International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

Research Article

Heart Disease Diagnosis System based on Multi-Layer Perceptron neural network and Support Vector Machine

Tabreer T. Hasan*, **Manal H. Jasim and Ivan A. Hashim**

University of Technology/ Department of Electrical Engineering, Baghdad-Iraq

Received 15 Aug 2017, Accepted 01 Oct 2017, Available online 07 Oct 2017, **Vol.7, No.5 (Sept/Oct 2017)**

Abstract

The area of medical information has advanced around organizing, preparing, storing, and transmit medical data for an assortment of purposes. One of these intentions is to create choice emotionally supportive networks that upgrade the human ability to analyze, treat, and evaluate forecasts of pathologic conditions. In this paper, heart disease diagnosis system has been built to classify two cases of heart conditions (Normal, Abnormal) in additional to classify five cases namely (Coronary Heart Disease, Angina Pectoris, Congestive Heart Failure, Arrhythmias, And Normal case), with high probability of classification. The proposed Heart disease diagnostic system consists of two types of database are used in the classification process; The online database which is taking from UCI learning data set repository for diagnosis heart disease and collected database from Ibn Al-Bitar Hospital Cardia Surgery and Baghdad Medical City. These databases consist of thirteen medical factors that are successful to diagnosis heart disease. Two heart diseases classifiers are proposed. They are; Multi-Layer Perceptron neural network (MLP), and Support Vector Machin (SVM). The simulation results show that, the MLP classifier has 98% accuracy of two heart diseases classification when the performance of this classifier was evaluated using collected database. While the accuracy of SVM classifier is reached 96%. Also, MLP has overcome from SVM classifier when classify four type of heart disease in additional to normal case for accuracy reached to 81%.

Keywords: *Heart disease, Multi-Layer Perceptron neural network (MLP), and Support Vector Machin (SVM).*

1. Introduction

Medical decision support system is a decision-support program which is designed to assist physicians and other health professionals with decision making tasks, such as determining diagnosis of patients' data. This approach helps employees make more informed medical decisions while working with their own physician [D. Vadicherla and S. Sonawane,2013]. Recently, the application of intelligent algorithms in medical field is the subject of researches, which mainly concentrates on modeling some of the human actions or thinking processes and recognizing diseases from a variety of input sources (e.g. patient's data cardiograms, Computerized Axial Tomography (CAT) / Magnetic Resonance Imaging (MRI) /Ultrasound scans, photomicrographs, etc.) [S. U. Ghumbre and A. A. Ghatol, 2012]. The researchers are used database for diseases such as breast cancer, Alzheimer, Malaria and Dengue Disease, etc. in order to diagnose the disease at less time and high accuracy based on various intelligent techniques as artificial neural networks, support vector machines, decision trees, Fuzzy Logic, etc. [S. Ergin and O. Kilinc,2014, M. S. Kidwai and A. M. Khan,2014, P. Sharma *et al* ,2013]. In modern times, the number of people suffering from heart disease is on a rise, and a large number of people dies every year due to the heart disease everywhere throughout the world [D. Vadicherla and S. Sonawane,2013]. Fig.(1) shows ten main factors that have led to the deaths over the world, as illustrated, heart diseases are being to become the main cause of death.

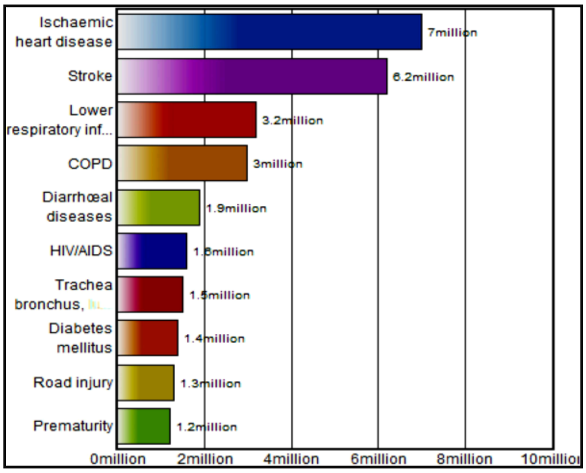

Figure 1: Ten leading causes of death in the world

*Corresponding author's ORCID ID: 0000-0002-6554-5947

For this reason, it was necessary to design a system that diagnoses heart disease because accurate diagnosis at an early stage followed by proper subsequent treatment can result in significant lifesaving. Unluckily, accurate diagnosis of heart illness at near the beginning stage is quite a challenging task because of complex interdependence on a variety of factors. For this reason, there is an urgent require to develop medical analytic decision support systems which can assist medical physicians in the diagnostic procedure [M. G. Feshki and O. S. Shijani,2016].

2. Literature Review

Diagnosis of heart disease using intelligent techniques attracted the attention of researchers. In the recent years, diagnostic systems have been implemented using many techniques. such as The Support Vector Machine (SVM), Neural Network, regression, decision trees(DT), Naive Bayesian(NB), etc. It based on several databases of patients from around the world to obtain a high accuracy diagnosis.

S. Bhatia *et al*, 2008 proposed a decision support system for heart disease classification based on SVM and integer-coded genetic algorithm (GA) for selecting the important and relevant features and discarding the irrelevant and redundant ones. The Cleveland heart disease database was taken from the University of California, Irvine (UCI) training database repository which was given by Detrano. This data consists of 303 cases included 5 classes, each with 13 diagnostic features. 250 of data are used for learning, and the rests are employed for testing the SVM. The results of the five-class classification 72.55% was obtained using only 6 out of 13 features, as against an accuracy of only 61.93% using all the features. As two-class classification, the suggested technique gives a classification accuracy of 90.57%. M. Gudadhe *et al*. 2010 suggested a decision support system for heart disease classification based on SVM by used UCI learning data set. 200 instances of dataset are used for training, and the rest are used for testing the SVM to classify the heart disease data into two classes, which shows presence of heart disease or absence of heart disease with 80.41% accuracy. A. Khemphila and V. Boonjing,2011 proposed a classification system using MLP with Back- Propagation learning algorithm depended on UCI dataset. The 60% (n=182) of data are used for training and 40% (n= 121) of data are used for testing. The thirteen attributes are reduced to 8 attributes by using feature selection algorithm to compare the classification accuracy. The accuracy of 8 features in the validation data set is 80.99% while the accuracy of 13 features in the validation data set is 80.17%. Hence, the accuracy differs between 13 features and 8 features in the validation data set is 0.82%. M. A. M. Abusharian *et al*., 2014 designed and implemented an automatic heart disease diagnosis system based on ANN and Adaptive Neuro-Fuzzy Inference Systems (ANFIS). The data set used is UCI learning data. The accuracy of diagnosis heart disease in two class present or absent from heart disease about 87.04%, 75.93 for ANN and ANFIS respectively. E.O. Olaniyi *et al*, 2015 suggested system was modeled on MLP and SVM. The 270 samples are divided in to 162 samples for training and 108 sample for testing. The results obtained are 85%, 87.5% for MLP and SVM respectively. From this experiment SVM is the best network for the diagnosis of heart disease. X. Liu *et al* .2017 achieved maximum classification accuracy of 92.59% by proposed system consist of two subsystems: The Relief F and Rough Set (RFRS) feature selection system and the classification system based on C4.5 classifiers. The dataset for this system obtained from the UCI database. 303 instances70% of data for training and 30%for testing. In order to summary the authors work table (1) show previous works.

