Utilization of an Artificial Neural Network in the Prediction of Heart Disease

Mohd Khalid Awang[1] and Fadzilah Siraj[2]

Faculty of Informatics,
University of Sultan Zainal Abidin, Gong Badak Campus,
21300 Kuala Terengganu, Terengganu, Malaysia
[1]khalid@unisza.edu.my, [2]fad173@uum.edu.my

Abstract

This paper intents to assess the application of artificial neural network in predicting the presence of heart disease, mainly the angina in patients. The prediction and detection of angina are significant in determining the most appropriate form of treatment for these patients. The development of the application involves three main phases. The first phase is the development of Heart Disease Management Information System (HDMIS) for data collection and patient management. Then followed by the second phase, which is the development of Neural Network Simulator (NNS) using back propagation neural network for training and testing. The final phase is the development of Prediction System (PS) for prediction on new patient's data. The best network model produced prediction accuracy of 88.89 percent. Apart from proving the ability of neural network technology in medical diagnosis, this study also shown how the neural network could be incorporated into the hospital information system as a prediction tool. As the pilot project, the application developed could be used as the starting point in building a medical decision support system, particularly in diagnosing the heart disease.

Keywords: *Prediction, Neural Network, Heart Disease*

1. Introduction

Worldwide, cardiovascular disorders and heart disease are considered as the number one killer. Even though the death rates from the diseases are now on the decrease because of changes in lifestyle, it is still a major cause of death and disability in both developing and developed countries. On the hand, World Health Organization (WHO) stated that, cardiovascular disorders claimed around 13.4 million people in 2008 [3]. People in the developing country including Malaysian are more exposed to the risk. According to Malaysian Ministry of Health [6], cardiovascular disease is the leading cause of premature death in hospitals; where about 35% of deaths from heart disease occurred in individuals aged less than 60 years. In 2010, there were about 25.4% of death was related cardiovascular disease. In addition to its negative impact on quality of life, increase the economic burden of cardiovascular disease individuals, families and communities, and reduce the productivity of the country.

2. Literature Review

The main function of the cardiovascular system is to supply vital oxygen as well as nutrients to all body cells and tissues will die within minutes without this supply [2]. A cardiovascular disorder is a medical terminology that describes all kinds of diseases that

related to our heart and circulatory system. The coronary heart disease (CHD) is one of the most dangerous heart problems, which occurs when the heart muscle receives insufficient oxygen because the coronary arteries fail to maintain a sufficient supply of blood [2].

The statistics presented reveal that heart disease and cardiovascular disorders are one of the world's most important causes of mortality, so improvement of diagnosis procedure would be very vital. Even though there is still considerable uncertainty about the cause of heart disease that make the diagnosis more difficult, it also reveals that the coronary anatomy can be predicted from finding and the initial encounter such smoking habits and cholesterol level [2]. [4] indicated that a lot of life could be saved and at the same time some of the total cost related to the diagnostic testing could be reduced if the initial assessments are reliable and further testing only be given to high risk patients. There have been previous efforts that use statistical methods to predict the presence of heart disease in patients; however, this method may not be able to identify all the hidden relationships between the various risk factors of the heart diseases that make the result not satisfactory.

Diagnosing heart disease is considered as a non-linear problem that shows the complex causal relationship between the variables [2]. However, there is a new computational paradigm called an artificial neural network, which is suitable for problems of extreme complexity not addressable with our conventional technologies, either by the conventional computer programming or statistical method. Several studies have shown that an artificial neural network can be successfully applied in diagnosing heart diseases [1]. Therefore, the purpose of this study is to evaluate the application of neural network in predicting the presence of heart disease, particularly the angina in patients. The prediction is mainly used for diagnosis angina in patients that are already diagnosed with myocardial infarction. The diagnosis of angina is important in determining the most appropriate form of treatment for myocardial infarction patients. Furthermore, diagnosis and management of angina is important for these patients since the angina can lead to the recurrent of myocardial infarction [2]. Nevertheless, the prediction system can also be used for new patients when they are admitted to the hospital with some sort of chest pain.

