow convergence, slow learning and training failures [20, 21].

To deal with the drawbacks of gradient-based learning methods, ELM was proposed by Huang, et al. [20]. In the ELM, the output weights of a single-hidden layer feedforward network (SLFN) are analytically computed by using the Moore-Penrose (MP) generalized inverse instead of iterative learning scheme. Fig. 2 shows structure of a SLFN using ELM. In this figure, l_{1m} , l_{2m} and l_{rm} , are weights vector connecting the i th hidden neuron and the input neurons, and w is the weight vector connecting the i th hidden neuron and output neuron and $f(.)$ is kernel function.

The most important properties of ELM are given as follows:

- The learning speed of ELM is extremely fast. Therefore, SLFNs can be trained by ELM much faster than classical learning methods.
- The ELM tends to reach both the smallest training error and the smallest norm of weights. Thus, the ELM tends to have good performance for neural networks.
- The ELM learning algorithm can be used to train SLFNs with non-differentiable kernel functions.
- The ELM tends to reach the solutions straightforward without such trivial issues [20].

The outputs of a SLFN with m neurons and f kernel function are written as follows:

$$o_j = \sum_{i=1}^m \beta_i f(l_i x_r + b_i) \tag{1}$$

That standart SLFN with m hidden neurons and f kernel function can approximate these m samples with zero error means that $\sum_{j=1}^m (o_j - t_j) = 0$, i.e., there exist w , l and b such that [11], [20];

$$t_j = \sum_{i=1}^m w_i f(l_i x_r + b_i), \quad j = 1, 2, \dots, m \tag{2}$$

The m set of equations can be written as follows:

$$H \times w = T \tag{3}$$

where; H is the hidden layer output matrix of the SLFN.

$$H(l_1, \dots, l_m, b_1, \dots, b_m, x_1, \dots, x_m) = \begin{bmatrix} f(l_1 x_1 + b_1) & \dots & f(l_m x_1 + b_m) \\ \vdots & \dots & \vdots \\ f(l_1 x_m + b_1) & \dots & f(l_m x_m + b_m) \end{bmatrix} \tag{4}$$

ELM method for SLFNs called ELM can be briefly explained as follows [20]:

Given a training set, kernel function $f(\cdot)$ and the number of hidden nodes m :

- Randomly assign input weight w and bias b , ($i=1, 2, \dots, m$).
- Find the hidden layer output matrix H .
- Find the output weight w .

4. Wavelet Kernel Based Extreme Learning Machines

The ELM learning algorithm has fast learning speed and good generalization performance. In recently years, number of researchers has increased [4, 7]. The initial parameters of hidden layer need not be tuned in ELM algorithm. In this algorithm, all nonlinear piecewise continuous functions are used as the hidden neurons. Because of this, for M optional different sample data set $\{(r_j, m_j) \mid r_j \in Q^l, m_j \in Q^k, j=1, \dots, M\}$, the output function in ELM by using K hidden neurons is

$$u_K(r) = \sum_{j=1}^K S_j v_j(r) = v(r)S, \tag{5}$$

In here, $v = [v_1(r), v_2(r), \dots, v_K(r)]$ is the output vector of the hidden layer with respect to the input r : $S = [S_1, S_2, \dots, S_K]$ is the vector of the output weights between the hidden layer of K neurons and the output neuron. v vector converts the data from input space to the ELM feature space [7]. In ELM structure, the training error and the output weights should be simultaneously minimized for decreasing the training error. Thus, generalization performance of neural networks improves.

$$\text{Minimize: } \|AS - C\|, \|S\| \tag{6}$$

There, Eq. (6) can be solved by using Eq.(7).

$$S = A^T \left(\frac{1}{E} + AA^T \right)^{-1} C, \tag{7}$$

In here, E is the regulation coefficient, A is the hidden layer

output matrix and C is the expected output matrix of samples. Thus, the output function of the ELM learning algorithm can be found as below:

$$u(r) = v(r)A^T \left(\frac{1}{E} + AA^T \right)^{-1} C \tag{8}$$

If the feature vector $v(r)$ is unknown, the kernel matrix of ELM based on Mercer's conditions can be given as below:

$$D = AA^T : k_{jz} = v(r_j)v(r_z) = b(r_j, r_z), \tag{9}$$

thus, the output function $u(r)$ of the wavelet kernel based extreme learning machine (WK-ELM) can be shown as:

$$u(r) = [b(r, r_1), \dots, b(r, r_M)] \left(\frac{1}{E} + D \right)^{-1} C, \tag{10}$$