Table 1 summary previous works

Authors	Algorithm	Class	Features	Accuracy
S. Bhatia et.al.,2008	GA+SVM	2	13	90.57%
		5	13	61.93%
			6	72.55%
M. Gudadhe et. al. 2010	SVM	2	13	80.41%
A. Khemphila and V. Boonjing,2011	MLP	2	13	80.17%
			8	80.99%
M. A. M. Abusharian *et al*., 2014	ANN	2	13	87.04%
	ANFIS			75.93%
E. O. Olaniyi and O. K. Oyedotun .,2015	MLP	2	13	85%
X. Liu *et al* .2017	RFRS+C4.5	2	13	92.59%

2. Heart Disease

Human heart is one of most important organs in the body. If the operation of a heart is not proper, then it will affect the other body parts of human such as Brain, Kidney, etc. It is nothing more than a pump, which pumps blood throughout the body. If the circulation of blood in the body is inefficient then the organs like brain suffers and if the heart stops working altogether then death may occur within minutes. Life is completely dependent upon the efficient working of heart. Risk factors are conditions or habits that make a person more likely to develop a disease. They can also increase the chances that an existing disease will get worse. Important risk factors for heart disease that you can do something about are (K. Khoshraftar *et al*,2016, B. Kaur and W. Singh,2014):

- Family history of heart disease
- Smoking
- Cholesterol
- High blood pressure
- Obesity
- Lack of physical exercise

There are many different types of heart disease. Some are congenital, but a majority of heart disease develop over the time and affect people later in life. Some of most common heart diseases are listed below (K. Sudhakar and M. Manimekalai,2014):

- Coronary heart disease
- Angina pectoris
- Congestive heart failure
- Cardiomyopathy
- Congenital heart disease
- Arrhythmias
- Myocarditis

In this work has been highlighted four types of heart diseases to be diagnosed namely (Coronary Heart Disease, Angina Pectoris, Congestive Heart Failure, and Arrhythmias).

3.1 The Online Heart Disease Data Set

The Cleveland database is exceptionally popular and has been extensively employ as a standard for heart illness identification system. The database contains of 303 instances with 13 medical factors that are obtained from UCI data set repository which was donated by Detrano (A. Asuncion and D. J. Newman,2007). Table (1) shows the 13 attributes description of the Cleveland data set. There are five attributes which are (Age, Trestbps, Chol, Thalach, and Oldpeak) have large magnitude compared with other attributes used in training set of classifier, as illustrated in Table (2) Thal has maximum value 7, hence the five attributes are normalized to be in the same level with Thal. Table (3) shows the distribution of UCI data disease records and the methods used to design the system. As demonstrated, two fundamental outputs are recognized where the value H0 denote normal heart disease and values H1, H2, H3 and H4 denote heart is abnormal.

Table 2 13 attributes description of the Cleveland data set

Age				
Sex	Male "0"		Female "1"	
CP: Chest pain type	typical angina "1"	atypical angina "2"	non-angina pain "3"	Asymptomatic "4"
Trestbps : Resting blood pressure in mm Hg.				
Chol: Serum cholesterol in mg/dl.				
Fbs: Indicator of whether fasting blood sugar was > 120 mg/dl				
Restecg : Resting electrocardiographic results	Normal "0"	ST-T wave abnormality "1"	probable or definite left ventricular hypertrophy "2"	
Thalach: Maximum heart rate achieved.				
Exang :Indicator of whether the angina is exercise induced	Yes "1"		No "0"	
Oldpeak: ST depression induced by exercise relative to rest.				
Slope: The slope of the peak exercise ST segment	up sloping "1"	Flat "2"	down sloping "3"	
Ca: Number of major vessels colored by fluoroscopy.				
Thal :Summary of heart condition	Normal "3"	fixed defect "6"	reversible defect "7"	

Table 3: Description of Cleveland Data set

Disease diagnosis	Classification	No .of records
H0	Normal	164
H1	Abnormal	55
H2		36
H3		35
H4		13
Total		303

3.2 Collected Data Set (CD)

Countless sources of info must be considered amid the analysis, in order to diagnosis the heart disease with high accuracy. So that the doctor depended on all symptoms are recorded, patient answers to questions, in additional to medical examination and laboratory results. UCI data include sufficient medical factors for the diagnosis of heart disease, so the data collected from Ibn al-Bitar hospital and Baghdad medical city based on those medical factors. However, some medical factors such as (Maximum heart rate, ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, and Number of major vessels colored by fluoroscopy) taken by exposing the patient to a particular effort has encountered a great difficulty to collect these factors. Therefore, these factors are replaced by the cardiologists into medical factors which are (heart rate, family history of HD, smoking, Echo finding for hypokinesia, and Previous attack of angina). Modified medical factors take into consideration the risk factor, family history, the possibility of previous angina in addition to recorded of echo in order to get sufficient medical factors as show in the table (4). These factors transforming into numerical representation in order to building diagnosis system.

There are three attributes which are (Age, BP, and HR) have large magnitude compared with other attributes used in training set of classifier, as illustrated in Table (4) CP has maximum value 4, hence the three attributes are normalized to be in the same level with CP. The heart disease gathered database utilized for testing and preparing the framework comprises of 200 cases accumulated from the Ibn al-Bitar doctor's facility and Baghdad medicinal city. Those data included four class of heart disease in addition to normal class. Table (5) show the distribution of the five classes of heart diseases in the database.

Table 4: Collected Data set (CD)

<table>
<tr><td>Age</td><td colspan="4"></td></tr>
<tr><td>Sex</td><td colspan="2">Male
"0"</td><td colspan="2">Female
"1"</td></tr>
<tr><td>CP: Chest pain type</td><td>typical angina
"1"</td><td>atypical angina
"2"</td><td>non-angina pain
"3"</td><td>Asymptomatic
"4"</td></tr>
<tr><td>BP :Resting blood pressure in mm Hg.</td><td colspan="4"></td></tr>
<tr><td>CHOL:Serum cholesterol >=240 mg/dl</td><td colspan="2">Normal
"0"</td><td colspan="2">Abnormal
"1"</td></tr>
<tr><td>FBS:fasting blood sugar was > 120 mg/dl</td><td colspan="2">No
"0"</td><td colspan="2">Yes
"1"</td></tr>
<tr><td>ECG:Resting electrocardiographic results</td><td>Normal
"0"</td><td>ST-T wave abnormality
"1"</td><td colspan="2">probable or definite left ventricular hypertrophy
"2"</td></tr>
<tr><td>HR:heart rate achieved.</td><td colspan="4"></td></tr>
<tr><td>EXANG:the angina is exercise induce</td><td colspan="2">No
"0"</td><td colspan="2">Yes
"1"</td></tr>
<tr><td>FH:Previous family history of HD</td><td colspan="2">No
"0"</td><td colspan="2">Yes
"1"</td></tr>
<tr><td>SM:smoking</td><td colspan="2">No
"0"</td><td colspan="2">Yes
"1"</td></tr>
<tr><td>HYP:Echo finding for hypokinesia</td><td colspan="2">No
"0"</td><td colspan="2">Yes
"1"</td></tr>
<tr><td>PERANG:Previous attack of angina</td><td colspan="2">No (negative)
"0"</td><td colspan="2">Yes (positive)
"1"</td></tr>
</table>

3.3 Division of data set

In this work, two type of dataset are used online dataset and CD, each dataset consists of 13 medical factors necessary to the diagnosis the heart diseases. These factors are extracted from the heart diseases data set and can be split into three categories: (i) the basic information of a patient such as (Age, Sex, Family history) (ii) the symptoms such as (Chest pain) (iii) the physical examinations and lab results such as (Blood pressure, Electrocardiogram, Echo, etc.). Large part of database records is used for training data and rest of them for testing data. Many divisions are employ for online database in order to be able to compared the performance of the proposed heart disease classifiers with previous works. These data divisions are illustrating in Table (6). While CD are divided into 50% of data for training and 50% for testing.

