Received: February 10, 2020

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Modeling for Predicting the Severity of Hepatitis Based on Artificial Neural Networks

Sara H. Muhi[1] **Hadeel N. Abdullah[1]*** **Bassam H. Abd[1]**

[1]*Electrical Engineering Department, University of Technology, Iraq.*
* Corresponding author's Email: *30002@uotechnology.edu.iq*

Abstract: This work aims to design an intelligent model capable of diagnosing and predicting the severity of the hepatitis of illness that assists physicians to make an accurate decision. The main contribution is achieved by adopting a new multiclass classifier approach based on a collected real database with new proposed features that reflect the precise situation of the disease. In this work, three artificial neural networks (ANNs) methods, namely, Back Propagation neural network (BPNN), Radial Basis Function Network (RBFN) and, K-nearest neighbor (KNN), used to forecast the level of hepatitis intensity. Real data Collected from the Gastroenterology and Liver Education Hospital of the City of Medicine in Baghdad used as modeling and forecasting samples, respectively, to compare the results of forecasting. The results show that the prediction result by the KNN network will be better than the two other methods in time record to reach an automatic diagnosis with an error rate of less than 1%. Diagnosis accuracy was 99.33% for 2-class and 88% for 5-class, which considered excellent accuracy.

Keywords: Prediction, Severity of hepatitis, Backpropagation, Radial basis function, K-nearest neighbour.

1. Introduction

Hepatitis is a deformation, inflammation, and damage of liver cells that affects all ages without exception is one of the chronic liver diseases that often may lead to the death of the patient. People with hepatitis need special care for a period to regain their health and reduce the risk of death by seeing a doctor and repeating tests from time to time. It can be more accessible and faster if it uses automated learning technology [1]. Therefore, the diagnosis of hepatitis has become one of the most critical challenges faced by researchers in the design of specialized medical diagnostic systems with high precision [2]. Regardless of the cause of hepatitis, which mentioned earlier, all of them lead after a period to damage the tissue of the liver and the death of many of the cells and the emergence of some nodes or scars in cells liver.

This advanced stage of the disease called cirrhosis, which is a complication of hepatitis, an awkward stage of stages the disease often kills the patient. In which the liver stops functioning. Whereas the early diagnosis of this disease can change the life of the patient because it is a silent disease is often without symptoms, but it develops and worsens until it reaches a stage where the rate of recovery from this disease is minimal then leads to death [3]. The doctors are prone to errors due to fatigue and work pressure or lack of experience. Therefore, many automatic classification systems have been developed to diagnose hepatitis, which allows the doctor to get a diagnosis correctly and quickly with the least effort and cost. Where it relies on the correct understanding and interpretation of biological analysis and clinical examination of the patient [4, 5] to obtain an accurate medical diagnosis with high accuracy, two main factors must be available: First, the necessary features of hepatitis disease should provide, whether visible symptoms, blood tests or sonar images. Second, choosing the correct and robust algorithm which does not take time to work.

ANNs have proven suitable for adequate diagnosis of various diseases. Also, their use makes the diagnosis more reliable and therefore increases patient satisfaction [6, 7]. Various categories of neural networks have been used with different characteristics in disease diagnosis. Many researchers recently focused on the use of neural networks for medical diagnosis. There was a fast, accurate, and fully automatic method of segmentation, a brain tumor based on deep neural networks [8]. H. Abdullah and M. Habtr have recommended brain cancer detection and identification system. Here, based on ANN, the result gives three-class according to feature extraction normal, benign, and malignant [9]. The advantages of ANNs help for efficient classification of given data. Thus it is adopted for the classification of heart disease dataset by considering the single and multilayer neural network modes [10]. A model of a heart rate regulation system built with the aid of FPGA techniques was proposed by H. Abdullah and B. Abd [11]. ANN has proven to be a very effective means of pattern recognition; this has made it very useful for diagnosing cancer in the very early stages [12].

In this paper, technology will harness in the service of medicine, especially in the area of diagnosis of diseases, because the diagnosis is the correct beginning of healing. The objectives of this work to satisfy the desired aims are: Study the previous works that dealt with this subject. Employ the available online database for the diagnosis of hepatitis disease to service the proposed system. Collect real database from the Gastroenterology and Liver Education Hospital in Baghdad Medical City to evaluate the performance of the proposed work. It was choosing intelligent classification techniques that provide the highest classification accuracy. Evaluating the performance of the proposed classifiers via MATLAB 2014a software based on the two databases.

The proposed diagnosis hepatitis system has many features, listed as follows:

- A new and extensive database of viral hepatitis was created that is employed to measure the performance of any classifier.
- A new classification of the severity of hepatitis was adopted, by dividing the collected database into five categories according to the severity of the disease: normal condition, chronic hepatitis, mild cirrhosis, moderate cirrhosis, acute cirrhosis.
- The proposed system achieved high diagnostic accuracy, despite the use of a more massive database than the UCI database and the use of features considered primitive compared to the features of UCI.

In the rest of the paper, Part 2 presents the relevant work, section 3 Introduce the proposed diagnosis hepatitis system, and explain the process and the purpose of collecting the new database that used by the proposed system, also, display the training results of each classifier, section 4 presents the experimental results, and it is discussions for the three suggested classifiers, also show a comparison with many techniques of previous studies, and section 5 concludes the work.

2. Related work

In recent years, technology has been harnessed to serve the medical field on a large scale for a diversity of purposes. One of the essential objectives is the development of decision support devices that increase the confidence of doctors to make sound decisions to diagnose and evaluate the patient's condition.

Many researchers have worked to develop medical decision support systems that diagnose hepatitis. The automated system of medical diagnosis would improve the medical services provided to the patient and reduce costs and diagnosis errors. ANNs are mightily applied in the medical field because of its benefit in multivariable classification problems with a high success rate.

A.H.Roslina and A.Norazia [13] improved classification prediction accuracy rate, by using only ten attributes out of 20 given from the UCI raw data. By gathering the Wrappers Method with SVM techniques instead of using SVM alone, they managed to increase the diagnostic accuracy rate from 72.73% up to 74.55%.