In here, $D = AA^T$ and $b(r, g)$ is the kernel function of Extreme Learning Machine. In ELM literature, some kernel functions, which are linear kernel, polynomial kernel, Gaussian kernel, and exponential kernel, appropriate to the Mercer condition. The readers can be found more details in [7, 9, 27, 28]. In this paper, wavelet kernel function, which is given on Eq.(11), is used for simulation and performance analysis for WK-ELM.

$$b(r, g) = \cos \left(w \frac{\|r - g\|}{x} \right) \exp \left(- \frac{\|r - g\|^2}{y} \right), \tag{11}$$

The training and testing performance of the wavelet kernel function given in Eq. (11) is better than the performances of linear kernel, polynomial kernel, Gaussian kernel, and exponential classical kernel functions respectively. In wavelet kernel, the adjustable parameters w , x and y play a major role in the performance of neural networks. Therefore, values of these parameters should be tuned carefully based on the solved problem. Moreover, in WK-ELM algorithms, the hidden layer feature mapping need not be known and the number of hidden neurons need not be chosen. However, the WK-ELM learning algorithm obtains similar or better generalization performance. Most important WK-ELM is more stable compared to traditional ELM and it is faster than support vector machine (SVM) [7, 9].

5. Genetic Algorithm

Genetic Algorithms is good at taking large, potentially huge search spaces and navigating them, looking for optimal combinations of things, solutions you might not otherwise find in a lifetime [22].

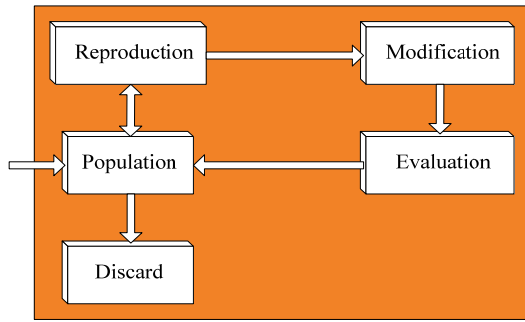


Fig. 3. Components of a GA.

The GA provides efficient, effective techniques for optimization and machine learning applications and they have been widely used in business, scientific and engineering fields. Fig. 3 shows components of a GA. As shown in this figure, components of a GA are encoding technique, initialization procedure, evaluation function, selection of parents, genetic operators, parameter settings. The GA begins with a set of solutions which are represented by individuals (chromosome). The individuals could be bit strings (0101 ... 1100), real numbers (43.2 - 33.1 ... 0.0 89.2), permutations of element (E11 E3 E7 ... E1 E15), lists of rules (R1 R2 R3 ... R22 R23), program elements (genetic programming) or any data structure. In reproduction, individuals of population are randomly selected with selection chances biased in relation to population evaluations. The evaluation step decodes a chromosome and assigns it a fitness measure. Also, the evaluation is the only link between a classical GA and the problem it is solving [23]. In genetic algorithm, the iterative process is repeated as long as during the new population is better than the old one. The higher the fitness value of an individual is, the more likely this individual, is reproduced for the next population. The iterative process is finished when some conditions (for example, number of individuals in the population, etc.) are satisfied [24].