Table 6: Pervious works data division of UCI database

References	Training data	Testing data
S. Bhatia et.al,2008	250	53
M. Gudadhe et.al. 2010	200	103
A. Khemphila and V.Boonjing,2011	182	121
M. A. M. Abusharian *et al.*, 2014	330	83
X. Liu *et al* .2017	212	91

4. Algorithms of Heart Disease Diagnosis System

Many algorithms are adopted for classification of heart disease. In this proposed work MLP, and SVM are used to diagnosis the heart disease. The Performance of each classification techniques is compared in terms of testing time, sensitivity, specificity, and accuracy of the system. The sensitivity measures the proportion of actual positives which are identified correctly. It diagnoses people living with heart disease correctly as heart disease positive. The sensitivity can be calculated as in Eq.(1)(J. Kuruvilla and K. Gunavathi,2014.)

$$\text{Sensitivity} = \frac{T_P}{T_P+F_N} \tag{1}$$

where:
T_P (True Positive): the number of samples that have heart disease and properly diagnosed
F_N (False Negative): the number of samples that were sick but healthy wrongly diagnosed
Eq.(2) measured proportion of people who are correctly identified as not having heart disease, i.e. A condition where a set of healthy people is correctly identified as not having the heart disease (Sunila *et al*,2012).

$$\text{Specificity} = \frac{T_N}{T_N+F_P} \tag{2}$$

where:

F_P (False Positive): the number of samples that are healthy but sick wrong diagnoses.

In addition to sensitivity and specificity, the performance can also be measured with Positive Prediction Value (PPV) which determined how well a positive test result actually determines that the disease is present. As illustrate in Eq.(3).

$$PPV = \frac{T_P}{T_P+F_P} \times 100 \tag{3}$$

Eq.4), measure the Negative Prediction Value (NPV), which determined how well a negative test result actually determines that the disease is absent (E. O. Olaniyi *et al*,2015).

$$NPV = \frac{T_N}{T_N+F_N} \times 100 \tag{4}$$

where:

T_N (True Negative): the number of samples that are healthy and properly diagnosed

Accuracy is a statistical measure of how well a classifier correctly identifies or excludes a condition. The accuracy is the proportion of true results (both true positive and true negative) in the population, it can calculated as in Eq.(5)(J. Kuruvilla and K. Gunavathi,2015).

$$\text{Accuracy} = \frac{(T_P+T_N)}{(T_P+T_N+F_N+F_P)} \quad (5)$$

4.1The multilayer perceptron (MLP)

MLP is one of the most frequently used neural network architectures in Medical Decision Support System (MDSS), and it belongs to the class of supervised neural networks. A typical MLP network consists of three or more layers of processing nodes: an input layer that receives external inputs, one or more hidden layers, and an output layer which produces the classification results Fig.(2).

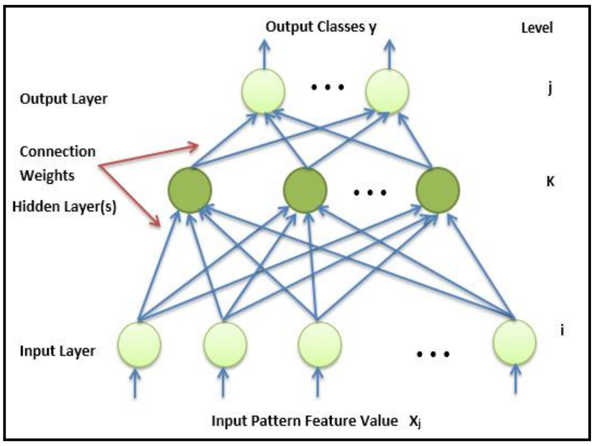

Figure2: Structure of a multilayer perceptron network.

The fundamental of the neural network is that when data are accessible at the input layer, the network neurons carry out calculations in the hidden layers until an output value is obtained at each of the output neurons. This output indication should be able to point out the suitable class for the input data. That is, one can expect to have a high output value on the right class neuron and low output values on all the rest. A node in MLP can be modeled as an artificial neuron as shown in Fig.(3), which computes the weighted sum of the inputs at the presence of the bias, and passes this sum through the activation function. The whole process is defined as follows (H. Yan *et al*,2006):

$$\text{Vj} = \sum_{i=1}^{P} W_{ji}X_i + \theta_j \quad (6)$$

$$y_j = f_j\,(V_j)$$

Where

V_j: The linear combination of inputs X_1; X_2. X_p
θ_j: The bias
W_{ji}: The connection weight between the input X_i and the neuron j
f_j (.): The activation function of the j the neuron
y_j: The output.
The sigmoid function is a common choice of the activation function, as defined in Eq. (7).

$$\text{F}(\alpha) = \frac{1}{1+e^{-\alpha}} \quad (7)$$

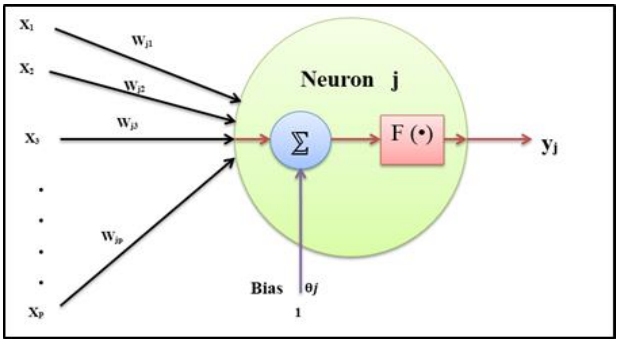

Figure 3: One node of MLP: an artificial neuron

Any network consists of training phase and testing phase. The reason of training phase is to update the weights value of the network based on the learning algorithm. The weights of the neural network are updated in a supervised mode using the most common algorithm known as the Back Propagation (BP) algorithm and according to Eq.(8)

$$\text{Wji}\,(\text{t}+1) = \text{Wji}\,(\text{t}) - \varepsilon \frac{\partial E_f}{\partial \text{Wji}}\,(t) \quad (8)$$

Where
ε : is the learning rate.
E_f : is the error function.
In the learning phase, the output of the feed forward neural network is computed for each input training pattern. The error between the computed output and desired output is used to update the weight of the network by back propagation algorithm (M. Azarbad *et al*,2012).

