Developing a Software for Diagnosing Heart Disease via Data Mining Techniques

Mustafa G. Saeed[a] and Yaser A. Jasim[b]

[a]Computer Science Dept. Cihan University- Sulaymaniyah\Iraq
[b]Accounting Dept. Cihan University-Erbil\Iraq
mostafa.swe@gmail.com, Yaser.a.Jasim@Cihanuniversity.edu.iq

KEYWORD	ABSTRACT
Data mining; Artificial Neural Network; Matlab R2016a and Heart disease	*This paper builds a data mining tool via a classification method using Multi-Layer Perceptron (MLP) with Backpropagation learning method and an algorithm of feature selection along with biomedical testing values for diagnosing heart disease. Addition to that, developing a prototype for heart disease diagnosing with a friendly-user graphical interface (GUI). The purpose to construct this software is that; clinical prosopopoeia is done in any event by doctor's experience. Despite that, some cases are reported negative diagnosis and treatment; therefore, patients are asked to take a number of tests for diagnosis. Moreover, not all the tests contribute towards an effective diagnosis of a disease, and by using data mining approach to diagnose heart disease that supports the doctors to make more efficient and subtle decisions.*

1. Introduction

Nowadays, Data Mining became the availability of a huge quantity of data which leads to the need of robust tool to analyze data for extracting useful knowledge; where researchers have long been concerned in applying tools of data mining and statistics to enhance the analysis of big data. Many diagnosis applications are applying data mining tools for giving successful outcomes, over the past ten years heart disease (HD) is the leading cause of death all over the world. Several researchers are using statistical and data mining tools to aid healthcare professionals in diagnosing heart disease [9,12].

Medical diagnosis is extremely significant, but on the other hand, it is a complex task that should be achieved efficiently and accurately. HD is diagnosed with a fixed form of CVD, including (1) stroke (2) heart attack and (3) chest pain, or with cardiovascular disease (CVD). Heart care case specifies 15 regular attributes to predict the disease issue. Addition to regular attributes, there are general attributes such as BMI (Body Mass Index), age, income, physician supply, education, ethnicity, and others used for prediction. An automated system for diagnosing medical cases would improve medical care and cost reduction [12].

Artificial Neural Network (ANN) which often known as "neural network" (NN) is one of the most important tools in data mining, which is a mathematical or computational module that imitates biological neural networks. In most cases, ANN is an adjustive system that modifies its structure depending on the internal or external information that flows over the network throughout the learning phase [10].

Mustafa G. Saeed and Yaser A. Jasim
Developing a Software for Diagnosing Heart Disease
via Data Mining Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 7 N. 3 (2018), 99-114
eISSN: 2255-2863 - http://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY NC DC

99

Moreover, as practical terms ANNs are non-linear tools of statistical data modeling. They are used to module convoluted relationships between input vectors and outputs or to locate patterns in a data set. Using ANNs as a tool, data warehousing organizations are gathering information from datasets which this process is recognized as data mining. The distinction between data warehouse and ordinary database; is that there are actual manipulation and cross-fertilization of the data providing users a broad information in decision making [10].

2. Related Works

During the last years, the data mining tool to analyze data have received a considerable amount of attention in the research community. Vijayashree and SrimanNarayanaIyengar proposed a system to brief about decision making support systems for the anticipation and diagnosis of heart disease powered by data mining and hybrid intelligent techniques [8]. A proposed system done by Abhishek Taneja uses data mining technologies to diagnose heart disease and developing a cost-effective treatment system [1]. Where Aditya Methaila, Kansal, Arya and Kumar intend to use data mining tools for Classification Modeling Techniques to predict and diagnose heart disease [2]. A system constructed by Safwan O. and Yaser A. Jasim uses artificial neural networks and data mining tool to diagnose and repair application errors [7].

3. Data Mining

Recently, with the enormous quantity of stored data in databases, repositories and files, it is increasingly significant and necessary, to develop robust means for analysis and exponent of such data and for extracting motivating knowledge that could support the decision-making [4]. The term *Data Mining*, which is known in popularly as *Knowledge Discovery in Databases* (KDD), referring to the implicit nontrivial extraction of data kept in databases, where previously was unknown and probably useful information [4].

3.1. Data Mining Process Model

The Knowledge Discovery in Databases (KDD) process consists of phases leading from a set of raw data to somewhat form of new knowledge or wisdom. The iterative process comprises of the following phases [13]:

1- **Data Cleaning**: or "Data cleansing", it is a level in which irrelevant data and noisy data are taken away from the collection.
2- **Data Integration**: at this level, multiple data sources, may be merged in a common source.
3- **Data Selection**: at this level, the decision is made for the relevant data to analyze and retrieve it from the data collection.
4- **Data Transformation**: or called "Data consolidation", it is a level in which the allocated data is transformed into appropriate forms for the mining procedures.
5- **Data Mining**: it is the pivotal level in which smart techniques are applied for pattern extract that is potentially useful.
6- **Pattern Evaluation**: in this level, forthrightly interesting patterns that represent knowledge are identified based on received measures.
7- **Knowledge Representation:** is the final level in which the retrieved knowledge is visually displayed to the user. This fundamental phase uses visualization techniques for supporting users to interpret and understand the data mining outcomes.