The basic genetic algorithm is summarized as follows:

- Step 1:** A random population of n individuals is generated which is a suitable solution for the problem. The numbers of n are 40 in this study.
- Step 2:** The fitness $f(x)$ of each individual x is evaluated in the population [28]. In these experimental studies, each of individuals in population is randomly formed.
- Step 3:** It is chosen two parental individuals from among the individuals has higher the fitness value in population. Then, the cross over operator is performed to these parental individuals. The purpose of cross over operator is obtaining the different individuals, which have higher fitness values.
- Step 4:** It is cross overed the parents with a crossover probability to form new individuals. If crossover is not performed, individual will be the exact copy of

Table 1. The attribute information of dataset

	Attribute information	Values
1	Age	10-80 by step 10 years
2	Sex	Male and female
3	Steroid	No, yes
4	Antiviral	No, yes
5	Fatigue	No, yes
6	Malaise	No, yes
7	Anorexia	No, yes
8	Liver big	No, yes
9	Liver firm	No, yes
10	Spleen palpable	No, yes
11	Spiders	No, yes
12	Ascides	No, yes
13	Varices	No, yes
14	Bilirubin	0.39, 0.80, 1.20, 2.00, 3.00, 4.00
15	Alk phosphate	33, 80, 120, 160, 200, 250
16	Sgot	13, 100, 200, 300, 400, 500,
17	Albumin	2.1, 3.0, 3.8, 4.5, 5.0, 6.0
18	Protime	10, 20, 30, 40, 50, 60, 70, 80, 90
19	Histology	No, yes
20	Class	Die, alive

parents.

- Step 5:** It is mutated each new individual with a mutation probability at each locus, which is the position in the individual.
- Step 6:** The new individuals are reconstructed from the new population.
- Step 7:** The genetic algorithm is stopped, if the end conditions is satisfied, and return the best solution in current population.
- Step 8:** It is gone to step 2 and used the newly generated population for a further run of the algorithm.

6. Data Description

The hepatitis data is obtained from UCI machine learning database. It contains 20 attributes including the class attribute [26]. The attribute information of dataset is given in Table 1.

The output shows whether patients with hepatitis are die or alive. Hepatitis dataset contains 155 data belonging to two different target classes. There are 20 features, 13 binary and 6 attributes with 6-8 discrete values. Also, the class distribution contains 32 cases for die and 123 cases for alive.

7. Application of GA-WK-ELM for Diagnosis of Hepatitis

This GA-WK-ELM method composes of three layer. In first layer of GA-WK-ELM, the hepatitis data is obtained from UCI machine learning database mentioned in Section 5. In second layer of GA-WK-ELM, the numbers of hidden

neurons of ELM and parameters of wavelet kernel function are optimized by GA.

In this study, the used parameters, which are coding type, mutation rate, mutation type, crossover rate, crossover type, population size, and chromosome size for GA design are given in Table 2. These parameter values are the best values found after much trial.

Table 2. The used parameters for GA design

The parameter type of GA	
Coding type	Binary encoding
Mutation rate	5 %
Mutation type	Bit inversion
Crossover rate	80 %
Crossover type	Two point crossover
Population size	40 chromosomes
Chromosome size	20 bits

The realized crossover sample in this study by using two point crossover can be given as below:

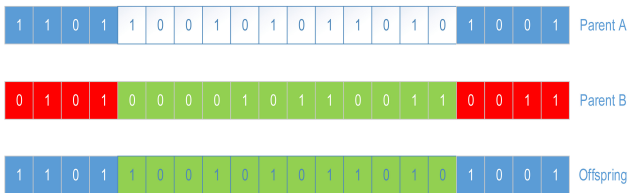


Fig. 4. The realized crossover sample in this study

The realized mutation sample in this study by using bit inversion can be given as below:

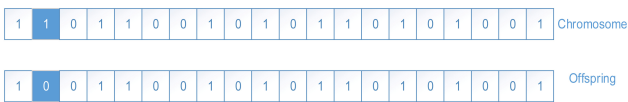


Fig. 5. The realized mutation sample in this study

In GA structure, a individual composes of total of 20 bits. These are:

- The first four bits (1st, 2nd, 3rd and 4th bits) of each of these individuals represent the w parameter values (between 1 and 16) of the wavelet kernel functions.
- The second four bits (5th, 6th, 7th and 8th bits) of each of these individuals represent the x parameter values (between 1 and 16) of the wavelet kernel functions,
- The third four bits (9th, 10th, 11th and 12th bits) of each of these individuals represent the y parameter values (between 1 and 16) of the wavelet kernel functions,
- Rest of 20 bits represents number of hidden neurons (between 5 and 132).