4.2 Support vector machine (SVM)

Support Vector Machines is a class of supervised learning algorithms proposed by Vladimir Vapnik within the area of statistical learning theory and structural risk minimization, have demonstrated to work successfully on various classification and forecasting problems. SVMs have been used in many pattern recognition and regression estimation problems and have been applied in medical diagnosis for diseases classification (J. Nayak *et al*,2015). The main goal in support vector machine is to have a maximum margin between different samples and keep the margin separated as illustrate in Fig.(4) (S. Bashir *et al*,2014). SVM consists of two types of classification linear and non-linear classifications (F. Melgani and L. Bruzzone,2004). Linear SVM find the Optimal Hyper plane for patterns[S. U. Ghumbre and A. A. Ghatol,2012]: Consider the training sample $\{(x_i, y_i)\}^N i = 1$ where x_i is the input pattern for the ith instance and y_i is the corresponding target output. With pattern represented by the subset y_i= +1 and the pattern represented by the subset y_i= -1 are linearly separable. The equation in the form of a hyper plane that does the separation is

$$W^T X + b = 0 \tag{9}$$

Where x is an input vector, w is an adjustable weight vector, and b is a bias.

$$W^T X_i + b \geq 0 \quad \text{For } y_i = +1 \tag{10}$$

$$W^T X_i + b < 0 \quad \text{For } y_i = -1 \tag{11}$$

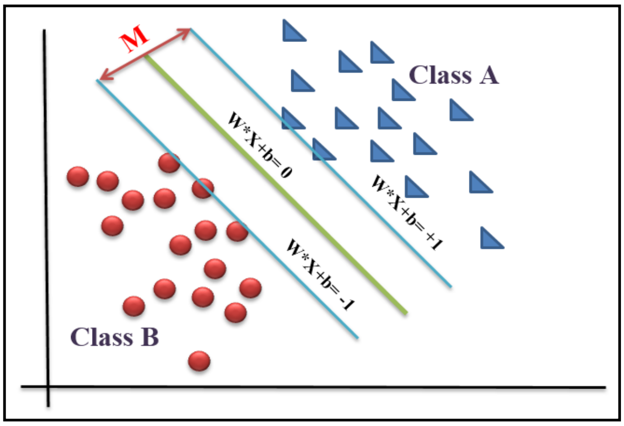

Figure 4: Support Vector Machine

The discriminant function gives an algebraic measure of the distance from x to the optimal hyperplane for the optimum values of the weight vector and bias, respectively.

$$g(x) = W_O^T \text{ X} + b_0 \tag{12}$$

SVM is capable of finding non-linear decision boundaries by using kernel functions there are two main steps [C.-W. Hsu and C.-J. Lin,2002]. In the first step the original input data is transformed into a higher dimensional space using kernel function (polynomial, Gaussian, radial basis function, exponential radial function, multi-layer perceptron, ex...). The second step searches for a linear separating hyper plane in the new space [F. Melgani and L. Bruzzone,2004]. As shown in Fig.(5).

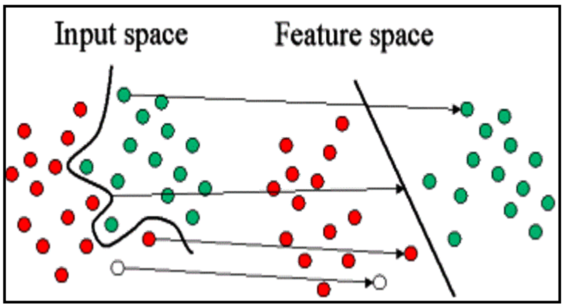

Figure 5: Nonlinear Support Vector Machine

5. The Proposed Heart Disease Diagnosis System

Diagnosis of heart disease using intelligent systems attracted the attention of researchers. The proposed systems implemented using many techniques Such as: (MLP, and SVM). It based on several databases of patients to obtain a high precision diagnosis.

5.1 Diagnosis System Based On MLP (Proposed System I)

The proposed MLP network consists of three layers input layer, hidden layer, and output layer. The input layer consists of 13 neurons that represent 13 medical factors of patients which applied to the neural network as source nodes. The number of hidden layer and the quantity of nodes in every layer are more significant parameters, which were experiment to decide the most excellent number of nodes that will have the capacity to train the pattern of the input data and provide the better accuracy for classification. In this work, two hidden layer are used with best number of neuron (5 neurons in each layers). The output layer consists of two neurons, which indicates the patient case. They represent one of neurons for normal patient and other neuron for abnormal patient as shown in Fig.(6).

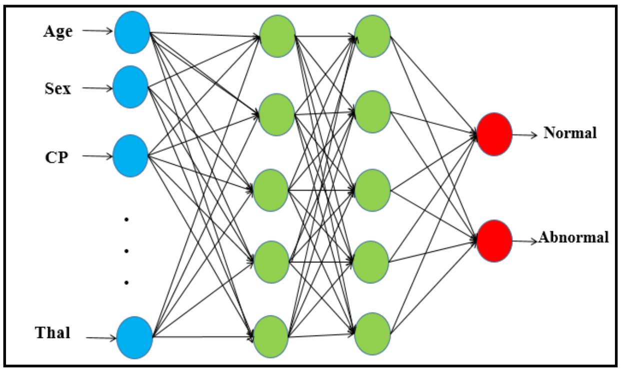

Figure 6: Structure of proposed MLP.

The neural network is learned by back propagation supervised algorithm. Tan-sigmoid is the activation function of the hidden layers, while the linear transfer function is used for output layer. A Back Propagation algorithm (Levenberg-Marquardt back propagation) is implemented for the learning of this network as show in Fig.(7).

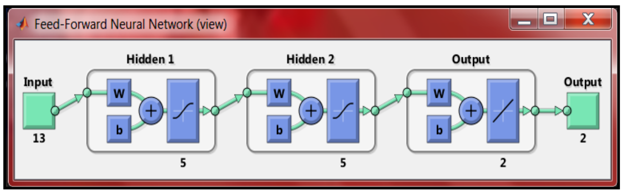

Figure 7: Two class MLP network

Figures ((8), (9), (10), (11), and (12)). Shows the training performance of the two class classification based on UCI database according to the division database shown in Table (6). As demonstrated from these figures the Mean Square Error performance (MSE) is less than 1×10^{-6} for all proposed classifiers. MSE performance of two class classification based on CD reached to 2.8205×10^{-5} as greatest performance at 21 epochs as illustrate in Fig.(13). The success classification rate is100.0% when the system is experienced using the similar learning database.