H. Chen, D. Liu, B. Yang, J. Liu, and G. Wang [14] proposed system which gathers (LFDA) and supporting support device (SVM) was achieved the highest degree of accuracy of classification (96.77%) when dividing the UCI data to 80% for training the system and 20% for testing the system for diagnosis of hepatitis. The proposed LFDA_SVM is compared to three existing methods, including SVM based on Basic Component Analysis (PCA_SVM) and SVM based on Fisherman Excellence Analysis (FDA_SVM) and the SVM standard concerning classification accuracy. Experimental results showed that LFDA_SVM significantly outperforms the other three methods. M. Neshat, A. Masoumi, M. Rajabi, and H. Jafari, [15] proposed a fuzzy Hopfield neural network (FHNN) that has been used as the detection of the severity of the hepatitis disease. The data was taken out from (UCI) database, where FHNN can classify hepatitis with an average precision of 92.05%.

B. Femina and S. Anto [16] planned to companied K Nearest Neighbor with Rough Set (RS-KNN) as classification. Rough Set to reduce the attributes number as possible to find the least subset. Dataset took from (UCI) repository dataset. This method achieves 84.52 % of accuracy.

S. Godara and R. Singh [17] used many techniques as Decision Tree, K-Nearest Neighbors, Radial Basis Function neural networks, ANNs, and SVM to test their efficacy in diagnosing each of the following diseases Hepatitis Wisconsin, Breast Cancer, Liver Disorder and cardiovascular Cleveland Heart disease datasets from archive UCI. Hepatitis diagnosis accuracy was 81.93%, 83.87%, 80.01%, 85.16% and 85.16% respectively.

In [18], an intelligent hybrid approach, combining the information gain method and the Adaptive neuro-fuzzy inference system (ANFIS), is proposed to diagnose killer hepatitis. The performance of this approach was estimated using the statistical process. The elevated results for classification accuracy, specificity, and sensitivity analysis for this system were 95.24%, 91.7%, and 96.17%, respectively.

Previous researches have improved accuracy as much as possible to obtain a correct diagnosis of hepatitis diagnosis using UCI data. However, the improvement in accuracy through factors that can significantly affect the accuracy of the diagnosis, which is disease features, number, and quality of training examples, has rarely been examined in previous research. To address this issue, this paper proposed a new multiclass classifier approach based on a collected real database with new proposed features that reflect the precise situation of the disease. The proposed Hepatitis diagnostic technology is adopted to help the doctor make the right decision quickly and accurately to reduce cost, time, and diagnosis. It is an innovative attempt at this diagnosis

3. The proposed method

3.1 Database acquisition

The University of California, Irvine (UCI) which most widely available on the internet used by many researchers, will be used in this work to test all proposed algorithms in the diagnosis of hepatitis [19-21]. These data are classified into two groups: the first group representing healthy people who did not have hepatitis, the second group is 123 people out of 155 with hepatitis. UCI data is based on 19 features, but some of these features have had difficulty collecting in Iraqi hospitals because it costs the patient a lot of pain, and the doctor costs the effort as well as the lack of some modern medical devices. Therefore, adopted in collecting data on consultation with specialist doctors on replaced some features with readily available features without the inconvenience and cost to the patient and the doctor.

Table 1. The classification of five classes of hepatitis disease of the collected database

Hepatitis disease diagnosis	**Number of patients**
Chronic	28
Cirrhosis Simple	24
Moderate Cirrhosis	86
Severe Cirrhosis	62
Normal	100
Total	300

Applied data in this paper for diagnosing the severity of hepatitis have chosen from the patients in the Gastroenterology and Liver Education Hospital of the City of Medicine in Baghdad. These data have 300 records, were 100 of which healthy and 200 infected with hepatitis. Each record has 15 fields (features) that were determined based on the consultation of expert physicians. Two of which are binary and the rest was a range of values, including blood tests and sonar examination. In this paper, the severity of hepatitis is divided into five classes. These classes are placed in Table 1 according to the specialist's diagnosis to compare the result together after designing the expert system and convention system.

The fields (features) for each record represent the specialist’s diagnosis that helps to diagnose the severity of hepatitis. These features have put into Table 2.

3.2 Division of data set

The database used in this work, whether the UCI data or the collected data, will be divided into two parts: part of these samples for training the networks and the other part for testing the networks. In the beginning, the UCI data divided according to a set of different divisions as in previous works, as shown in Table 3. These divisions will be used to train and test the proposed classifiers and make a fair comparison with the other techniques employed to diagnose hepatitis in previous researches with different factors such as data size, learning cycle, and processing time to achieve high diagnostic accuracy and estimated error. While the collected data will be divided into 50% for training and 50% for testing the proposed work.

Table 2. The features of hepatitis disease for the collected database

<table>
<tr><th>No.</th><th>collected database</th><th colspan="4">The description</th></tr>
<tr><td>1</td><td>Age</td><td colspan="4">ranges from 13 to 70 years old</td></tr>
<tr><td>2</td><td>Gender</td><td colspan="2">Male 1""</td><td colspan="2">Female 2""</td></tr>
<tr><td rowspan="2">3</td><td rowspan="2">Plt</td><td colspan="4">Platelets (PLT) are small fragments of cells that are essential to healthy blood coagulation</td></tr>
<tr><td colspan="4">450 – 150</td></tr>
<tr><td rowspan="2">4</td><td rowspan="2">Hb</td><td colspan="4">Haemoglobin Concentration (Hb)</td></tr>
<tr><td colspan="4">16 – 12</td></tr>
<tr><td rowspan="2">5</td><td rowspan="2">WBC</td><td colspan="4">White blood cells (WBCs)</td></tr>
<tr><td colspan="4">1100 – 4000</td></tr>
<tr><td rowspan="2">6</td><td rowspan="2">ALT</td><td colspan="4">Alanine Aminotransferase (ALT)</td></tr>
<tr><td colspan="4"><=37</td></tr>
<tr><td rowspan="2">7</td><td rowspan="2">INR</td><td colspan="4">International Normalized Ratio</td></tr>
<tr><td colspan="4"><=1</td></tr>
<tr><td rowspan="2">8</td><td rowspan="2">Liver big</td><td colspan="4">abnormal liver size</td></tr>
<tr><td colspan="2">YES 1""</td><td colspan="2">NO 2""</td></tr>
<tr><td rowspan="2">9</td><td rowspan="2">Spleen Palpable</td><td colspan="4">Abnormal spleen size</td></tr>
<tr><td colspan="4"><=12</td></tr>
<tr><td rowspan="2">10</td><td rowspan="2">Ascites</td><td colspan="4">Ascites is the abnormal buildup of fluid in the abdomen</td></tr>
<tr><td>None 1</td><td>Mild 2</td><td>Moderate 3</td><td>Sever 4</td></tr>
<tr><td rowspan="2">11</td><td rowspan="2">Bilirubin</td><td colspan="4">Bilirubin is a chemical dye produced in the liver, spleen. The analysis of bilirubin is an analysis in which the process of testing the overall amount of bilirubin; which is responsible for the patient's yellow color</td></tr>
<tr><td colspan="4"><=1.2</td></tr>
<tr><td rowspan="2">12</td><td rowspan="2">Alk</td><td colspan="4">Alkaline Phosphatase Test (ALP)</td></tr>
<tr><td colspan="4">116 – 46</td></tr>
<tr><td rowspan="2">13</td><td rowspan="2">SGOT/AST</td><td colspan="4">Aspartate Aminotransferase (AST)</td></tr>
<tr><td colspan="4"><=37</td></tr>
<tr><td rowspan="2">14</td><td rowspan="2">Albumin</td><td colspan="4">Blood serum contains significant amounts of protein. Albumin is the essential protein found in the blood</td></tr>
<tr><td colspan="4">5 - 3.4</td></tr>
<tr><td rowspan="2">15</td><td rowspan="2">PT</td><td colspan="4">Prothrombin time</td></tr>
<tr><td colspan="4">< =13</td></tr>
</table>