3.2. Types of Database in Information Systems

Data mining is set for using and studying database, which includes an object-oriented database, transactional database, relational database, data warehouses and object-relational database; where table 1 depicts these types of database. Moreover, time-series database, semi-structured and unstructured repositories like the world wide web (WWW), multimedia database, textual database, advanced database like spatial database [12,13].

Table 1: Data Mining Challenges Presented via Different Types of Database

Database Types	Challenges Faces Datamining
Flat Files	This is a simple type of data files, where its format is binary or text with a framework known as the data-mining algorithm used to be applied in various cases. The data can be scientific measurements, time-series data, transactions, etc.
Relational Databases	Data mining algorithms that use relational databases can be more multilateral than data mining algorithms which are written for flat files. While data mining takes benefits of SQL for data consolidation, transformation and selection it outdoes what SQL could offer, like comparing, predicting, detecting divergence, etc.
Data Warehouses	A data warehouse provides a preference to analyze data of similar various sources. In other meaning, data from the various repositories would be cleaned, transformed, loaded and integrated together; to ease the decision-making process and facilitate multi-dimensional views, where data warehouses usually modeled via a multi-dimensional data structure.
Transaction Databases	One exemplary data mining analysis on data is the known as "market basket analysis" or "association rules" in which associations among items that occur together with or in sequence are taken in consider for studies.
Multimedia Databases	Multimedia is characterized via its high dimensionality, which raises more challenges for data mining. Data mining of multimedia repositories could require computer graphics, computer vision, natural language processing methodologies, image interpretation.
Spatial Databases	Like spatial database that exhibits new take on data mining algorithms.
Time-Series Databases	Data mining for this kind of databases commonly comprises the survey of trends and correlations among different variables evolutions, as well as the Anticipation of trends and movements of variables in time.
World Wide Web	Data mining in web mining or world wide web tries to address all issues located in web and often divided into web structure mining, web usage mining and web content mining.

4. Artificial Neural Networks

Artificial Neural Network (ANN) also known as "neural network" (NN), where artificial neural networks are relatively modest electronic nodes based on the neural pattern of the brain [14]; in different words, is a mimic of the biological neural system. It is also a computational model or mathematical model depending on biological neural networks. ANN consists of a connected set of processes information and artificial neurons that uses a connectionist method for computation as a figure (1) depicts its structure. Mostly, ANN is a system that adapts itself to changing its structure depending on the internal or external information that flows throughout the network structure during the learning stage [10].

A neural network is a broadly distributed parallel processor that in nature stores experiential knowledge and make it available for use. It looks like the brain [3]:

Mustafa G. Saeed and Yaser A. Jasim
Developing a Software for Diagnosing Heart Disease
via Data Mining Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 7 N. 3 (2018), 99-114
eISSN: 2255-2863 - http://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY NC DC

101

1. Knowledge is gained via the network through the learning phase.
2. The strengths of interneuron connection are known as synaptic weights and used for storing the knowledge. The procedure used to achieve the learning process is named a learning algorithm.
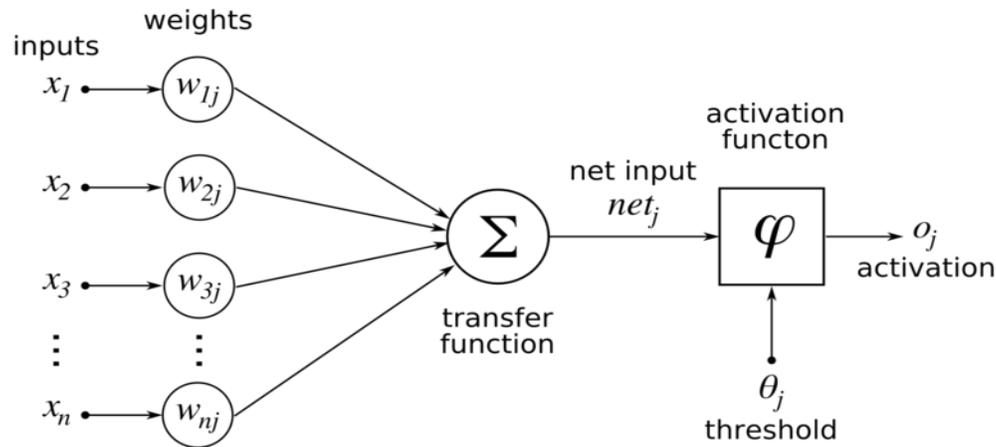

Fig. 1: Structure of Artificial Neural Network

4.1. Backpropagation Neural Network

Backpropagation algorithm is the most popular neural network algorithms; After selecting the network weights in a random manner, using backpropagation algorithm to compute the corrections. The algorithm can be illustrated in the following four steps:

1- Feed-forward computation
2- Backpropagation to the output layer
3- Backpropagation to the hidden layer
4- Weight updates

The algorithm is obstructed when the error function value becomes small [5].