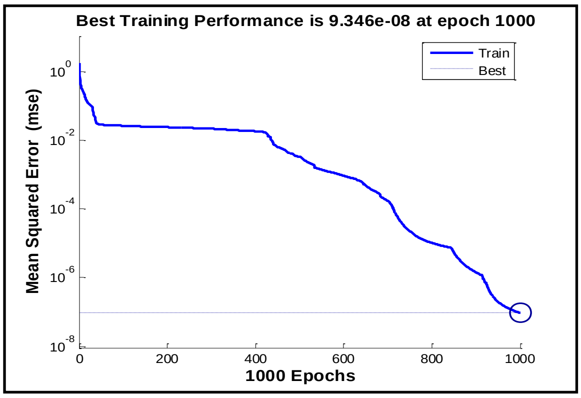

Figure 8: Training performance of MLP for 220 training samples of UCI

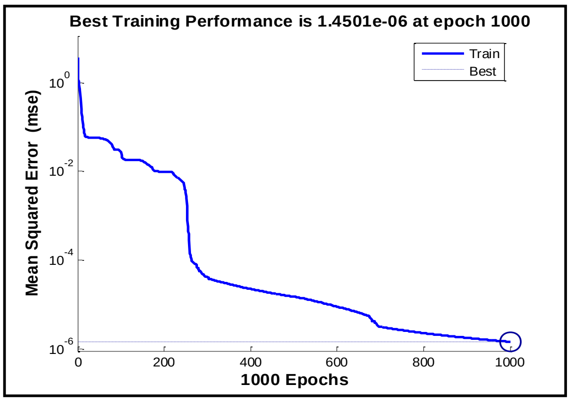

Figure 9: Training performance of MLP for 212 training samples of UCI

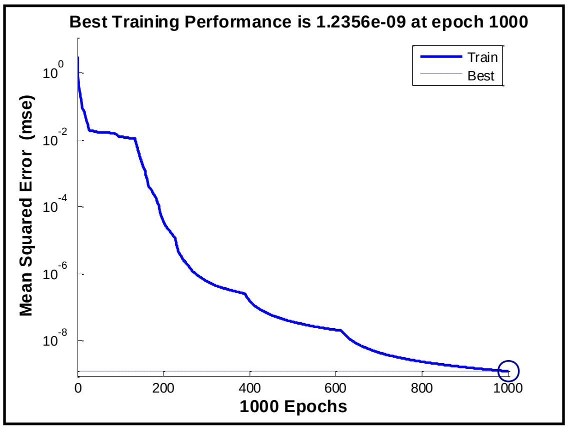

Figure 10: Training performance of MLP for 200 training samples of UCI

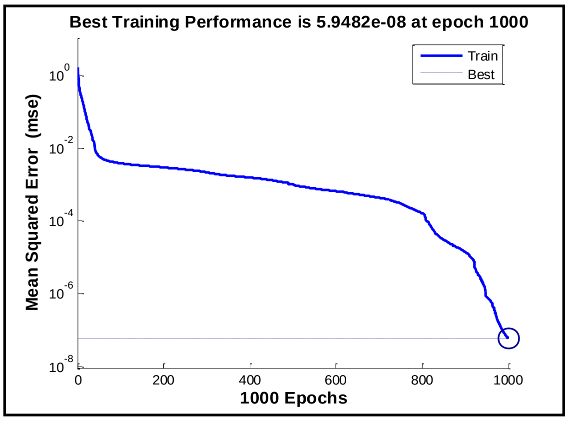

Figure 11: Training performance of MLP for 182 training samples of UCI

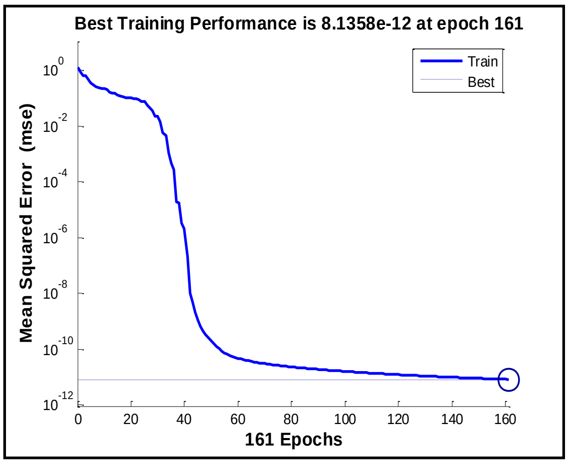

Figure 12: Training performance of MLP for 250 training samples of UCI

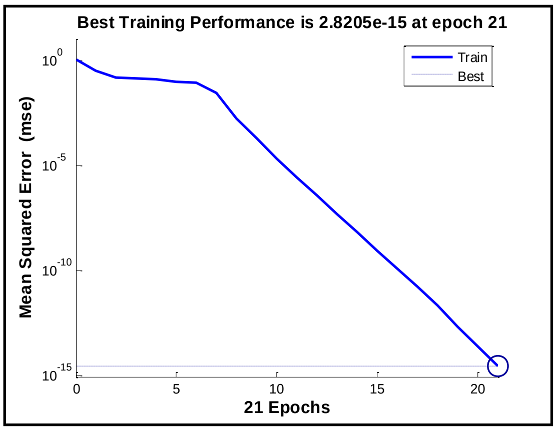

Figure 13: Training performance of MLP for 100 training samples of CD

The Diagnosis System Based on MLP for 5 class classification consists of The input layer with 13 neurons, two hidden layer with 5 neurons in each layers, and the output layer that have 5 neurons. The output neurons point to the objective of the system. They stand for one of neurons for normal patient and other neurons for different heart disease of abnormal patient as shown in Fig.(14).

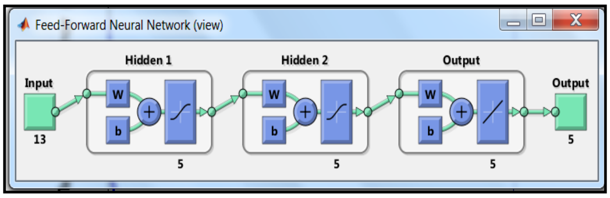

Figure 14: Five class MLP neural network

As previously mentioned the CD consists of four type of heart disease in additional to normal case. In the multiclass classification the output of the proposed system represented in to "0" for normal patient, "1" represented to Coronary heart disease, "2" represented to Angina pectoris, ”3"represented to Congestive heart failure, and "4" represented to Arrhythmias, finally only one result output indicated to the patient's case. The performance result for 5 class classification shown in the Fig.(15). As demonstrated the best training performance is 1.0049×10^{-8} at epoch 438.

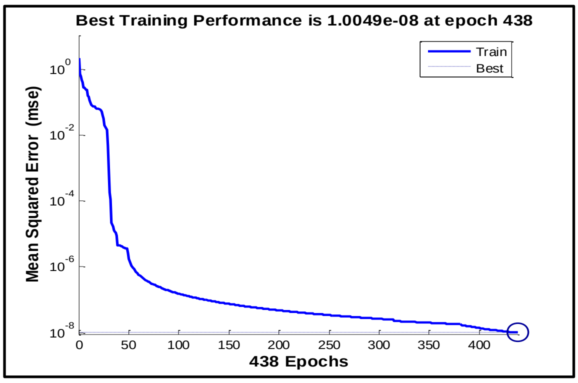

Figure 15: Training performance of MLP for 5 class 100 training samples of CD

5.2 Diagnosis system based on SVM (proposed system II)

SVM has been applied to the auxiliary diagnosis of heart disease. That is to get the separating hyperplane of both heart disease patients and healthy people. Although the idea of SVM was initially suggested to binary recognition, varieties have been proposed to employ SVM for multiclass problem as well. The most popular methods for multiclass classification problems are one against one (OVO) and one against all (OVA). The first is includes the construction two binary classifier, single for every pair of whole classes, the last class is evaluated a pre-defined selection mechanism. The next one versus all (OVA) technique, there is a binary classifier for every class to split the members of that class from all other classes. MATLAB program was written for multiclass SVM classifier by using OVA method for the training process. The linear kernel function is used to evaluate the system accuracy. The system was tested on the same training set. the training accuracy and training time of the system based on division of training data illustrate in Table (7).

Table 7: SVM training results

No. of Training Data	No. of Class	Training Time	Training Accuracy
250	2	0.1744 sec	84.4%
200	2	0.1561 sec	85%
182	2	0.1541 sec	81.319%
220	2	0.1556 sec	83.636%
212	2	0.323 sec	82.076%
100	2	0.0825 sec	95%
	5	0.161 sec	74%

6. Performance Result

6.1 Performance result of heart disease diagnosis system based on MLP (proposed system I)

The performance result of two class diagnosis system based on MLP according to UCI data base shown in Table (8). As illustrate the high accuracy for diagnosis heart disease when using 27% of UCI data (n=83 samples) and 17% of UCI data (n=53 samples) which gives accuracy rate of (91.566% and 90.57%) respectively. The reason of high accuracy is large number of samples are used to training the network, while a few samples are used to testing the network, for this reason the lower accuracy diagnosis is 81.818% when using 40% of data (n=121samples) for testing the system. The higher ability of MLP to predict the patient with heart disease when Sensitivity is 0.947. Furthermore, the high Sensitivity to predict the patient without heart disease is 0. 744.The result of normal and abnormal heart disease diagnosis system based on MLP according to CD, when using 50% of data samples for testing shown in Table (9). As compared with performance result of UCI data, CD gives highest accuracy 98% in additional to Sensitivity, specificity, NPV, and PPV.