Table 3. Previous works data division of UCI database

Reference	Training Data	Testing Data
[20,21]	100	55
[22]	95	60
[23,24]	116	39
[14,16,25]	132	23
[26]	108	47

3.3 The proposed networks model for hepatitis diagnosis

Currently, many different data classification technologies are available. New technologies were developed either to improve the performance of a previous technology or to find a new method that could better solve the classification problem. With classification technology, the primary purpose of a classifier is to classify the collected data into appropriate categories, and when the newly collected data category is unknown, a new category can also be assigned. Fig. 1 show the proposed diagnosis system.

3.3.1. Proposed diagnosis system based on BPNN

The proposed network consists of the three basic layers for each dataset division, as shown in Fig. 2. The input layer that binds the network to surrounding environment the number of nodes in this layer represents the number of medical factors for patients to diagnosis hepatitis disease. The second layer is the

Received: February 10, 2020

hidden layer, which may be of one or more levels, the best accuracy of the classification is depend on the number of levels and the number of neurons in each level. The best accuracy has been obtained in this work is when using a hidden layer with two levels and five neurons in each level.

Finally, the third and last layer is the output layer, from which know the class to which the person belongs according to the features produced in the inputs. In this proposed neural network, the activation function in both the input layer and the output layer is the linear transmission function, while the activation function used in the hidden layer is sigmoid.

In this work, to classify the UCI database into two categories, healthy and infected, the proposed network consists of three layers. An input layer that has 19 neurons to represent the healing properties of patients, as shown in Fig. 3.

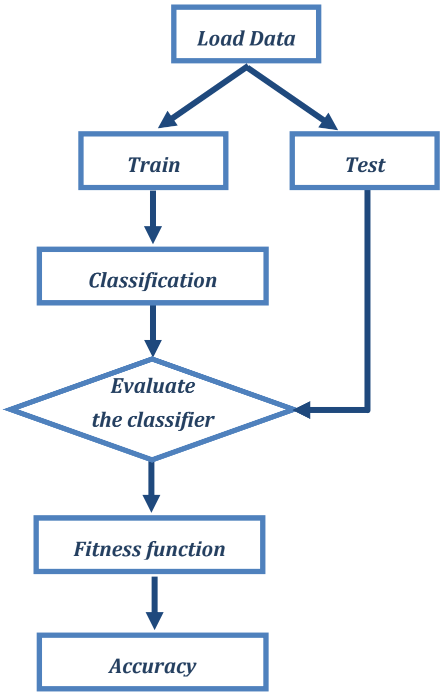

Figure. 1 Proposed hepatitis disease diagnosing system

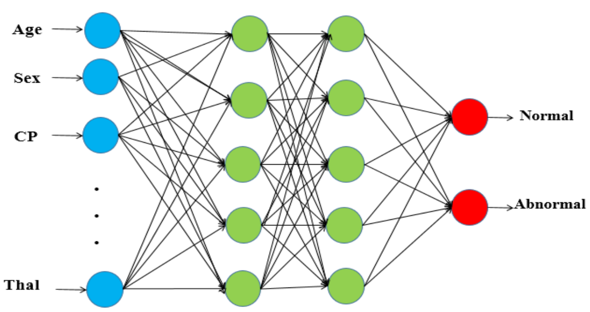

Figure. 2 A proposed diagnosing system based on BPNN for two-class classification

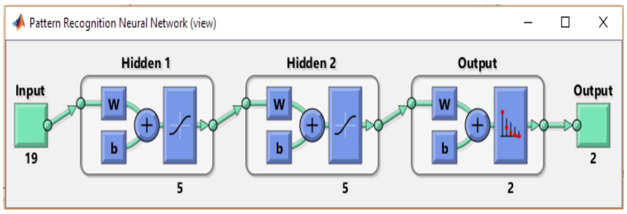

Figure. 3 Topology structure of BPNN for two-class classification using UCI data

Figs. 4-7 display the performance of the proposed BPNN in the training phase according to database division.

The performance of (MSE) Mean Square Error in these figures is a lesser amount of1×10^{-6}, and the accuracy of the correct classifications is 100% when the training group itself tests the system for all database division.