4.2. Patterns Recognition in Artificial Neural Network

Automatic (machine) recognition, classification, a grouping of patterns and description are significant issues in a variety of scientific and engineering disciplines such as biology, remote sensing, psychology, artificial intelligence, marketing, medicine and computer vision. A pattern might be a (human face, speech signal, fingerprint image or a handwritten cursive word). Given a pattern, its classification/ recognition could depends on one of the following tasks:

1. Supervised classification (e.g., discriminate analysis) in which the input pattern is specified as a predefined class member,
2. Unsupervised classification (e.g., clustering) in which the pattern is dedicated to a hitherto unknown class [6].

4.3. Artificial Neural networks in data mining

Neural networks are non-linear tools for statistical data modeling. They are used to module complex relationships among input vectors and outputs or to look for patterns in data. Using neural networks as a tool, data

information system companies are gathering information from datasets in the process called data mining. The difference between this data information system and traditional databases is that there is cross-fertilization and actual control of the data which supports users to make an accurate and efficient decision [10].

5. System Implementation, Tests, and Empirical Outcomes

This section depicts the implementation of artificial neural network and the data mining system also; the features diagnosis the heart disease. The database is received from Cleveland Clinic Foundation database; which consists of (296) samples and each one has (76) attributes which the network training process is based on them. The database contains (286) persons, the outcome is represented by two cases (Healthy, Sick) for each person; figure (2) shows the main phase of the heart diagnosis system implementation.
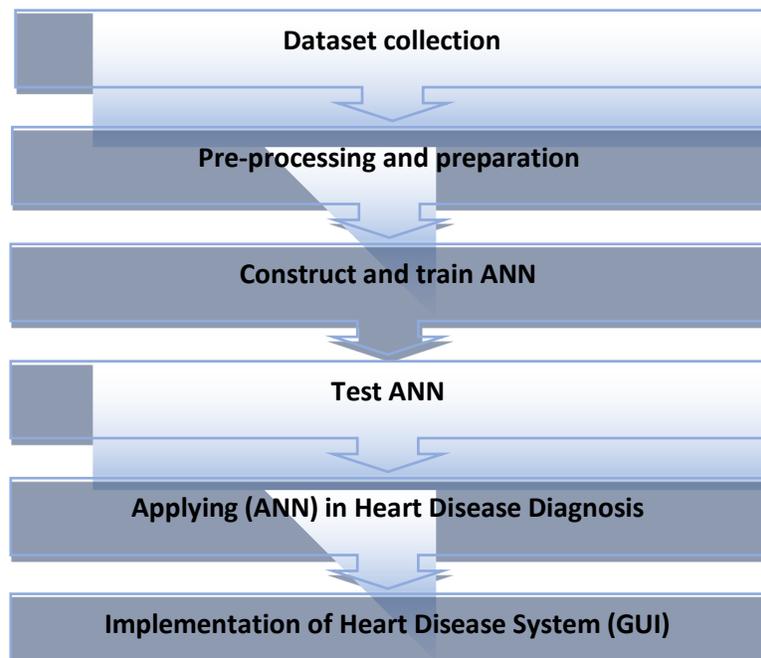
Dataset collection

Pre-processing and preparation

Construct and train ANN

Test ANN

Applying (ANN) in Heart Disease Diagnosis

Implementation of Heart Disease System (GUI)

Fig. 2: General Structure of Heart Diagnosis System

5.1. Data Preprocessing

Filtering and cleaning the data probably could be necessarily carried out to the data and data-mining algorithm dedicated to avoiding the creation of fraudulent or inappropriate patterns or rules. Firstly; the preprocessing selects an attribute for allocating a subset of attributes with a good predicting capability. The database is a collection of attributes that symbolize feature for each person and it aids the doctors to diagnose diseases. The database consists of (76) attribute, where (13) of them are applied for each person as shown in the table (2), while the goal will decide that the person is sick or not.

In first phase data is saved as an Excel file and it includes two types of features; text data and numeric data as coded in Matlab R2016a programming language ([numericData, textData] = xlsread(Cleveland.xlsx') ;). Where the file contains (14) attributes and (8) of them are symbolic while (6) are numeric as illustrated in table (2); which every symbolic feature should be converted to numeric values.