Table 8: Performance of two class diagnosis system based on MLP according to UCI database

Testing Data	T_N / F_P	F_N / T_P	Test time Sec	Specificity	Sensitivity	NPV%	PPV%	Accuracy%
53	25	2	0.718	0.893	0.92	92.59	88.46	90.57
	3	23						
103	53	11	1.223	0.883	0.744	82.81	82.05	82.52
	7	32						
121	63	12	1.811	0.863	0.75	84	78.26	81.82
	10	36						
83	40	2	1.264	0.889	0.947	95.24	87.81	91.57
	5	36						
91	55	7	2.989	0.917	0.774	88.71	82.76	86.81
	5	24						

Table 9: Performance of two class diagnosis system based on MLP according to CD

Testing Data	T_N / F_P	F_N / T_P	Test time Sec	Specificity	Sensitivity	NPV%	PPV%	Accuracy%
100	50	0	1.52	0.96	1	100	96	98
	2	48						

The ability of the proposed system to diagnosis heart disease for multiclass classification is illustrated in Table (10). This Table shows the performance result of the system according to used 17% of UCI database and 50% of CD to testing the system. As clearly seem the highest multiclass accuracy is 81% according to CD, it is higher than multiclass accuracy of proposed system based on UCI data which is 60.377%.

Table 10: Performance of five class diagnosis system based on MLP

Testing Data	T_N / F_P	F_N / T_P	Test time Sec	Specificity	Sensitivity	NPV%	PPV%	Accuracy%
100	49	1	1.517	0.875	0.969	98	82.05	81
	7	32						

6.2 *Performance result of heart disease diagnosis system based on SVM (proposed system II)*

The performance result of proposed system II prove that the SVM is considered as good algorithm for diagnosis heart disease into two classes as shown in Table (11). The result show the proposed system can achieve very high accuracy rate (>90%) expect the diagnosis accuracy when using 34% of UCI database (n=103 samples) for testing the system which is 82.524%, this is due to random selective of the training and testing samples. As it seems clearly the highest accuracy is 93.407% in additional to high specificity, and PPV at small testing time is 0.06 sec when using 30% of UCI database (n=91 samples) for testing, while maximum sensitivity and NPV are achieved 100% at 53 samples used for testing the system.

Table 11: Performance of two class diagnosis system based on SVM according to UCI database

Testing Data	T_N / F_P	F_N / T_P	Test time Sec	Specificity	Sensitivity	NPV%	PPV%	Accuracy%
53	27	0	0.039	0.844	1	100	80.77	90.57
	5	21						
103	51	13	0.073	0.911	0.723	79.69	87.18	82.52
	5	34						
121	70	5	0.09	0.946	0.894	93.33	91.3	92.56
	4	42						
83	38	4	0.079	0.905	0.902	90.48	90.25	90.36
	4	37						
91	58	4	0.061	0.967	0.871	93.55	93.1	93.41
	2	27						

The performance of proposed system II when using 50% of CD database is give the maximum accuracy 96% in additional to all performance parameters result as illustrated in Table (12).

Table 12: Performance of two class diagnosis system based on SVM according to CD

Testing Data	T_N / F_P	F_N / T_P	Test time Sec	Specificity	Sensitivity	NPV%	PPV%	Accuracy %
100	46	4	0.09	1	0.93	92	100	96
	0	50						

As demonstrated in Table (13) the SVM is bad algorithm for analysis multiclass classification of heart disease which gets 60% classification accuracy depending on CD.

Table 13: Performance of five class diagnosis system based on SVM

Testing Data	T_N / F_P	F_N / T_P	Test time Sec	Specificity	Sensitivity	NPV%	PPV%	Accuracy%
100	33	7	0.1	1	0.8	83	96	60
	1	27						

7. Comparison Result of Heart Disease Proposed Systems

The proposed classifiers MLP, and SVM are compared based on testing time, specificity, sensitivity, NPV%, PPV%, and accuracy of the result obtained. The performance comparison results of the classifiers when 250 samples of UCI data are used for training and 53 samples of data for testing is shown in Fig.(16).

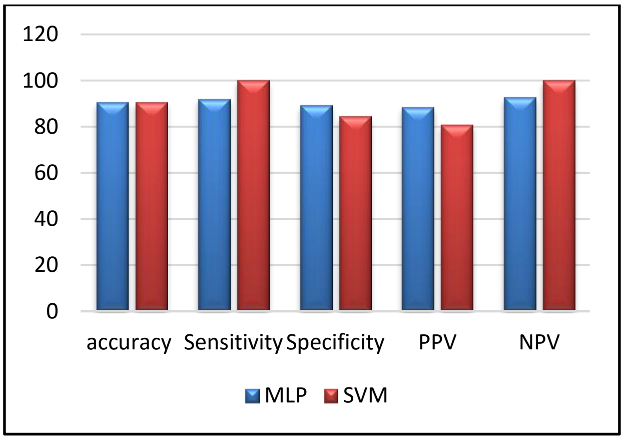

Figure 16: The performance comparison of 2 class heart disease at 53 testing samples of UCI

The high sensitivity is more important to identify a present of disease so that sensitivity of SVM is 100% which are higher than the sensitivity of MLP. While the PPV of MLP is higher than SVM. Furthermore, the SVM is very good for determine the disease when the test is negative at NPV equal 100%. But the specificity of MLP is comparatively high compared with SVM classifier which is 90%. In general, the accuracy of MLP, and SVM are the same value (90.57%).

Fig.(17) show the performance comparison of the classifiers when 200 samples of UCI data are used for training classifiers and 103 instances are used for testing the accuracy of the proposed system. As illustrated from this figure, the accuracy of MLP and SVM is 82.524 % in additional to that MLP has good sensitivity and NPV.

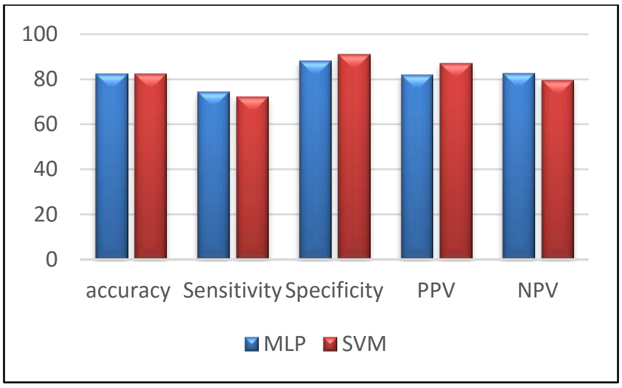

Figure 17: The performance comparison of 2 class heart disease at 103 testing samples of UCI

The performance comparison results of the classifiers when 121 instances are used for testing the classifier is shown in Fig.(18).

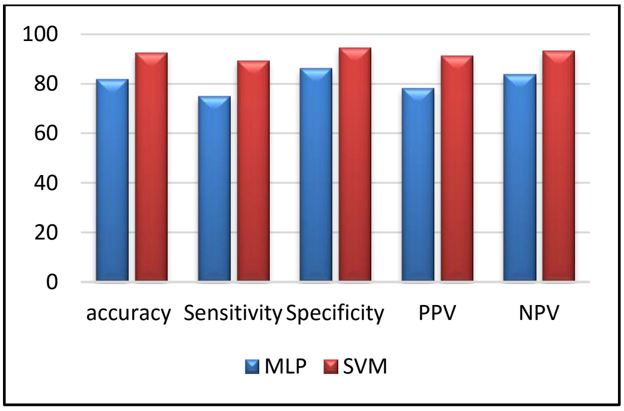

Figure 18: The performance comparison of 2class heart disease at 121 testing samples of UCI

As demonstrated, the SVM is the best classifier to diagnosis disease which gives high accuracy (92.562%) at good testing time (0.09sec) which is slightly greater than the MLP testing time (1.811sec). In additional to that, SVM can measured quantity of populace who properly recognized as not have heart illness when the specificity equal to 94.6%. Furthermore, the sensitivity of SVM is higher than other classifier. Also, it is good to determines how appositive test in fact determine that the illness is present (PPV 91.304%) and how negative check really determine that the sickness is absent (NPV 93.333%).