3. Research Methodology

Since the required data for the study is not available in a ready to use format, the effort started with the development of Management Information System to provide a fundamental tool for data gathering and management. Once the data are collected, they must be prepared into certain format before being passed through the neural network. Before the development of a neural network simulator, the neural network model and architecture must be defined. Then the network is trained and tested and the results are stored in the weight file, which was required by the prediction system. Finally, all works are integrated to form a single application model. Basically, there are seven major stages involved in building the application as listed below:

A. The development of the Heart Disease Management Information System (HDMIS)
B. Data preparation
C. Neural network modeling
D. The development of Neural Network Simulator (NNS)
E. The development of Prediction System (PS)
F. The integration of all systems (MIMIS, NNS, PS)

A. The development of the Heart Disease Management Information System (HDMIS)

The development of the application began with the development of Heart Disease Management Information System. The development of this system is crucial because it serves as the initial step for data collection. Furthermore, this system is also used as the real system for data entry as well as the input for both Neural Network Simulator and Prediction System. The Neural Network Simulator will extract the new data from this system for further retraining and retesting whilst the Prediction System uses new patient data for prediction purpose.

B. Data preparation

One of the most important parts in data preparation is to determine the best variables that contribute to the decision-making. The data selection step requires some detailed knowledge of the problem domain and the underlying data [5]. Therefore, the selections of the variables are based on the advice of the doctors and also the review of the literatures. Even though there are quite a number of variables entered into the Myocardial Infarction Management Information System, only eight are identified as the important variables that contribute to the prediction of angina. Those variables are the major risk factors of angina. They are as follows:

Table 1. The Risk Factors

Field Name	Data Type	Values	Representation and Normalization Method
Age	Continuous Numeric	0 to maximum value	Scale 0 to 1 New Value = Old Value / Max Value
Family history of premature CAD	Categorical	YES, NO	YES = 1 NO = 0
Smoking	Categorical	YES, NO	YES = 1 NO = 0
Cholesterols	Categorical	YES, NO	YES = 1 NO = 0
Hypertension	Continuous Numeric	0 to maximum value	Scale 0 to 1 New Value = Old Value / Max Value
Diabetes mellitus	Categorical	YES, NO	YES = 1 NO = 0
Hypercholesterolemia	Categorical	YES, NO	YES = 1 NO = 0

C. Neural Network Modeling

Once the model and architecture of the neural network are chosen, the next step is to train and test neural network with the selected data. The training and testing require the neural network simulator. Even though there are many kinds of commercial neural network software available in the market, a neural network simulator based on the model and architecture was developed. The main purpose of the development of this simulator is to extract the trained neural network model weights for predicting the new data set. However, some commercial software provides the model weights of the trained neural network, but it involves a lot of

interpretations. On the other hand, the model weight written by the Neural Network Simulator in this is presented in easy-to-read and ready to use without requiring a lot of interpretation. The neural network simulator uses the back propagation-learning algorithm, which is not differing from other commercial software.

D. The development of Neural Network Simulator (NNS)

Once the data preparation is complete and the neural network model and architecture is selected, the next important step is to train the neural network. Then the trained network has to go through a testing phase to see the network performance. The main purpose of neural network training is to teach the network to make a good generalization about the new data set which the network never seen before. Therefore, the network should not memorize the network patterns, but must learn the relationship among the variables involves in the network. The performance of the network is based on the testing results and the training result. The network with the highest testing results and with the lowest training results is considered as the best network model. Normally the neural network training and testing involve few stages as described in the Figure 2. It started with the selection of the best-hidden unit, followed by selection of learning and momentum rate, the activation function and finally the best stopping criteria are defined.