Table 2: Feature Extraction and Encoding

Attributes	Values	Numeric
Age	Numeric	Numeric
Gender	Male, Female	1,0
Chest Pain Type	Angina, Abnormal, No Tang, Asymp	1-4
Blood Pressure	Numeric	Numeric
Cholesterol	Numeric	Numeric
Fasting Blood Sugar	<120 True, >120 False	1,0
Resting ECG	Normal, Abnormal, Hyp	0,1,2
Maximum Heart Rate	Numeric	Numeric
Induced Angina	True, False	1,0
Old Peak	Numeric	Numeric
Slope	Up, Flat, Down	1,2,3
Number Colored Vessels	0,1,2,3	0,1,2,3
Thal	Number, Fix, rev	3,6,7
Concept Class	Healthy, Sick	1,0

Feature number two in the database (Gender) has a symbolic value (Female, Male) as displayed in Matlab code; figure (3) shall depict this code. This feature is converted to (zeros, ones) which male take (one) as its value while female take (zero) as its value.

```
gender=textData(2:end,2);
 [row col]=size(gender);
Gender=zeros(row,1);
for indr=1:row
    temp=gender{indr};
    if strcmp(temp,'male')==1
        Gender(indr)=1;
    end
end
```

Fig. 3: Matlab Code for Gender Feature

The third Feature in the database (Chest pain type) includes (4) symbolic values (asymptomatic, typical angina, non-angina pain, atypical angina) as Matlab code shows in figure (4). This feature is converted into (one, two, three and four) while typical angina takes (one) as its value, atypical angina take (two), non-angina pain takes (three) as its value and asymptomatic takes four as its value.

```
cp=textData(2:end,3);
 [row col]=size(cp);
Cp=zeros(row,1);
for indr=1:row
    temp=cp{indr};
    if strcmp(temp,'angina')==1
        Cp(indr)=1;
    elseif strcmp(temp,'abnang')==1
        Cp(indr)=2;
    elseif strcmp(temp,'notang')==1
        Cp(indr)=3;
    elseif strcmp(temp,'asympt')==1
        Cp(indr)=4;
    end
end
```

Fig. 4: Matlab Code for Chest Pain Type Feature

Feature number six in the database (fasting blood sugar) holds two symbolic values (True and False) as the Matlab code depicts in figure (5). Where its converted to (one and two) while true takes the value (one) and false takes (zero) as its value.

```
 [row col]=size(fbs);
Fbs=zeros(row,1);
for indr=1:row
    temp=fbs{indr};
    if strcmp(temp,'')==1
        Fbs(indr)=1;
    end
end
```

Fig. 5: Matlab Code for Fasting Blood Sugar Feature

The seventh Feature in the database (resting electrocardiographic results) has three symbolic values (normal, showing probable\definite left ventricular hypertrophy, having ST-T wave abnormality) as shown in the Matlab code in figure (6). Where this feature must be transported to (zero, one, two); therefore, typically normal takes (zero) as its value, having ST-T wave abnormality takes (one) as its value and showing probable or definite left ventricular hypertrophy allocates (two) as its value.

Mustafa G. Saeed and Yaser A. Jasim
Developing a Software for Diagnosing Heart Disease
via Data Mining Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 7 N. 3 (2018), 99-114
eISSN: 2255-2863 - http://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY NC DC

105

```
restecg=textData(2:end,7);
 [row col]=size(restecg);
Restecg=zeros(row,1);
for indr=1:row
    temp=restecg{indr};
    if strcmp(temp,'norm')==1
        Restecg(indr)=0;
    elseif strcmp(temp,'abn')==1
        Restecg(indr)=1;
    elseif strcmp(temp,'hyp')==1
        Restecg(indr)=2;
    end
end
```

Fig. 6: Matlab Code for Resting Electrocardiographic Results Feature

Feature number nine in database (exercise-induced angina) contains two symbolic values (Yes and No) as shown in the Matlab code in figure (7). This feature is converted into (zero and one), which yes takes (one) as its value and no takes (zero) as its value.

```
exang=textData(2:end,9);
[row col]=size(exang);
Exang=zeros(row,1);
for indr=1:row
    temp=exang{indr};
    if strcmp(temp,")==1
        Exang(indr)=1;
    end
end
```

Fig. 7: Matlab Code for Exercise-Induced Angina Feature

Eleventh Feature (the peak exercise ST segment slope) has three symbolic values (flat, up sloping, downsloping) as the Matlab code illustrates in figure (8). This feature will be converted to (one, two, three) which up sloping takes (one) as its value, flat takes (two) as its value and down sloping takes (three) as its value.