The 40 number of these individuals are randomly chosen for initial population. Thus, it is purposed to obtain the best possible performance from ELM classifier. The training

and testing dataset for proposed GA-WK-ELM method is obtained from UCI dataset. These dataset compose of 155 data. The randomly selected 100 of 155 data are used for training of classifier whereas remaining data is used for testing of classifier. For the optimum numbers of hidden neurons of ELM and parameters values of wavelet kernel function, the results of proposed GA-WK-ELM method are given. Further, a comparison is performed with previous studies to show validity of proposed GA-WK-ELM method. From results, the proposed GA-WK-ELM method is quite powerful tool for automatic diagnosis of hepatitis and may work in real-time systems.

The block diagram of the proposed GA-WK-ELM method is given in Fig. 6. In these applications, a 3-fold cross-validation schema was applied where the two-fifth data were used for training the proposed GA-WK-ELM method and the remaining other data were used as the test data set. This strategy was applied for three times and the average values were calculated for determining the performance of proposed GA-WK-ELM method.

In genetic algorithm, the mean training error was used as fitness function. This mean training error was obtained from result of training of ELM for each of individuals by using parameters represented by these individuals. In the test process of GA-WK-ELM, the numbers of hidden neurons of ELM and parameters of wavelet kernel function are optimized by GA. As said previously, the dataset has 19 relevant features except for class attribute and includes a total of 155 cases. Thus, it is a matrix with dimension of 155 x 19. The parameters of ELM are optimized by GA.

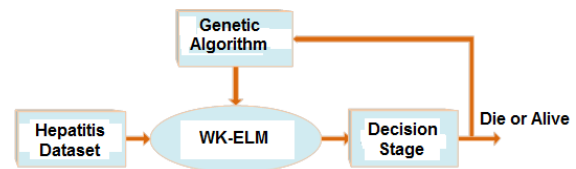


Fig. 6. The block diagram of GA-WK-ELM based optimal diagnosis system

Table 3. Coding for kernel functions

Each of w, x, y wavelet kernel parameter values	Coding
1	0 0 0 0
2	0 0 0 1
3	0 0 1 0
4	0 0 1 1
5	0 1 0 0
6	0 1 0 1
7	0 1 1 0
8	0 1 1 1
9	1 0 0 0
10	1 0 0 1
11	1 0 1 0
12	1 0 1 1
13	1 1 0 0
14	1 1 0 1
15	1 1 1 0
16	1 1 1 1

Table 4. Coding for number of hidden neurons

The number hidden neurons	Coding
5	0 0 0 0 0 0 0 0
6	0 0 0 0 0 0 0 1
7	0 0 0 0 0 0 1 0
8	0 0 0 0 0 0 1 1
⋮	⋮
⋮	⋮
132	1 1 1 1 1 1 1 1

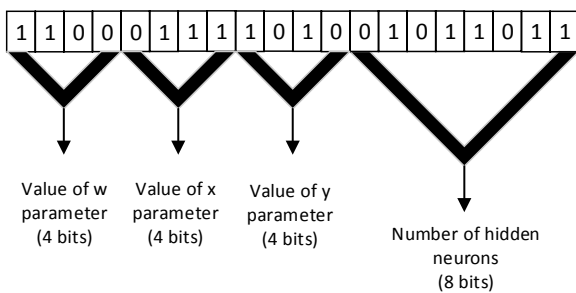


Fig. 7. An example for individuals of the population

Firstly, 40 random individuals which are formed from a total of 20 bits are chosen as the initial population. Coding for parameters of wavelet kernel functions and the number of hidden neurons are given in Table 3 and Table 4, respectively.

An example for individuals of the population is shown as follows. The first four bits (1st, 2nd, 3rd and 4th bits) of each of this individual represent the w parameter values (between 1 and 16) of the wavelet kernel functions. The second four bits (5th, 6th, 7th and 8th bits) of each of this individual represent the x parameter values (between 1 and 16) of the wavelet kernel functions. The third four bits (9th, 10th, 11th and 12th bits) of each of this individual represent the y parameter values (between 1 and 16) of the wavelet kernel functions, the rest of 20 bits of this individual represents number of hidden neurons (between 5 and 132).