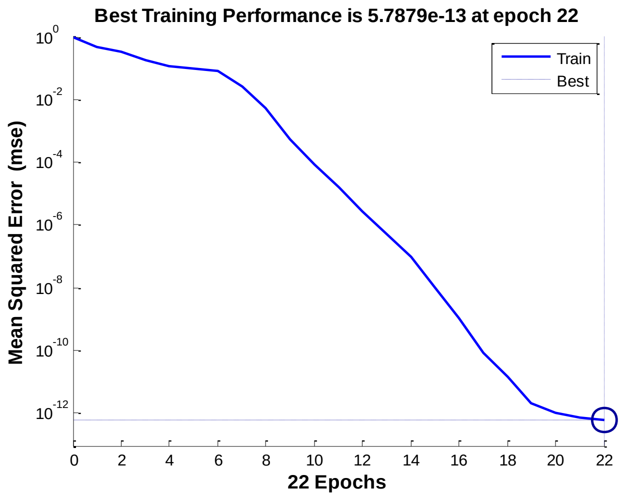

Figure. 4 Training performance of BPNN for 100 training samples of UCI

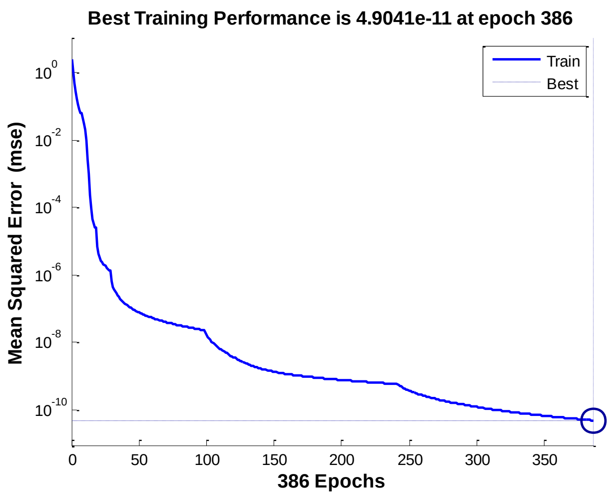

Figure. 5 Training performance of BPNN for 95 training samples of UCI

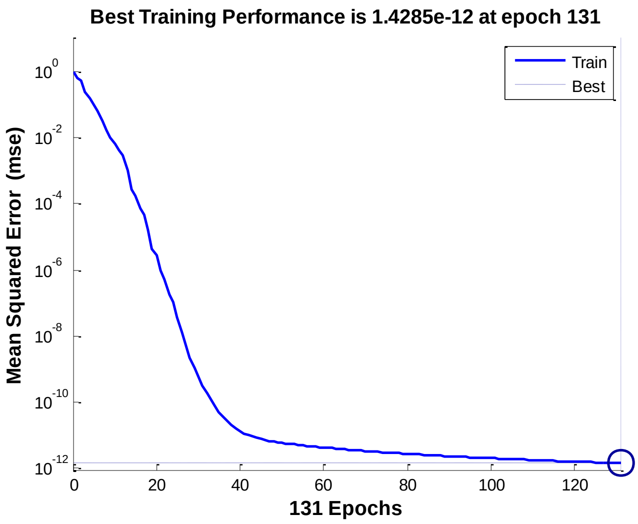

Figure. 6 Training performance of BPNN for 116 training samples of UCI

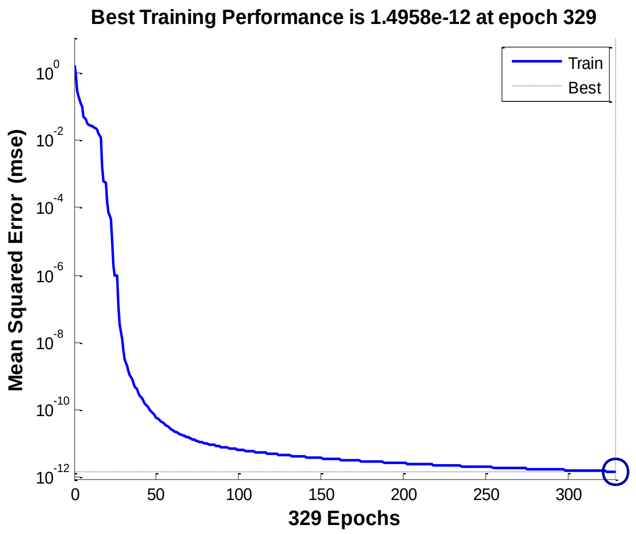

Figure. 7 Training performance of BPNN for 132 training samples of UCI

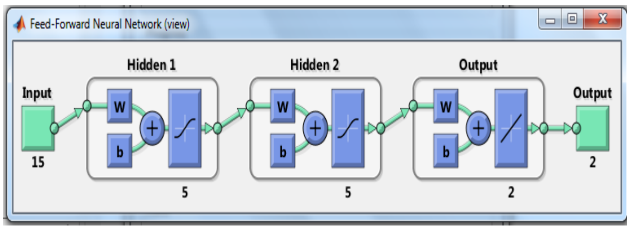

Figure. 8 Topology structure of BPNN for two-class classification using collected data

When using 50% of the collected data to train this network to classify two-class, as shown in Fig. 8, the MSE performance reached3.0109×10^{-15}, at epoch 23, as shown in Fig. 9. Also, the precision of the accurate classifications is 100% when the training group itself tests the system.

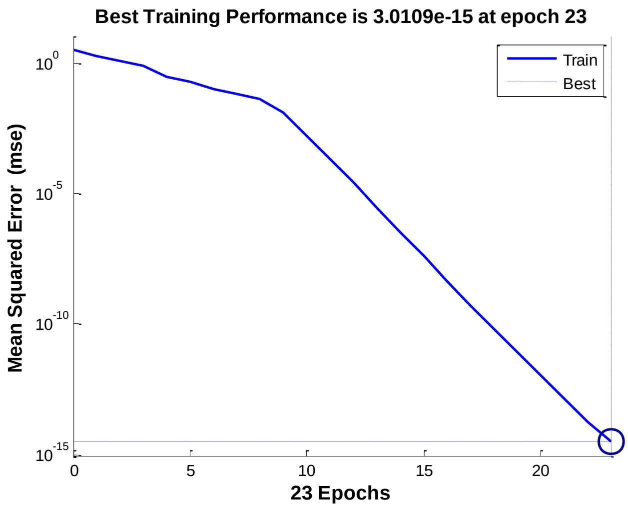

Figure. 9 Training performance of BPNN for 150 training samples of collected data

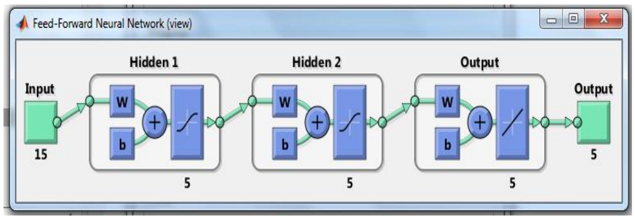

Figure. 10 Topology structure of BPNN for five-class classification using collected data

To classify collected data into five categories, the proposed network shown in Fig. 10. Five neurons in the output layer representing the five classes (four of which represent the level of hepatitis disease intensity as well as the normal state). The performance result for the five categories shown in Fig. 11.