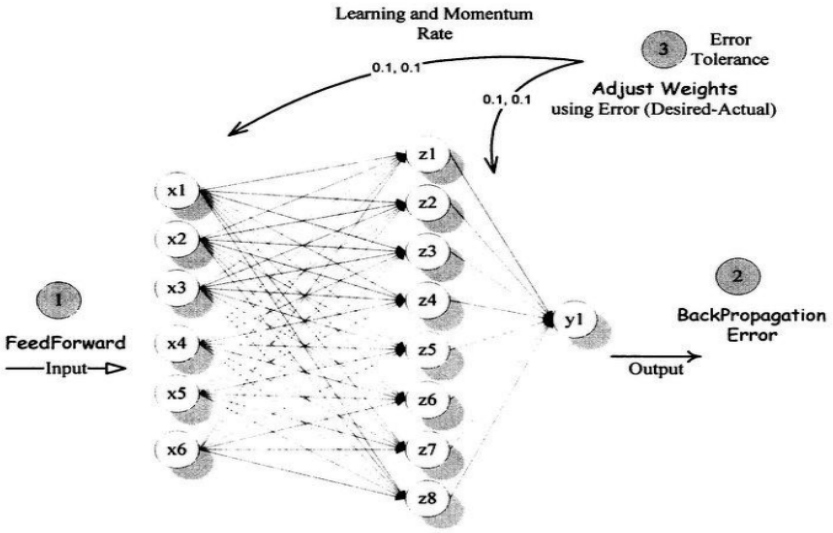

Figure 1. Neural Network Model

E. The development of Prediction System (PS)

Once the performance of the neural network is satisfied, the next step is to deploy and use it to score new patient's data. The development of Prediction System involves 4 major steps, which are as follows:

- Step 1: Reading input parameters from Myocardial Infarction Management Information System.
- Step 2: Automatic data preprocessing and representation.
- Step 3: Extract the model weights from trained neural network and feed forward the inputs through the network.
- Step 4: Calculate the output and post processing (scale) it for prediction purposes.

F. The integration of all systems (HDMIS, NNS, PS)

Once the individual system completed and tested, it will be integrated into one single system. For the purpose of the study, the Heart Disease Management Information System will serve as the main system for the integration. The Neural Network Simulator is integrated with the Maintenance Module whilst Prediction System is integrated with the Registration Module as shown in Figure 2.

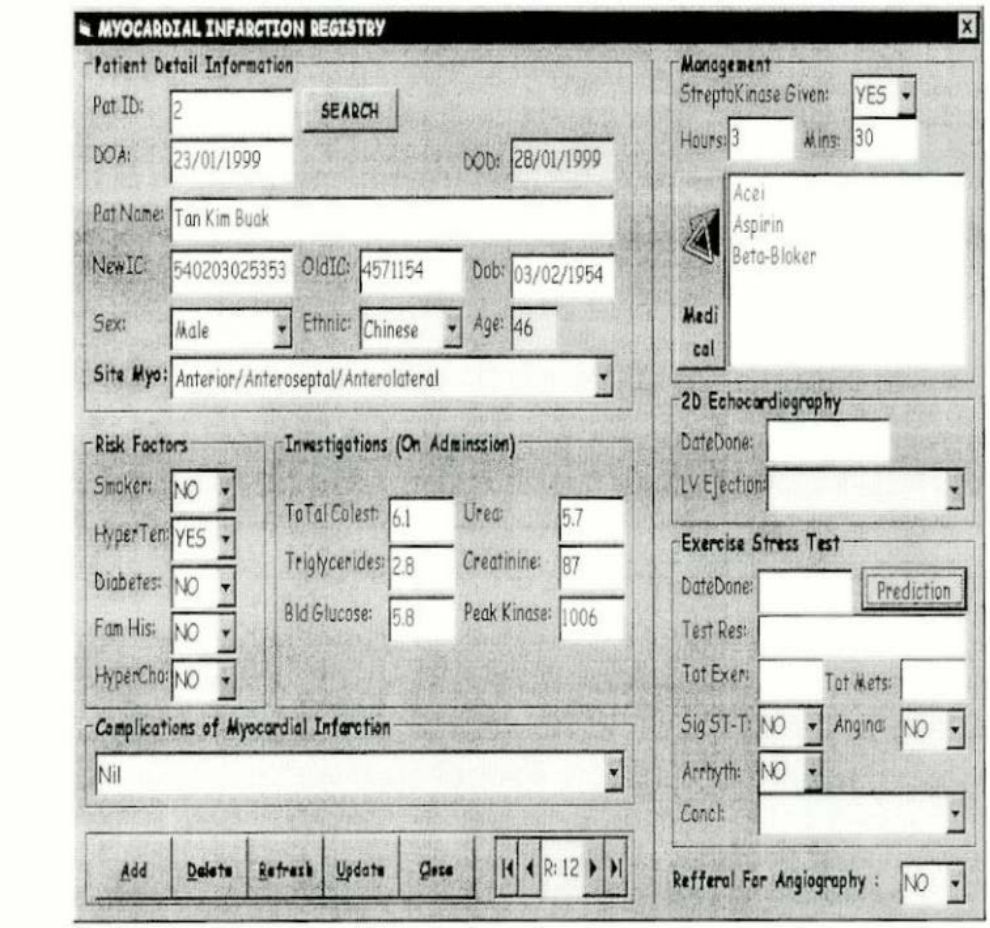

Figure 2. Prediction Interface

4. Result and Discussion

Hidden Unit 5 and 8 (both with training result of 83.33% and testing accuracy of 88.89%) have been chosen since they produce the highest testing accuracy and with the lowest possible training result. Hidden Units 11 also shows the same performance of training and testing, but has not been chosen because it requires more training and testing resources compared to Hidden Unit 5 and 8 [5].