Performance of the proposed GA-WK-ELM method is calculated by three evaluation methods as classification accuracy, sensitivity and specificity analysis and Receiver Operating Characteristic (ROC) curves. The classification accuracies for the datasets are found using the following equation.

$$Correct\ accuracy\ (C) = \frac{\sum_{k=1}^{|C|} assess(c_k)}{|C|}, \quad c_k \in C \quad (12)$$

$$assess(C) = \begin{cases} 1 & \text{if } classify(c) = c.m \\ 0 & \text{otherwise} \end{cases}$$

$$assess(c) = \begin{cases} 1, & \text{if } classify(c) = c.d \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

where, C is the set of hepatitis data to be classified (the test set), c.m is the class of item c and classify(c) returns the classification of the by ELM. The sensitivity and specificity analysis are obtained as follows:

$$sensitivity(\%) = \frac{TP}{TP + FN} \quad (14)$$

$$specificity(\%) = \frac{TN}{FN + TN} \quad (15)$$

Where; TP, TN, FP and FN are true positives, true negatives, false positives and false negatives, respectively.

8. Obtained Results

In this study, an automatic intelligent system for diagnosis of hepatitis disease is presented. Performance of the proposed GA-WK-ELM method is also evaluated by classification accuracy, sensitivity and specificity analysis and ROC curve.

The correct hepatitis diagnosis performances of classic ELM classifiers, which have different types of kernel function and the number of hidden neurons, are compared with the performance of the proposed GA-WK-ELM method, which has optimum values of wavelet kernel w, x, y parameters and numbers of hidden neuron in these experimental studies. The results of these performance comparisons are given in Table 4. As shown in this table, the best classification accuracy of the proposed GA-WK-ELM method is found as 97.5623% in case of values of w, x, y wavelet kernel function parameters, which are 15, 3 and 10, and hidden neuron with 86.

To show validity of proposed GA-WK-ELM method, compared results with previous studies using same dataset are also given in Table 5. As shown in this table, the highest correct hepatitis diagnosis rate is obtained as 97.56% by using proposed GA-WK-ELM method. Because, the optimum values of these parameters and the numbers of hidden neurons of ELM were obtained by using Genetic Algorithm (GA). Also, it is clearly shown that the obtained results from GRNN and RBF have been deteriorated in case of without GA. Similarly, other methods presented in Ref [3, 6, 7, 9] and [10] use reduced feature vector for input of classifier. In this study, after finding optimum parameters, GA does not need to use and then the WK-ELM can be directly used. The correct hepatitis diagnosis performance comparing of GA-WK-ELM method with Kernel ELM classifiers, which have different types of kernel function and the number of hidden neurons is given in Table 5.

The obtained hepatitis diagnosis accuracies by statistical evaluation criterias are given in Table 7.

A comparison with previous studies including different database is also given in Table 8. As shown in this table,

Table 5. The correct hepatitis diagnosis performance comparing of GA-WK-ELM method with Kernel ELM classifiers, which have different types of kernel function and the number of hidden neurons.

Used method	Type of the kernel function	Value of w wavelet kernel parameter	Value of x wavelet kernel parameter	Value of y wavelet kernel parameter	The number of hidden neurons	Accuracy (%)
Kernel ELM	Sigmoid	-	-	-	87	89.09
GA-WK-ELM	Wavelet	8	5	3	23	95.25
GA-WK-ELM	Wavelet	11	4	5	45	95.45
GA-WK-ELM	Wavelet	7	10	12	92	96.45
GA-WK-ELM	Wavelet	15	3	10	86	97.56
Kernel ELM	Poly	-	-	-	57	89.09
Kernel ELM	Tangent sigmoid	-	-	-	259	92.72
Kernel ELM	Sigmoid	-	-	-	143	90.90
Kernel ELM	Poly	-	-	-	191	89.09
Kernel ELM	Tangent sigmoid	-	-	-	18	92.72
Kernel ELM	Sigmoid	-	-	-	81	89.09
Kernel ELM	Tangent sigmoid	-	-	-	204	92.72

Table 6. The comparison results of proposed GA-WK-ELM method and previous studies using same dataset.