3.3.2. Proposed diagnosis system based on RBFN

Powerful feedforward neural networks used in several uses such as time series prediction, approximation, system control, and classification. It consists of three primary layers: the input layer (input vector), the hidden layer (RBF neurons), and the output layer (class nodes), as shown in Fig.12.

The activation function is a commonly Gaussian function of hidden layer neurons: as shown in Eq. (1) [27]:

$$f(x) = \sum_{j=1}^{m} w_j h_j(x) \tag{1}$$

Where: wj is the weight between the hidden layer and the output layer;

Received: February 10, 2020

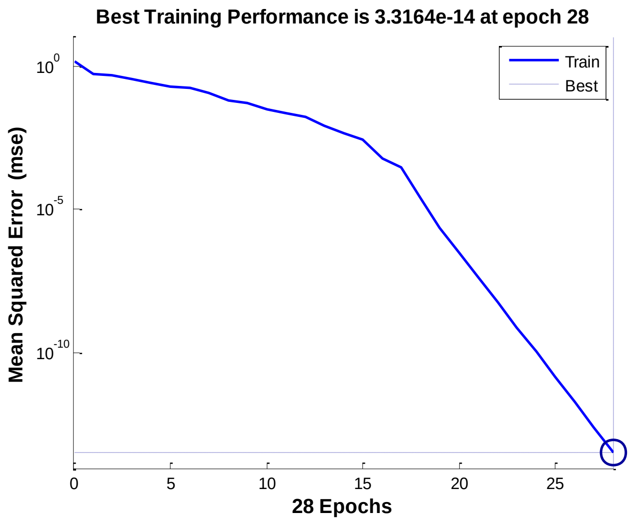

Figure. 11 Training performance of BPNN for five-class

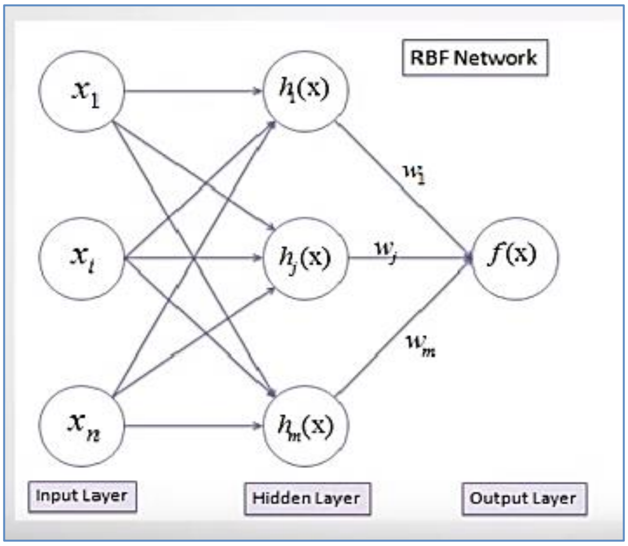

Figure. 12 Topology structure of an RBF-NN

$$h(x) = \exp^{\left(-\frac{(x_i - c_j)^2}{\sigma_1^2}\right)} \tag{2}$$

C_i is the center of a radial basis function of the neuron, σ, stander deviation.

According to the Gaussian function, we must find the center and sigma for each neuron in the hidden layer and corresponding weight. This three-parameter defined separately at each RBF unit. So the k-means clustering algorithm used to find the cluster centers and K-nearest neighbor used to calculate the width or radius of the bell-shape (σ) appropriately.

The standard methods to find the width or radius of the gaussian curve is normalization method or P-nearest neighbor rule. For every node j the width σ_j is calculated via the P-nearest neighbour heuristic:

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{p}(c_j - c_i)^2}{p}} \tag{3}$$

Where: $c_1, c_2, \ldots, c_p$ are the p nearest node centres to the hidden node j. The parameter p is selected so that a large amount of nodes is activated when an input vector is presented to the neural network model.

When using UCI data, the input layer consists of 19 nodes (represented the disease features), while the numbers of nodes in hidden layer depend on the number of the centers which varies according to division of database, as shown in Table 4, since the UCI repository classify the data into normal and abnormal so the number of nodes in the output layer two nodes only one node for each class.

When using 50% from the collected data for training, the input layer was 15 nodes, and the number of centres of the hidden layer was 19 RBF neurons in two-class classification and 32 RBF neurons in five class classification. So the output layer was two nodes and five nodes in 2 classes and five classes, respectively, as shown in the Table 4.

3.3.3. Proposed diagnosis system based on KNN

KNN is a simple algorithm, and the best classification methods used in most medical diagnostics. A fundamental feature is that it maintains stored data and cannot be lost, and there is no training stage.

The principle of its operation is that we must store in the network part of the known data targets and upon entering a foreign element in the network, a class of this element can be predicted by calculating the distance from one of these equations (Euclid, Cosine, Correlation, Exaggeration, and Manhattan) between a strange element and the elements stored in the grid according to a value as determined by the programmer, which must be strange. Eq. (3) explain the Manhattan distance calculation [28].

$$d = \sum_{j=1}^{n} |a_j - b_j| \tag{4}$$

Where n is standing for the feature size. The class of the new sample evaluated can be identified by majority voting of the K nearest samples classes.

In this work, used KNN as classification system that works on the principle of measuring the distance between all the data in which the network was trained and any new element entering the network by city block distance equation, through the minimum

Table 4. RBFN training results

No. of Training Data		No. of Class	No. of centers	Training time	Training accuracy
UCI	116	2	17	0.012915 sec	94.8%
	95	2	14	0.047476 sec	97.9%
	100	2	16	0.013416 sec	95%
	108	2	17	0.027123 sec	96.3%
	132	2	18	0.014554 sec	96.2%
CD	150	2	19	0.046357 sec	96%
		5	32	0.118287 sec	86.67%

Table 5. KNN training results

No. of Training Data		No. of Class	Training time	Training accuracy
UCI	116	2	0.008301 sec	100%
	100	2	0.007956 sec	100%
	95	2	0.008728 sec	100%
	108	2	0.004428 sec	100%
	132	2	0.004440 sec	100%
CD	150	2	0.007941 sec	100%
		5	0.008103 sec	100%

Predicted \ **Actual**	*Positives (1)*	*Negatives (0)*
Positives (1)	*True positives (TP)*	*False Positives (FP)*
Negatives (0)	*False Negatives (FN)*	*True Negatives (TN)*

Figure 13. Confusion matrix

distance based on the value of K=1, the new element is classified into the category to which it belongs is based on the nearest one neighbor.