Mustafa G. Saeed and Yaser A. Jasim
Developing a Software for Diagnosing Heart Disease
via Data Mining Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 7 N. 3 (2018), 99-114
eISSN: 2255-2863 - http://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY NC DC

106

```matlab
slope=textData(2:end,11);
 [row col]=size(slope);
Slope=zeros(row,1);
for indr=1:row
   temp=slope{indr};
   if strcmp(temp,'up')==1
      Slope(indr)=1;
   elseif strcmp(temp,'flat')==1
      Slope(indr)=2;
   elseif strcmp(temp,'down')==1
      Slope(indr)=3;
   end
end
```

Fig. 8: Matlab Code for Slope of Peak Exercise ST Segment Feature

Feature number thirteen (Thal) has three symbolic values (fixed defect, normal, and reversible defect) as the Matlab code depicts in figure (9). Where this feature is converted to (three, six, seven); therefore, normal takes (three) as its value, a fixed defect takes (six) as its value and reversible defect takes (seven) as its value.

```matlab
thal=textData(2:end,13);
whos thal
[row col]=size(thal);
Thal=zeros(row,1);
for indr=1:row
   temp=thal{indr};
   if strcmp(temp,'norm')==1
      Thal(indr)=3;
   elseif strcmp(temp,'fix')==1
      Thal(indr)=6;
   elseif strcmp(temp,'rev')==1
      Thal(indr)=7;
   end
end
```

Fig. 9: Matlab Code for Thal Feature

A feature known as (Dhd) is the target that will be compared with the input data for ANN training as the Matlab code shows in figure (10).

```
dhd=textData(2:end,14);
 [row col]=size(dhd);
Dhd=zeros(row,1);
for indr=1:row
    temp=dhd{indr};
    if strcmp(temp,'sick')==1
        Dhd(indr)=1;
    end
end
```

Figure 10: Matlab Code for ANN Target

Moreover, table (3) represents origin database which is initial to date preprocessing, and table (4) represents database after data preprocessing where it converts every symbolic feature to numeric in order to ANN accepts only numeric value as input vectors.

Table 3: Sample of Origin Database Before Data Preprocessing

Age	Gender	Chest Pain Type	Resting Blood Pressure	Cholesterol	Fasting Blood Sugar	Resting Electrocardiographic Results	Achieved Maximum Heart Rate	Exercise-Induced Angina	Depression Induced by Exercise relative to rest	The Peak Exercise ST Segment Slope	Number of Major Vessels	Thal	Diagnosis of Heart Disease
63	Male	Angina	145	233	True	Hyp	150	Fal	2.3	Down	0	Fix	Buff
67	Male	Asympt	160	286	Fal	Hyp	108	True	1.5	Flat	3	Norm	Sick
67	Fem	Asympt	120	229	Fal	Hyp	129	True	2.6	Flat	2	Rev	Sick
37	Male	Notang	130	250	Fal	Norm	187	Fal	3.5	Down	0	Norm	Buff

Table 4: Sample of Database After Data Preprocessing Stage

Age	Gender	Chest Pain Type	Resting Blood Pressure	Cholesterol	Fasting Blood Sugar	Resting Electrocardiographic Results	Achieved Maximum Heart Rate	Exercise-Induced Angina	Depression Induced by Exercise	The Peak Exercise ST Segment Slope	Number of Major Vessels	Thal	Diagnosis of Heart Disease
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
67	0	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0

5.2. Design and Test of Artificial Neural Network (BP)

The input layer includes thirteen nodes as an input vector, which represent features also there are (13) nodes in the hidden layer, and one node is represented for output vectors that depict the number of persons in the system's database.

The dataset contains (296) samples and its distributed as (246) samples for training and (50) samples for testing. In this level, the network has been trained and as following figure (11) shows the performance diagram where figure (12) illustrates; results are acquired.
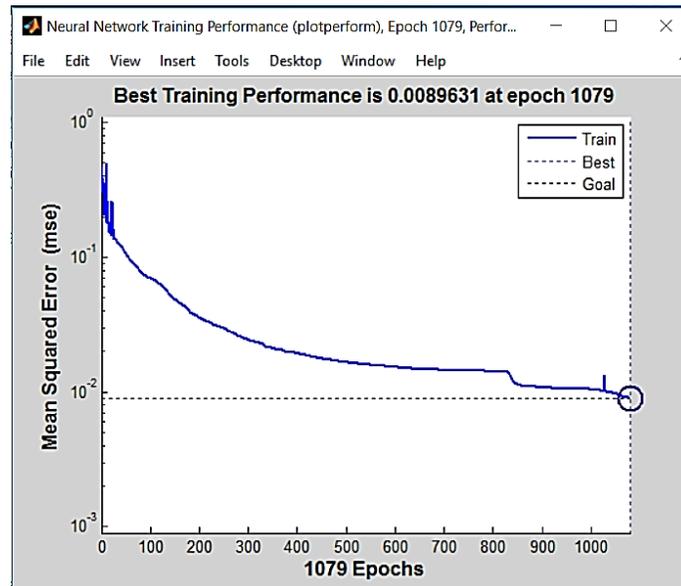