References	Methodology	Accuracy (%)
Ref [2]	ANN	81.37
	RBF	85.00
Ref [5]	ANN+LM	91.87
Ref [6]	GA+GRNN	93.55
	GA-RBF	96.77
Ref [7]	LDA+ANFIS	94.16
Ref [3]	PCA+ANN	89.60
Ref [9]	FS+AIRS	92.59
Ref [10]	PCA+LSSVM	92.59
This study	GA-WK-ELM	97.56

Table 7. The obtained hepatitis diagnosis accuracy by statistical methods.

Method	Classification Accuracy (%)
Sensitivity Analysis	96.89
Specificity Analysis	98.23
Average	97.56

the high classification accuracy has been obtained by proposed GA-WK-ELM method [22-25].

In this study, ROC curves and AUC values are calculated by using TP, TN, FP and FN are true positives, true negatives, false positives and false negatives, respectively [22-27]. A ROC curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate against the false positive rate at various threshold settings. (The true-positive rate is also known as sensitivity in biomedical informatics or recall in machine learning. The false-positive rate is also known as the fall-out and can be calculated as 1 - specificity). The ROC curve is thus the sensitivity as a function of the fall-out.

In general, if the probability distributions for both detection and false alarm are known, the ROC curve can be generated by plotting the cumulative distribution function (area under the probability distribution from to) of the

Table 8. Comparing the correct diagnosis rate of GA-WK-ELM based expert system and previous methods, which used different hepatitis database.

Method	Accuracy %	Reference
Logdisc	77.7	Statlog
IncNet	77.6	Norbert Jankowski
DIPOL92	77.6	Statlog
Linear Discr. Anal.	77.5-77.2	Statlog; Ster & Dobnikar
SMART	76.8	Statlog
GTO DT (5xCV)	76.8	Bennet and Blue
ASI	76.6	Ster & Dobnikar
MLP+BP	76.4	Ster & Dobnikar
MLP+BP	75.8 ± 6.2	Zarndt
LVQ	75.8	Ster & Dobnikar
RBF	75.7	Statlog
NB	75.5-73.8	Ster & Dobnikar;
kNN, k=22, Manh	75.5	Karol Grudziński
MML	75.5 ± 6.3	Zarndt
SNB	75.4	Ster & Dobnikar
BP	75.2	Statlog
kNN, k=18, Euclid, raw	74.8 ± 4.8	WD-GM
CART DT	74.7 ± 5.4	Zarndt
CART DT	74.5	Statlog
DB-CART	74.4	Shang & Breiman
ASR	74.3	Ster & Dobnikar
CART	72.8	Ster & Dobnikar
Kohonen	72.7	Statlog
kNN	71.9	Ster & Dobnikar
QDA	59.5	Ster, Dobnikar
BFGS quasi Newton	77.08	(Yildirim et al. 2003)
Levenberg-Marquardt	77.08	(Yildirim et al. 2003)
RBF	68.23	(Yildirim et al. 2003)
GRNN	80.21	(Yildirim et al. 2003)
GDA- LS-SVM	79.16	(K. Polat et al., 2008)
Gradient descent	77.60	(Yildirim et al. 2003)
SVM (5xCV)	77.6	Bennet and Blue
C4.5	76.0 ± 0.9	Friedman, 5xCV
Semi-Naive Bayes	76.0 ± 0.8	Friedman, 5xCV
Naive Bayes	74.5 ± 0.9	Friedman, 5xCV
GA-WK-ELM	92.7273	(used in this study)

detection probability in the y-axis versus the cumulative distribution function of the false-alarm probability in the x-

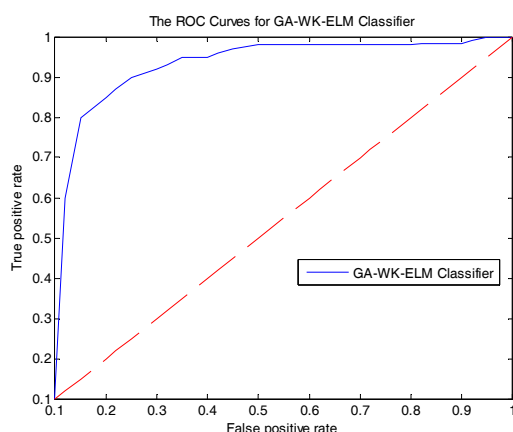


Fig. 8. The ROC curve of proposed GA-WK-ELM method for intelligent hepatitis diagnosis.

axis.