Figure (19) show the performance comparison of the proposed systems to diagnosis heart disease when using 83 samples of UCI data for testing. It is clearly seen, the SVM reach to accuracy (90.361%) but MLP has higher accuracy (91. 566%).In general, all classifiers can consider good classifiers to diagnosis heart disease and the accuracy result have only different (1.205). In spite of, the small difference, MLP is best classifier for diagnosis heart disease.

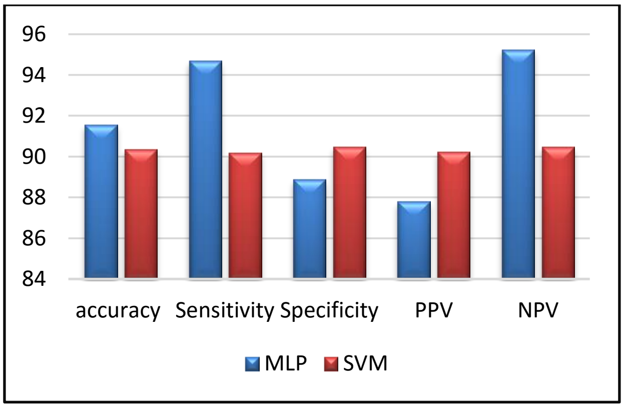

Figure 19: The performance comparison of 2 class heart disease at 83 testing samples of UCI

When using 91 instances of UCI data for testing the systems, the performance comparison result of the classifiers for two class heart disease is shown in Fig.(20). From this figure, the SVM can considered very good classifier to diagnosis heart disease which give maximum accuracy (93.407%) in 0.061 sec with high sensitivity, specificity, PPV, and NPV.

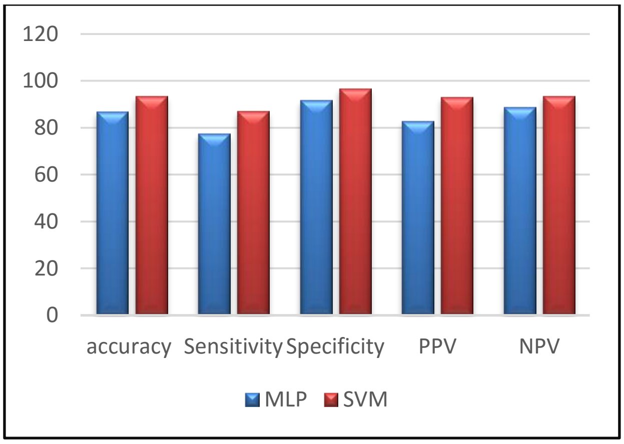

Figure 20: The performance comparison of 2 class heart disease at 91 testing samples of UCI

The performance comparison of the proposed systems to diagnosis normal and abnormal patient from heart disease by using 50% of CD for test the system is show in Fig.(21). As demonstrated, all classifiers have very good performance, it is clearly the MLP has maximum accuracy which is 98%.

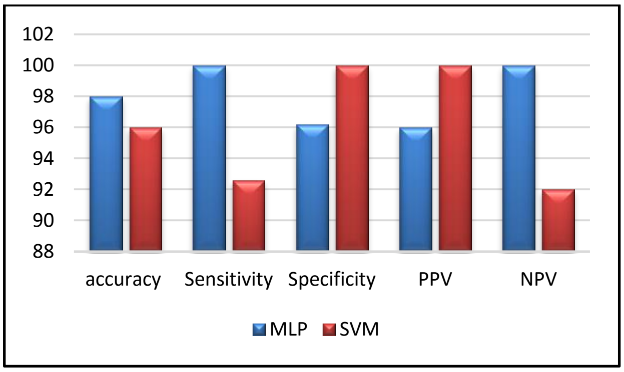

Figure 21: The performance comparison of 2 class heart disease at 100 testing samples of CD.

As mentioned the collected data (CD) included four type of heart disease in additional to normal case. In the multiclass classification the output of the proposed system represented in to **"0"** for normal patient, **"1"** represented to Coronary heart disease, **"2"** represented to Angina pectoris,**"3"** represented to Congestive heart failure, and **"4"** represented to Arrhythmias, finally only one result output indicated to the patient's case. So that, the performance comparison of the classifiers for 5 class classification is shown in the Fig.(22). The result of the proposed systems show the MLP is good classifier to predicate heart disease while the SVM is bad classifier for get accuracy 90. MLP is considered as good accuracy to multiclass classification. But SVM can consider as bad algorithm for multiclass heart disease which has accuracy reached to 60%.

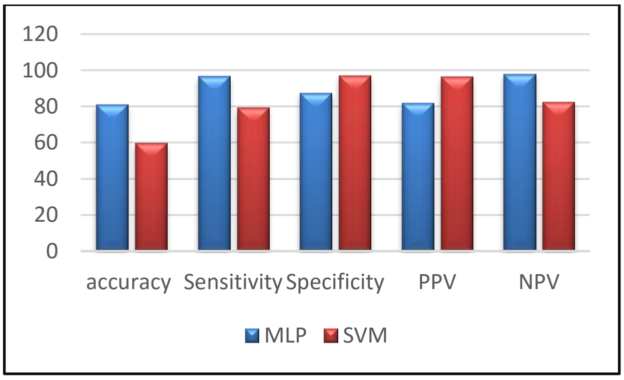

Figure 22: The performance comparison of five class heart disease at 100 testing samples of CD.

8. Performance Comparison with Previous Work

A lot of studies have been made to build the precision of finding heart illness. This investigation has been assessed utilizing the particular database information from UCI dataset repository which was given by Detrano. The correctness of each along with compare those with consequences of this work is shown in Table (14).

GA and SVM are adopted in [S. Bhatia et.al,2008] to diagnosis heart disease into two class according to division of UCI database which are 250 samples for training and 53 samples for testing, the accuracy of these system is 90.57%. When compared it with accuracy result of this work, the MLP, and SVM have the same accuracy of pervious work. When using 200 samples of UCI data for training and 103 samples of data for testing, M. Gudadhe et. al. 2010[10] reached to diagnosis accuracy 80.4% by using diagnosis system based on SVM. While in this work all proposed system reached to the higher accuracy (82.524%) based on the same division of data. The MLP classifier in the diagnosis system in [A. Khemphila and V. Boonjing,2011] used 60% of UCI data for training and 40% of data for testing, the accuracy result of this system is 80.41%. This accuracy is smaller compared with the accuracy of the proposed system based on SVM that give maximum accuracy (92.562%). M.A.M. Abusharian *et al*., 2014 adopted two approaches to diagnosis heart disease ANN and ANFSI and the maximum accuracy reached to 87.04% of ANN, when compared this result with the proposed system of this work, the MLP has better accuracy (91.566%) SVM has accuracy (90.361%). The hybrid classification system in [X. Liu *et al* .2017] based on RFRS and C4.5 are used to diagnosis heart disease into two classes according to division of UCI database. The accuracy of this system is 92.59%, which is smaller than the accuracy result of this work based on SVM that give maximum accuracy which is 93.4075.