Finally the best network model obtained consists of 6 input units, one hidden layer with 8 hidden units and one output unit. The learning rate of 0.1 and the momentum rate of 0.1 have been selected. The best stopping criteria is obtained at 1000 epochs and the network best perform with the binary sigmoid activation function. The network model has been able to

perform quiet well with the prediction accuracy of 88.99% when presented with the training data set. The final network model is presented in the following table.

Table 2. The Best Network Model

Network Model	
Model	Multi-Layer Perceptron with Back Propagation Algorithm
Input Units	6
Hidden Layers	1
Hidden Units	8
Output Unit	1
Learning rate	0.1
Momentum Rate	0.1
Activation Function	Binary Sigmoid
Stopping Criteria	1000 epochs

5. Conclusion

Basically, the main objective of the study, which is to evaluate the application of neural network technology in diagnosing the heart disease, particularly the presence of an angina in patients has been successful. The performance of back propagation neural network with the prediction accuracy of 88.89% is satisfactory. With the high accuracy of prediction, the neural network model can be applied in medical application, particularly in predicting the angina in patients. However, the current prediction system is not reliable enough because of the lack of patient's data. More cases need to be considered, trained and tested by the application. With the collection of further patient's data, it is expected that the accuracy of the network will improve and at the same time increase the predicted reliability up to the point where the method could become clinically useful.

This information could then be used to support a decision whether or not a patient should go through the EST. Apart from proving the ability of neural network technology in medical diagnosis; this study also shows how the neural network can be integrated into a management information system as a prediction tool. The application development can be used as the starting point in building a medical decision support system, particularly in diagnosing the heart disease. Most of Hospital Information Systems implemented in this country do not take into consideration the decision making support. In this paper, we have shown that artificial intelligent particularly neural network could be used an integrated with the current information system to provide intelligent diagnosis and prediction.

The Neural Network Simulator had only implemented the back propagation algorithm with the binary sigmoid activation function. In future, further study could be carried out to test different neural network models such as the Radial Basis Function as for comparison. In conclusion, this study has succeeded in initiating a pilot project that utilized neural network in predicting the presence of an angina in patients.

References

[1] R. F. Harrison, S. J. Marshall and R. L. Kennedy, "A connectionist aid to the early diagnosis of myocardial infarction", Proceed. Third European Conf. Artificial Intelligence in Medicine, Maastricht, The Netherlands, **(1991)**.
[2] M. Calnan, "Preventing Coronary Heart Disease: Prospects, Policies and Politics", Routledge, London, **(1991)**.
[3] V. L Roger, D. M. Lloyd-Jones, R. J. Adams, J. D. Berry, T. M. Brown, M. R. Carnethon, S. Dai, G. de Simone and E. S. Ford, Heart disease and stroke statistics--2011 update: a report from the American Heart Association, vol. 123, no. 4, **(2011)**, pp. e18-e209.
[4] D. B. Proyor, L. Shaw, C. B. McCants, K. L. Lee, D. B. Mark and F. E. Harrell, "Value of the History and Physical in Identifying patients at Increased Risk of Coronary Artery Disease", Ann Int Med, vol. 118, **(1993)**, pp. 81-90.
[5] J. P. Bigus, "Data Mining with Neural networks: Solving Business Problems -from application development to decision support", McGraw Hill, New York, **(1996)**.
[6] Ministry of Health, Annual Report. Ministry of Health Malaysia, **(2011)**.

Author

Mohd Khalid Awang is a Senior Lecturer in the Department of Computer Science at Sultan Zainal Abidin University (UniSZA) since 2006. He has received his M.Sc. degree in Information Technology from University Utara Malaysia (UUM) in 2000, and his B.Sc. degree in Computer Science from Indiana University at Bloomington (IUB) in 1994. His research interests include: Artificial Intelligence, Data Mining, and Ensemble Methods