ROC analysis provides tools to select optimal models and to discard suboptimal ones independently from the class distribution or the cost context. ROC analysis is related in a direct and natural way to cost/benefit analysis of diagnostic decision making. In here, the ROC curve of GA-WK-ELM is given by using obtained best TP, TN, FP and FN values in Fig. 6. The calculated AUC value of ROC curves by using GA-WK-ELM classifier can be given as 0,9624.

9. Discussion and Conclusion

This paper proposes an GA-WK-ELM based optimal intelligent system for diagnosis of hepatitis. The proposed GA-WK-ELM hepatitis diagnosis system has advantages such as directly using of feature vector, fast training and testing time as well as generalization capability over conventional neural networks with back propagation. The proposed GA-WK-ELM method is carried out in two stages as classification and optimization of classifier's parameters. The feature vector from hepatitis dataset is used as input to WK-ELM classifiers. In Wavelet Kernel Extreme Learning Machine (WK-ELM) structure, there are three adjustable parameters of wavelet kernel. These parameters and the numbers of hidden neurons play a major role in the performance of ELM. Therefore, values of these parameters and numbers of hidden neurons should be tuned carefully based on the solved problem. In this study, the optimum values of these parameters and the numbers of hidden neurons of ELM were obtained by using Genetic Algorithm (GA). The output of WK-ELM decides make a classification. To obtain the best possible classification performance, the optimum values of parameters of classifier are determined by GA. The feasibility of the developed technique has been tested by data set of 192 test cases. The proposed GA-WK-ELM technique shows effective classification performance when

compared with previous studies depending on directly using of feature vector and training time as shown in Tables V-VIII and Fig. 8.

References

- [1] Esin Dogantekin, Akif Dogantekin, Derya Avci, Automatic hepatitis diagnosis system based on Linear Discriminant Analysis and Adaptive Network based on Fuzzy Inference System, *Expert Systems with Applications* 36 (2009) 11282–11286.
- [2] Esin Dogantekin, Akif Dogantekin, and Derya Avci. "Automatic hepatitis diagnosis system based on linear discriminant analysis and adaptive network based on fuzzy inference system", *Expert Systems with Applications* 36(8), pp. 11282-11286, 2009.
- [3] Tahseen A. Jilani, Huda Yasin, Madiha Mohammad Yasin, "PCA-ANN for Classification of Hepatitis-C Patients", *International Journal of Computer Applications* (0975-8887), Volume 14-No.7, February 2011.
- [4] Ultrasound and Magnetic Resonance Medical Images Textures Analysis Using Neural Networks, [http:// www.intermag.kiev.ua/projects/neuraln.html](http://www.intermag.kiev.ua/projects/neuraln.html), (last accessed: 20 January 2012).
- [5] M. Serdar Bascil, Feyzullah Temurtas, "A Study on Hepatitis Disease Diagnosis Using Multilayer Neural Network with Levenberg Marquardt Training Algorithm", *J Med Syst* (2011) 35:433-436.
- [6] Adeli, Mahdieh, and Hassan Zarabadipour. "Automatic disease diagnosis systems using pattern recognition based genetic algorithm and neural networks", *Int. J. Phys. Sci.* v6 i25 (2011): 6076-6081.
- [7] Bin Li, Xuewen Rong and Yibin Li, "An Improved Kernel Based Extreme Learning Machine for Robot Execution Failures", Hindawi Publishing Corporation The Scientific World Journal, Volume 2014, Article ID 906546, pp. 7, <http://dx.doi.org/10.1155/2014/906546>.
- [8] Peng Guan, De-Sheng Huang, Bao-Sen Zhou, Forecasting model for the incidence of hepatitis A based on artificial neural network, *China World Journal of Gastroenterol*; 10(24), 2004, pp. 3579-3582.
- [9] Kemal Polat, Salih Gunes, Hepatitis disease diagnosis using a new hybrid system based on feature selection (FS) and artificial immune recognition system with fuzzy resource allocation, *Digital Signal Processing* 16 (2006), pp. 889-901.
- [10] Duygu Çalisir, Esin Dogantekin, A new intelligent hepatitis diagnosis system: PCA-LSSVM, *Expert Systems with Applications*, Volume 38, Issue 8, pages: 10705-10708, 2011.
- [11] Al-Shayea, Qeethara Kadhim. "Artificial neural networks in medical diagnosis", *International Journal of Computer Science Issues* 8.2 (2011): 150-154.
- [12] Amato, Filippo, et al. "Artificial neural networks in