Figure 11: Performance Diagram

Mustafa G. Saeed and Yaser A. Jasim
Developing a Software for Diagnosing Heart Disease
via Data Mining Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 7 N. 3 (2018), 99-114
eISSN: 2255-2863 - http://adcaij.usal.es
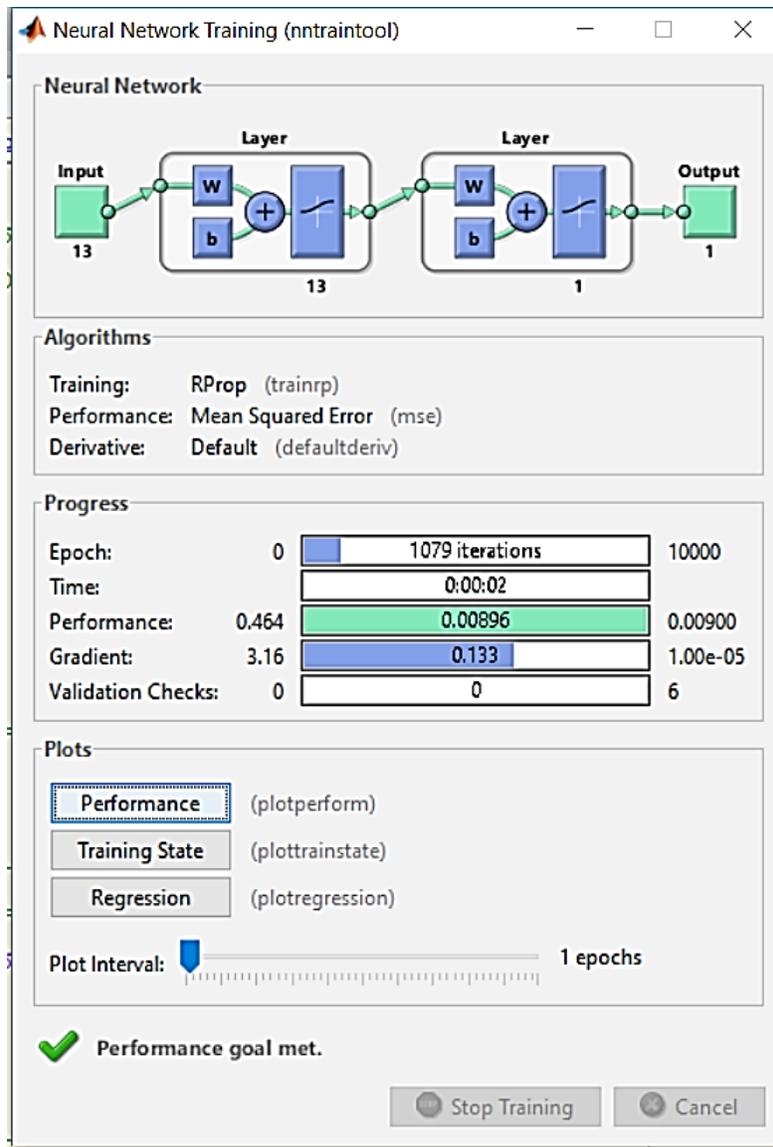Ediciones Universidad de Salamanca - CC BY NC DC

109

Figure 12: Artificial Neural Network Training Phase

Iteration = 1079
Time = 0.00:02s
Performance = 0.00896 Mean Square Errors

The neural network was examined using (50) samples. Moreover, the correlation coefficient function (CORR2) and mean squared normalized error performance function (MSE) was used while the outcome of the neural network was (0.9) after normalizing actual output of ANN as the Matlab code shows in figure (13).

Mustafa G. Saeed and Yaser A. Jasim
Developing a Software for Diagnosing Heart Disease
via Data Mining Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 7 N. 3 (2018), 99-114
eISSN: 2255-2863 - http://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY NC DC

110

```
out=sim(p.net,tdata)
for ind=1:length(out)
    if out(ind)> 0.9
        out(ind)=1;
    else
        out(ind)=0;
    end
end
e=target-out
msea=mse(e)
corr=corr2(target,out)
```

Figure 13: Matlab Code for Testing (50) Samples on ANN

1- Mean Squared Normalized Error = 0.0196
2- Correlation Coefficient = 0.9610 (Correlation coefficient values between (0,1); if the result is close to (1) that means good matching between the target and the actual output of ANN).

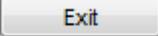## 5.3. Graphical User Interface of Application

The graphical user interface of diagnosis system is designed using Matlab GUIDE Tools, where the system interface contains three GUI, first of all; Main Interface that runs by calling the main function, this manifest as shown in figure (14).