This network trained using the UCI database that classifies the data normal and abnormal according to database divisions in previous researches. It also used 50% of the collected data to the training this algorithm to classify data in two-categories and five-categories of hepatitis disease, as shown in Table 5. The MATLAB program was written to evaluate the performance of the classification algorithm by testing it on the same training data by calculated the accuracy and speed (time training) of the system in all cases.

4. Results and performance analysis

To assess the effectiveness of this work, we will calculate some known and reliable criteria for all proposed networks: specificity, PPV, NPV, sensitivity, and accuracy of system diagnostics. The sensitivity measures the percentage of real positives, which are recognized correctly. It diagnoses persons living with hepatitis disease correctly as hepatitis disease positive.

Specificity used to measure the percentage of persons who are correctly identified as there is no hepatitis, any condition where a group of healthy people is recognized correctly as there is no hepatitis [29].

The prediction results of a classifier can be represented by the Confusion Matrix. The confusion matrix was obtained to calculate sensitivity, specificity, and accuracy. Fig. 13 demonstrates the confusion matrix. Accuracy in classification problems is the number of correct predictions made by the model over all kinds of predictions made.

$$Accuracy = \frac{(T_P + T_N)}{(T_P + T_N + F_P + F_N)} \times 100\% \tag{5}$$

$$Sensitivity = \frac{T_P}{T_P + F_N} \times 100\% \tag{6}$$

$$Specficity = \frac{T_N}{T_N + F_P} \times 100\% \tag{7}$$

As well as specificity and sensitivity, performance can also be evaluated using the Positive Prediction Value (PPV), which deduces how a positive test result has succeeded in deducing the presence of a disease. Also, Negative-prediction value (NPV), concluded well how to conclude the negative test result that this disease is absent.

$$PPV = \frac{T_P}{T_P + F_P} \times 100\% \tag{8}$$

$$NPV = \frac{T_N}{T_N + F_N} \times 100\% \tag{9}$$

The results obtained based on the BPN network when using the UCI database to diagnosis hepatitis in two-class cases illustrated in Table 6. Also, when using 50% of CD samples for testing the classifier to recognize the data into two-class is shown in Table 6. As illustrated in Table 7, the Classification accuracy of the suggested BPN classifier is decreasing to 86.667% for multiclass diagnosis hepatitis disease when tested by using 50% of CD.

Table 6. Performance results of two class diagnosis system based on BPN according to UCI and CD database

Type of Data	Testing Data samples	T_P / F_N	F_P / T_N	Test time Sec	Specificity (%)	Sensitivity (%)	NPV (%)	PPV (%)	Accuracy (%)
UCI	60	11	4	0.013594	91.489	84.615	95.556	73.333	90
		2	43						
	55	5	4	0.011809	91.837	83.333	97.826	55.556	90.9091
		1	45						
	39	6	2	0.011757	93.93	100	100	75	94.8718
		0	31						
	23	6	0	0.015622	100	85.714	94.118	100	95.6522
		1	16						
	47	8	2	0.013859	94.594	80	94.594	80	91.4894
		2	35						
CD	150	98	2	0.012966	96.154	100	100	98	98.6667
		0	50						

Table 7. Performance results of five class diagnosis system based on BPN according to CD

Testing Data samples	T_P / F_N	F_P / T_N	Test time Sec	Specificity (%)	Sensitivity (%)	NPV (%)	PPV (%)	Accuracy (%)
150	80	20	0.012518	71.429	100	100	80	86.6667
	0	50						

Table 8. Performance results of two class diagnosis system based on RBFN according to UCI and CD database

Type of Data	Testing Data samples	T_P / F_N	F_P / T_N	Test time Sec	Specificity (%)	Sensitivity (%)	NPV (%)	PPV (%)	Accuracy (%)
UCI	60	11	4	0.013594	91.489	84.615	95.556	73.333	90
		2	43						
	55	5	4	0.011809	91.837	83.333	97.826	55.556	90.9091
		1	45						
	39	6	2	0.011757	93.93	100	100	75	94.8718
		0	31						
	23	6	0	0.015622	100	85.714	94.118	100	95.6522
		1	16						
	47	8	2	0.013859	94.594	80	94.594	80	91.4894
		2	35						
CD	150	98	2	0.012966	96.154	100	100	98	98.6667
		0	50						

Table 9. Performance results of five class diagnosis system based on RBFN according to CD

Testing Data samples	T_P / F_N	F_P / T_N	Test time Sec	Specificity (%)	Sensitivity (%)	NPV (%)	PPV (%)	Accuracy (%)
150	99	1	0.000088	97.561	90.826	80	99	92.667
	01	40						

Table 10. Performance results of two class diagnosis system based on KNN according to UCI and CD database

Type of Data	Testing Data samples	T_P / F_N	F_P / T_N	Test time Sec	Specificity (%)	Sensitivity (%)	NPV (%)	PPV (%)	Accuracy (%)
UCI	60	13	2	0.002699	95.745	100	100	86.666	96.6667
		0	45						
	55	11	1	0.003277	97.727	100	100	91.667	98.1818
		0	43						
	39	7	1	0.001971	96.875	100	100	87.5	97.4359
		0	31						
	23	6	0	0.002356	100	100	100	100	100
		0	17						
	47	10	0	0.002743	100	100	100	100	100
		0	37						
CD	150	99	1	0.001907	98.039	100	100	99	99.333
		0	50						

Table 11. Performance results of five class diagnosis system based on KNN according to collected data (CD)

Testing Data samples	T_P / F_N	F_P / T_N	Test time Sec	Specificity (%)	Sensitivity (%)	NPV (%)	PPV (%)	Accuracy (%)
150	99	1	0.001907	98.039	100	100	99	99.333
	0	50						

The rest performance parameters results of the RBFN classifier illustrated in Table 8. As revealed in Table 9, the RBFN is an approximately suitable algorithm for diagnosis multiclass classification of hepatitis disease, which gets 82% diagnosis accuracy depending on 50% of CD.

The performance results of the proposed KNN classifier are displayed in Table 10. This network obtained very high classification results when using the UCI database and CD. Table 11 shows the efficiency of the KNN algorithm for five class recognition based on CD.