Table 14: Comparison result with previous work of two class heart disease

References	Algorithm	Training data	Testing data	Accuracy %	Present work Accuracy %
S. Bhatia *et al*,2008	GA+SVM	250	53	90.57	MLP 90.57
					SVM 90.57
M. Gudadhe *et al.* 2010	SVM	200	103	80.41	MLP 82.524
					SVM 82.524
A. Khemphila and V. Boonjing,2011	MLP	182	121	80.17	MLP 81.818
					SVM 92.562
M. A. M. Abusharian *et al.*, 2014	ANN	220	83	87.04	MLP 91.566
	ANFIS			75.93	SVM 90.361
X. Liu *et al* .2017	RFRS+C4.5	212	91	92.59	MLP 86.813
					SVM 93.407

Conclusion

The diagnosis system is proposed in this work to assist physicians to diagnosis heart condition by converting medical factors of the patients into numerical representation. The simulation results show that, the proposed MLP classifier has 98% accuracy of two heart diseases classification when the performance of this classifier was evaluated using collected database. While the performance for SVM classifier is reached 96%. Also, MLP has overcome from SVM classifier when classify four type of heart disease in additional to normal case for accuracy reached to 81%. According to UCI database the MLP and SVM is considered best classifier for classify heart disease. Furthermore, a reasonable comparison is completed among the two suggested system and with the most excellent past work systems under similar conditions. The comparison unmistakably demonstrated that the best two proposed recognizers (MLP and SVM) give higher grouping exactness when compared and the best recognizer of past works under a similar learning and testing database division.

References

D. Vadicherla and S. Sonawane. (2013) Decision Support System for Heart Disease Based On Sequential Minimal Optimization in Support Vector Machine, International Journal of Engineering Sciences and Emerging Technologies, 4 (2), pp. 19-26.

S. U. Ghumbre and A. A. Ghatol. (2012). Heart Disease Diagnosis Using Machine Learning Algorithm, in *Proceedings of the International Conference on Information Systems Design and Intelligent Applications*, India.

S. Ergin and O. Kilinc, (2014). A new feature extraction framework based on wavelets for breast cancer diagnosis, *Computers in Biology and Medicine,* 51, pp. 171-182.

M. S. Kidwai and A. M. Khan. (2014). An Efficient Algorithm for The Diagnosis of Alzheimer Disease by Using MATLAB, *International Journal of Emerging Trends and Technology in Computer Science,* 3 (2), pp. 297_298.

P. Sharma, D. Singh, M. K. Bandil and N. Mishra. (2013), Decision Support System for Malaria and Dengue Disease Diagnosis (DSSMD)," *International Journal of Information and Computation Technology,* 3 (7), pp. 633-640.

D. Vadicherla and S. Sonawane. (2013), Classification of Heart Disease Using SVM and ANN, *International Journal of Research in Computer and Communication Technology,* 2 (9), pp. 694-701.

M. G. Feshki and O. S. Shijani, "Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network," in *Artificial Intelligence and Robotics (IRANOPEN)*, Qazvin, Iran, 2016.

J. S. Sonawane, D. R. Patil and V. S. Thakare. (2013), Survey on Decision Support System for Heart Disease, *International Journal of Advancements in Technology,* 4 (1), pp. 89-96.

S. Bhatia, P. Prakash and G. N. Pillai. (2008), SVM Based Decision Support System for Heart Disease Classification with Integer-Coded Genetic Algorithm to Select Critical Features, in *Proceedings of the World Congress on Engineering and Computer Science*, San Francisco, USA.

M. Gudadhe, K. Wankhade and S. Dongre. (2010), Decision Support System for Heart Disease based on Support Vector Machine and Artificial Neural Network," in *International Conference on Computer and Communication Technology (ICCCT)*, Allahabad, Uttar Pradesh, India.

A. Khemphila and V. Boonjing. (2011), Heart disease Classification using Neural Network and Feature Selection, in *International Conference on Systems Engineering*, Las Vegas, NV, USA.

M. A. M. Abushariah, A. A. M. Alqudah, O. Y. Adwan and R. M. M. Yousef. (2014), Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches, *Journal of Software Engineering and Applications,* 7 (12), pp. 1055-1064.

E. O. Olaniyi, O. K. Oyedotun and K. Adnan. (2015), Heart Diseases Diagnosis Using Neural Networks Arbitration Arbitration, *international journal of Intelligent Systems and Applications,* 7 (12), pp. 75-82.

X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang and Q. Wang. (2017), A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method, *Computational and Mathematical Methods in Medicine, p*p. 11.

K. Khoshraftar, A. V. Roudsari, S. Omidfar and P. Turkmen. (2016), The study of data mining issue in the medical field, in *International Conference on Research in Science and Technology*, London_England.

B. Kaur and W. Singh. (2014), Review on Heart Disease Prediction System using Data Mining Techniques, *International Journal on Recent and Innovation Trends in Computing and Communication,* 2 (10), pp. 3003_3008.

K. Sudhakar and M. Manimekalai. (2007), Study of Heart Disease Prediction using Data Mining, *International Journal of Advanced Research in Computer Science and Software Engineering,* 4 (1), pp. 1157-1160

A. Asuncion and D. J. Newman, "UCI Machine Learning Repository," 2007. [Online]. Available: http:// www.ics.uci.edu/ $\sim$mlearn/ {MLR}epository.html.

J. Kuruvilla and K. Gunavathi. (2014), Lung cancer classification using neural networksfor CT images, *Computer Methods and Programs in Biomedicine,* 113 (1), pp. 202–209.

Sunila, P. Panday and N. Godara. (2012), Decision Support System for Cardiovascular Heart Disease Diagnosis using Improved Multilayer Perceptron, *International Journal of Computer Applications,* 45 (8), pp. 12-20.

J. Kuruvilla and K. Gunavathi. (2015), Lung cancer classification using fuzzy logic for CT images, *International Journal Medical Engineering and Informatics,* 7 (3), pp. 233–249.

H. Yan, Y. Jiang, J. Zheng, C. Peng and Q. Li. (2006), A multilayer perceptron-based medical decision support system for heart disease diagnosis, *Expert Systems with Applications,* 30 (2), pp. 272–281.

M. Azarbad, S. Hakimi and A. Ebrahimzadeh. (2012), Automatic Recognition of Digital Communication Signal, *International Journal of Energy, Information and Communications,* 3 (4), pp. 21-34.

J. Nayak, B. Naik and H. S. Behera. (2015), A Comprehensive Survey on Support Vector Machine in Data Mining Tasks: Applications and Challenges, *International Journal of Database Theory and Application,* 8 (1), pp. 169_186.

S. Bashir, U. Qamar and M. Y. Javed. (2014), An Ensemble based Decision Support Framework for Intelligent Heart Disease Diagnosis, in *International Conference on Information Society*, London, UK.

F. Melgani and L. Bruzzone. (2004), Classification of Hyperspectral Remote Sensing Images With Support Vector Machines, *IEEE Transactions on Geoscience and Remote Sensing,* 42 (8), pp. 1778 - 1790.

S. U. Ghumbre and A. A. Ghatol. (2012), Heart Disease Diagnosis Using Machine Learning Algorithm, in *Proceedings of the International Conference on Information Systems Design and Intelligent Applications*, Visakhapatnam, India.

C.-W. Hsu and C.-J. Lin. (2002), A Comparison of Methods for Multiclass Support Vector Machines, *IEEE Transactions On Neural Networks,* 13 (2), pp. 415 – 425