Figure 14: Main GUI of Heart Disease Diagnosis System

The main user interface contains three action event buttons as mentioned below:

1- **Heart Disease Diagnosis Button** calls the GUI interface, which contains the information of diagnosing heart diseases.
2- **Heart Disease Database Button** ⟨Heart Disease Database⟩ calls the GUI interface for the database, which contains the report for each patient.

Mustafa G. Saeed and Yaser A. Jasim
Developing a Software for Diagnosing Heart Disease
via Data Mining Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 7 N. 3 (2018), 99-114
eISSN: 2255-2863 - http://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY NC DC

111

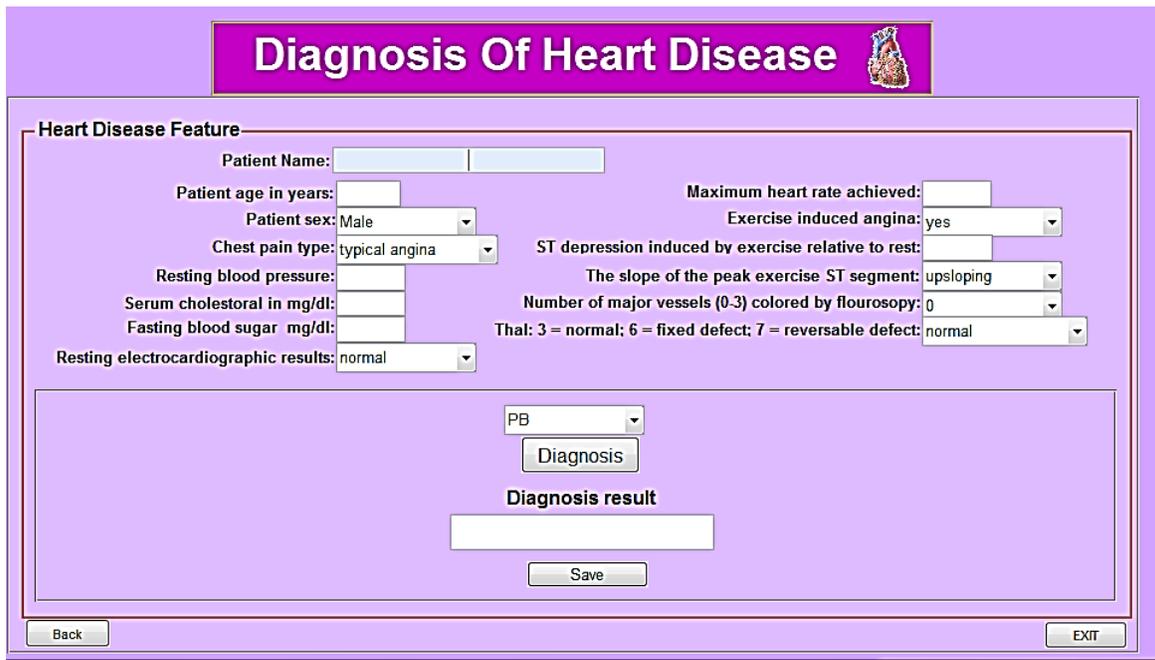**3-** Exit **Exit Button** closes the application.

Figure 15: Diagnosis of Heart Disease Interface

On the other hand, there is a subinterface titled as Diagnosis of Heart Disease **(**DHD**)** as shown in figure (15). Here in this Interface, there are text boxes that contain symbolic or numeric values of the patient, and select box that includes a multi choice of values as a figure (16) depicts, where it contains four buttons:
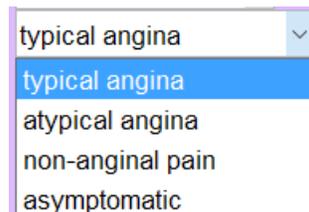

Figure 16: Multi-Choice Feature Selection

1) **Diagnosis Button** Diagnosis which diagnosis the patient states and report if (sick or healthy). Therefore, if the user didn't fill all fields, the system shows a warning message as shown in figure (17).

Mustafa G. Saeed and Yaser A. Jasim
Developing a Software for Diagnosing Heart Disease
via Data Mining Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 7 N. 3 (2018), 99-114
eISSN: 2255-2863 - http://adcaij.usal.es
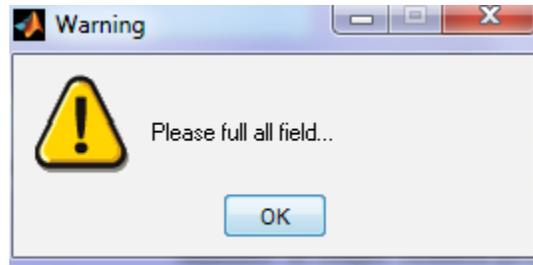Ediciones Universidad de Salamanca - CC BY NC DC

112

Figure 17: System Error Message

2) [Save] **Save Button** when the user clicks on this button the system saves the diagnosis data of the patient in the table; then later the user can retrieve its database.