- medical diagnosis”, *Journal of applied biomedicine* 11.2 (2013): 47-58.
- [13] Malmgren, Helge, Magnus Borga, and Lars Niklasson, eds. *Artificial Neural Networks in Medicine and Biology: Proceedings of the ANNIMAB-1 Conference, Göteborg, Sweden, 13-16 May 2000*. Springer Science & Business Media, 2012.
- [14] Cho, Jae-Hoon, Dae-Jong Lee, and Myung-Geun Chun, “Parameter Optimization of Extreme Learning Machine Using Bacterial Foraging Algorithm”, *EESRI*, pp: 742-747, 2007.
- [15] Guang-Bin, Huang, et al. “Extreme learning machine for regression and multiclass classification.” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 42.2 (2012): 513-529.
- [16] Guang-Bin Huang, et al., “Extreme Learning Machines: A Survey,” *International Journal of Machine Learning and Cybernetics*, 2 (2011) 107-1221.
- [17] Guang-Bin Huang, et al., “Optimization Method Based Extreme Learning Machine for Classification”, *Neurocomputing*, 74 (2010) 155-163.
- [18] Huang, Guang-Bin, Lei Chen, and Chee-Kheong Siew, “Universal Approximation Using Incremental Constructive Feedforward Networks with Random Hidden Nodes”, *IEEE Transactions on Neural Networks*, 17 (2006) 879-892.
- [19] Guang-Bin Huang, et al., “Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks,” 2004 International Joint Conference on Neural Networks (IJCNN'2004), (Budapest, Hungary), July 25-29, 2004.
- [20] Guang-Bin Huang, et al., “Extreme Learning Machine: Theory and Applications”, *Neurocomputing*, vol. 70, pp. 489-501, 2006.
- [21] Liang, Nan-Ying, et al., “A Fast and Accurate On-line Sequential Learning Algorithm for Feedforward Networks”, *IEEE Transactions on Neural Networks*, 17 (2006) 1411-1423.
- [22] Guo, Ling, et al., “Automatic feature extraction using genetic programming: An application to epileptic EEG classification”, *Expert Systems with Applications*, Volume 38, Issue 8, August 2011, Pages 10425-10436.
- [23] Birendra Biswal, et al., “Non-stationary power signal processing for pattern recognition using HS-transform”, *Applied Soft Computing*, Volume 9, Issue 1, January 2009, Pages 107-117.
- [24] Jiang, Lingli, et al., “Using bispectral distribution as a feature for rotating machinery fault diagnosis”, *Measurement*, Volume 44, Issue 7, August 2011, Pages 1284-1292.
- [25] UCI Machine Learning Repository. Available from: <http://www.ics.uci.edu/~mllearn/MLRepository.html> (last accessed: 20 September 2014).
- [26] Jiu-Yao Wang, “Wavelet kernel extreme learning classifier”, *Microelectron Comput* 10:73-76-80, 2013.
- [27] Shifei Ding, et al., “A wavelet extreme learning machine”, *Neural Comput & Applic*, DOI 10.1007/s00521-015-1918-8, Published online: 09, May, 2015.



Derya Avcı She received B.S and Master degree in Computer Science from Firat University. She received Ph.D degree in Electrical and Electronic Engineering from Firat University. Her research interests are artificial intelligence, image processing and signal processing.