After calculated the testing time, specificity, sensitivity, NPV%, PPV%, and classification accuracy for each network separately, a fair comparison under the same conditions as the same data and same division of samples is done to find the best classifier from the three proposed classifiers in order to be implemented using FPGA kit in future.

Fig. 14 shows the performance comparison results among the three proposed classifiers when using 95 samples for training and 60 samples for testing. Accuracy is the most important measure of the classifier's performance, where the results were good for everyone, but the best is KNN 96.6667% with testing time is 0.0026sec, while BPN accuracy was 90% and RBF accuracy was 85%. PPV ratio is 86.666% for KNN, while BPN and RBFN result at a very low rate of 73.333% and 46.667%, respectively.

The performance comparison of three classifiers when using 50% (150samples) of collected data for testing the efficiency of the system to diagnosis hepatitis disease into a two-class case is illustrated in Fig. 15. Fig. 16 shows how the results were compared between the three algorithms in the multiclass classification when using 150 samples (50%) from collected data to recognize the five-class of hepatitis disease.

As demonstrated, the accuracy of K-NN reached 88%. It considers as good accuracy to multiclass classification compared with other classifiers at good testing time 0. 0034sec. PB also has excellent sensitivity, and NPV is 100% but with predicate accuracy lower than K-NN and testing time slower than KNN.

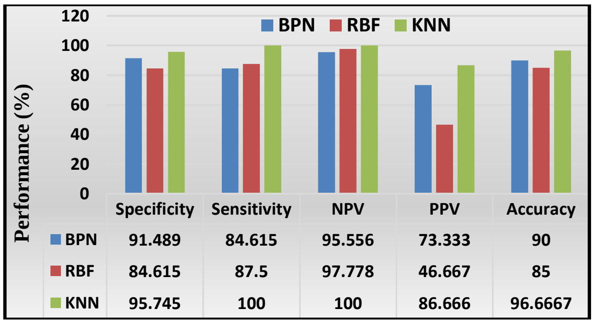

Figure. 14 The performance comparison of two-class hepatitis diagnosis system at 60 testing samples of UCI

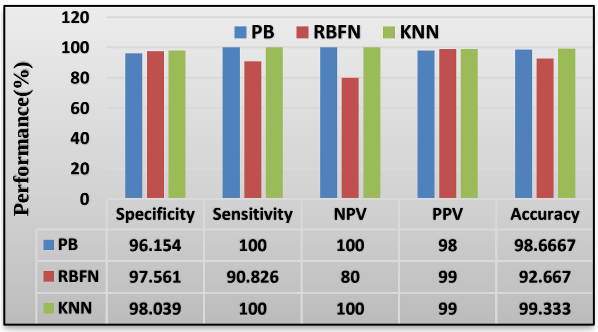

Figure. 15 The performance comparison of two-class hepatitis disease at 150 testing samples of CD

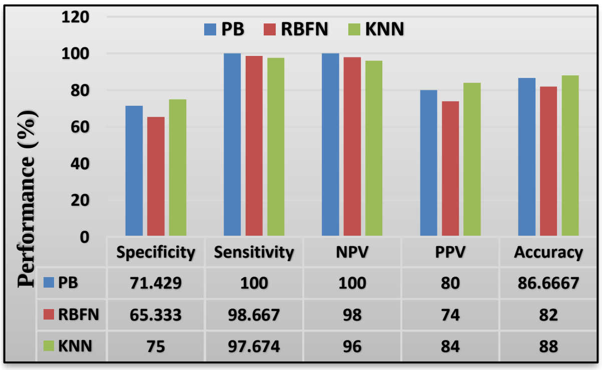

Figure. 16 The performance comparison of five-class hepatitis disease at 150 testing samples of CD

4 Conclusion

KNN makes timely predictions by calculating the similarity between the input sample and each training instance. If the training data is much greater than a number of features, KNN is better than others. There is no training involved in KNN. During the test, k neighbors with a minimum distance will participate in the classification. BPNN and RBFN need large training data and lots of hyperparameter settings compared to KNN to achieve sufficient accuracy. Previous researches have improved accuracy as much as possible to obtain a correct diagnosis of hepatitis diagnosis using UCI data. Table 12 shows the accuracy of each work and comparing these results with the results of the current study.

By considering all the above, we concluded that the RBFN very fast but not sufficiently accurate. Also, BPNN has good forecast accuracy but somewhat slow. While KNN network achieving the best results of diagnosis quickly and accurately in the case of classification into two class 99.33% and in the case of classification into five class were accuracy 88%, which is considered as good accuracy to multiclass classification.

References

[1] Y. Zhang, Y. Zhao, X. Lin, A. Wang, and D. Che, “Modeling for the prediction of Hepatitis B incidence based on integrated online search indexes”, *Informatics in Medicine Unlocked*, Vol. 10, pp. 143-8, 2018.

[2] S. Anto, S. Chandramathi, “An expert system based on SVM and hybrid GA-SA optimization for hepatitis diagnosis”, *International Journal of Computer Engineering in Research Trends*, Vol. 2, No. 7, pp. 437-443, 2015.

[3] E. A. Saad, “Non-invasive Assessment of Liver Fibrosis Using Serum Markers”, *Journal of Pharmaceutical, Chemical, and Biological Sciences*”, Vol. 2, No. 2, pp. 59-76, 2014.

[4] H.N. Abdullah, B.H. Abd, and S. H. Muhi, “High-Resolution Systems for Automated Diagnosis of Hepatitis”, In: *Proc. of the Third Scientific Conference of Electrical Engineering*, pp. 39-44, 2018.

[5] D. Avci, “An Automatic Diagnosis System for Hepatitis Diseases Based on Genetic Wavelet Kernel Extreme Learning Machine”, *Journal of Electrical Engineering & Technology*, Vol. 11, No. 4, pp. 993-1002, 2016.

[6] R. Nanjundegowda and V.A. Meshram, “Arrhythmia Detection Based on Hybrid Features of T-wave in Electrocardiogram”, *International Journal of Intelligent Engineering and Systems*, Vol. 11, No. 1, pp.153-162, 2018.

[7] J. Jiang, H. Wang, J. Xie, X. Guo, Y. Guan, and Q. Yu, “Medical knowledge embedding based on recursive neural network for multi-disease diagnosis”, *Artificial Intelligence in Medicine*, Vol. 103, 2020.