3) **Back Button** [Back] returns the user to the main Interface.

4) [Exit] **Exit Button** closes the DHD Interface by the user.

Addition, another Interface titled **Report Page** is shown in figure (18):

Figure 18: Report Interface

This Interface displays three buttons:

1. **Search Button** [search] the user searchs for the patient field from the table that contains information about patients.

2. **Back Button** [Back] returns the user to the main Interface.

3. [Exit] **Exit Button** closes the report Interface.

Mustafa G. Saeed and Yaser A. Jasim
Developing a Software for Diagnosing Heart Disease
via Data Mining Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 7 N. 3 (2018), 99-114
eISSN: 2255-2863 - http://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY NC DC

113

6. Conclusion

This paper explains how feature selection helps increasing computational efficiency and enhancing classification accuracy. Moreover, the complexity of the system is decreased via reducing the dataset. At the same time, it also can save repository size, decrease computational requirement, save health checklist costing and reduces the number of attributes, which is necessary must be taken from patients. A total of (76) attributes that were available and (12) attributes that are highly relevant in predicting heart disease from the Cleveland Center dataset were selected.

This study illustrated that data mining techniques are used effectively to predict and model heart disease cases. The outcome can be used as an aid tool by cardiologists supporting them to make a more consistent diagnosis of heart disease. In addition, the artificial neural network is effectively used in heart diseases diagnosis.

Designing a GUI tool to diagnose heart disease made it easy for unprofessional users to control and use the designed software.

7. Reference

1. Abhishek Taneja, 2013, "Heart Disease Prediction System Using Data Mining Techniques", Oriental Journal of Computer Science & Technology, Vol. 6, No. (4), PP. 457-466, India.
2. Aditya Methaila, Prince Kansal, Himanshu Arya and Pankaj Kumar, 2014, "Early Heart Disease Prediction Using Data Mining Techniques", Maharaja Surajmal Institute of Technology New Delhi, India.
3. Bajpai, S., Jain, K., and Jain, N., 2011, "Artificial Neural Networks". International Journal of Soft Computing and Engineering (IJSCE) ISSN, 2231-2307.
4. Chen M., Han J., and Yu, P., 1996, "Data Mining: An Overview from A Database Perspective", IEEE Transactions on Knowledge and Data Engineering, 8(6), 866-883.
5. Cilimkovic, M., 2007, "Neural Networks and Back Propagation Algorithm", Institute of Technology Blanchardstown, Blanchardstown Road North Dublin, Ireland.
6. Dutt, V., Chadhury, V., and Khan, I., 2011, "Different Approaches in Pattern Recognition", Computer Science and Engineering, Vol. 1, Issue 2, PP 32-35, India.
7. Hasoon, Safwan O., Jasim, Yaser A., 2013, "Diagnosis Windows Problems Based on Hybrid Intelligence Systems", Journal of Engineering Science & Technology (JESTEC), Vol. 8, Issue 5, Pages 566-578, School of Engineering, Taylor's University, Malaysia.
8. J. Vijayashree and N.Ch.SrimanNarayanaIyengar, 2016, "Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques: A Review ", International Journal of Bio-Science and Bio-Technology, Vol.8, No.4, PP. 139-148, India.
9. Shouman, M., Turner, T., and Stocker, R., 2012, "Using Data Mining Techniques in Heart Disease Diagnosis and Treatment", In Electronics, Communications and Computers (JEC-ECC), Japan-Egypt Conference on (pp. 173-177), IEEE.
10. Singh, Y., and Chauhan, A., 2009, "Neural Networks in Data Mining", Journal of Theoretical and Applied Information Technology, Vol. 5, No. 6, 36-42.
11. Srinivas, K., Rao, G. R., and Govardhan, A., 2010, "Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques", In Computer Science and Education (ICCSE), 2010 5th International Conference on (pp. 1344-1349), IEEE.
12. Tan, P., Steinbach, M., and Kumar, V., 2013, "Data Mining Cluster Analysis: Basic Concepts and Algorithms", Introduction to data mining, India.
13. Zaïane, O., 1999, "Principles of Knowledge Discovery in Databases", Department of Computing Science, University of Alberta.
14. Sonali. B. Maind, 2014, "Research Paper on Basic of Artificial Neural Network", International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 2, Issue: 1, pp. 96-100, India.

Mustafa G. Saeed and Yaser A. Jasim
Developing a Software for Diagnosing Heart Disease
via Data Mining Techniques

ADCAIJ: Advances in Distributed Computing
and Artificial Intelligence Journal
Regular Issue, Vol. 7 N. 3 (2018), 99-114
eISSN: 2255-2863 - http://adcaij.usal.es
Ediciones Universidad de Salamanca - CC BY NC DC

114