[8] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. M. Jodoin, and H. Larochelle, “Brain tumor segmentation with deep neural networks”, *Medical Image Analysis*, Vol. 1, No. 35, pp. 18-31, 2017.

[9] H. N. Abdullah and M. A. Habtr, “Brain Tumor Extraction Approach in MRI Images Based on Soft Computing Techniques”, In: *Proc. of the*

8th International Conference on Intelligent Networks and Intelligent Systems, pp. 21-24, 2015.

[10] Y. Khourdifi and M. Bahaj, "Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization", *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 1, pp. 1-11, 2019.

[11] H. N. Abdullah and B. H. Abd, "A simple FPGA system for ECG RR interval detection", In: *Proc. of the* 11th Conference on *Industrial Electronics and Applications*, pp. 1379-1382, 2016.

[12] S. Agrawal and J. Agrawal, "Neural network techniques for cancer prediction: A survey", *Procedia Computer Science*, Vol. 60, pp. 769-774, 2015.

[13] A. H. Roslina and A. Noraziah, "Prediction of Hepatitis Prognosis Using Support Vector Machines and Wrapper Method", In: *Proc. of the Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 2209-2211, 2010.

[14] H. L. Chen, D. Y. Liu, B. Yang, J. Liu, and G. Wang, "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, pp. 11796-11803, 2011.

[15] M. Neshat, A. Masoumi, M. Rajabi, and H. Jafari, "Diagnosing Hepatitis Disease by Using Fuzzy Hopfield Neural Network", *Annual Research & Review in Biology*, Vol. 4, No. 17, pp. 2709-2721, 2014.

[16] B. Femina and S. Anto, "Disease diagnosis using rough set-based feature selection and K-nearest neighbor classifier", *International Journal of Multidisciplinary Research and Development*, Vol. 2, No. 4, pp. 664-668, 2015.

[17] S. Godara and R. Singh, "Evaluation of Predictive Machine Learning Techniques as Expert Systems in Medical Diagnosis", *Indian Journal of Science and Technology*, Vol. 9, No. 10, pp. 0974-5645, 2016.

[18] W. Ahmad, A. Ahmad, A. Iqbal, M. Hamayun, A. Hussain, G. Rehman, S. Khan, U. U. Khan, D. Khan, and L. Huang, "Intelligent hepatitis diagnosis using adaptive neuro-fuzzy inference system and information gain method", *Soft Computing*, Vol. 23, No. 21, pp.10931-10938, 2019.

[19] https://archive.ics.uci.edu/ml/machine-learning databases/hepatitis/hepatitis.data

[20] S. Ansari, I. Shafi, A. Ansari, J. Ahmad, and S. I. Shah, "Diagnosis of liver disease induced by hepatitis virus using artificial neural networks", In: *Proc. of the 14th International Multitopic Conference*, pp. 8-12, 2011.

[21] R. R. Janghel, A. Shukla and K. Verma, "Soft Computing Based Expert System for Hepatitis and Liver Disorders", In: *Proc. of International Conference on Engineering and Technology*, 2016.

[22] E. Dogantekin, A. Dogantekin, and D. Avci, "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, pp. 11282-11286, 2009.

[23] M. Neshat, M. Sargolzaei, A. N. Toosi and A. Masoumi, "Hepatitis Disease Diagnosis Using Hybrid Case-Based Reasoning and Particle Swarm Optimization", *ISRN Artificial Intelligence*, pp. 1-6, 2012.

[24] M. S. Uzer, N. Yilmaz, and O. Inan, "Feature Selection Method Based on Artificial Bee Colony Algorithm and Support Vector Machines for Medical Datasets Classification", *The Scientific World Journal*, pp. 1-10, 2013.

[25] S. Mishra, B. K. Mishra, and H. K. Tripathy, "A Neuro-Genetic Model to Predict Hepatitis Disease Risk", In: *Proc. of International Conference on Computational Intelligence and Computing Research*, pp. 1-3, 2016.

[26] N. P. Husain, N. N. Arisa, P. N. Rahayu, A. Z. Arifin, and D. Herumurti, "Least Squares Support Vector Machines Parameter Optimization Based on Improved Ant Colony Algorithm for Hepatitis Diagnosis", *Jurnal Ilmu Komputer dan Informasi* (*Journal of Computer Science and Information*), Vol. 10, No. 1, pp. 43-49, 2017.

[27] R. Cheruku, DR. Edla, and V. Kuppili, "Diabetes classification using radial basis function network by combining cluster validity index and bat optimization with novel fitness function", *International Journal of Computational Intelligence Systems*, Vol. 10, No. 1, pp. 247-265, 2017.

[28] Y. Hamid, B. Ranganathan, L. Journaux, Q. Farooq, and S. Muthukumarasamy, "An Improvised k-NN Respecting Diversity of Data for Network Intrusion Detection", *International Journal of Intelligent Engineering and Systems*, Vol. 10, No. 3, pp.409-417, 2017.

[29] N. G. Siddappa and T. Kampalappa, "Adaptive Condensed Nearest Neighbor for Imbalance Data Classification", *International Journal of*

Intelligent Engineering and Systems, Vol. 12, No. 2, pp.104-113, 2019.

Table 12. Comparison result with previous works of two-class hepatitis disease according to the UCI database

References	**Algorithm**	**No. of features**	**Accuracy %**	**Training data**	**Testing data**	**Proposed Work Accuracy %**	
[22]	LDA+ANFIS	8	94.16	95	60	BPN	90
						RBFN	85
						KNN	96.66
[5]	GA-WK-ELM	19	96.642	100	55	BPN	90.9091
[21]	BPN		88.0727			RBFN	90.9091
	PNN		80.0000				
	CLN		66.3636				
	LVQ		92.7273			RBFN	98.1818
	Elman Networks		86.1818				
[20]	(FFNN)		91.33				
	(GRNN)		92				
	Self- (SOM)		Unable to diagnose				
[24]	ABCFS + SVM	11	94.92	116	39	BPN	94.8718
						RBFN	94.8718
[23]	(CBR-PSO)	19	93.25			KNN	97.4359
[25]	(MLP-GS)	9	82.58	132	23	BPN	95.65
[16]	RS-KNN	4	84.52			RBFN	86.9565
[14]	LFDA_SVM	19	96.77			KNN	100
[26]	IACA-LSSVM	10	91.6	108	47	BPN	91.489
	LFDA-LSSVM	8	83.3			RBFN	87.234
	LSSVM	19	79.1			BPN	100