

MULTILAYER PERCEPTRON BASED ACTIVATION FUNCTION ON HEART ABNORMALITY ACTIVITY

J. Adnan^{1,*}, N. G. N. Daud¹, A. S. N. Mokhtar¹, F. R. Hashim¹, S. Ahmad¹, A. F. Rashidi¹ and
Z. I. Rizman²

¹Department of Electrical and Electronic Engineering, Faculty of Engineering, National
Defence University of Malaysia, Sg. Besi Camp, 57000 Kuala Lumpur, Malaysia

²Faculty of Electrical Engineering, Universiti Teknologi MARA, 23000 Dungun, Terengganu,
Malaysia

Published online: 10 September 2017

ABSTRACT

This paper is aimed to detect the heart abnormality by using multilayer perceptron (MLP) network. Several data from the electrocardiogram (ECG) signal is extracted to be set as the input parameters. In order to get the best result, several activation function are used, such as unipolar sigmoid, bipolar sigmoid, tangent hyperbolic and conic section function. The result obtained is then compared among the activation function techniques.

Keywords: heart abnormality; multilayer perceptron; activation function; electrocardiogram.

Author Correspondence, e-mail: jaafar@upnm.edu.my

doi: <http://dx.doi.org/10.4314/jfas.v9i3s.33>

1. INTRODUCTION

Heart abnormality activity or heart disease is a very common disease that people in the world have, especially in Malaysia. Heart abnormality or heart disease will be occurred when the blood flow into the heart is blocked by a blood clot [1]. The present of the blood clot will



make the artery, which connected to the heart become narrow. If the blood clots totally block the main artery to the heart, blood supply will eventually stopflowing into it. This will make the heart 'starved' for oxygen because oxygen is important for pumping the blood back to the whole body. In a short time, the heart muscle will become weak and supply less blood to the whole body because of lack of oxygen; this causes the heart attack [2, 23].

One of the most effective ways to detect the heart abnormality is by using electrocardiogram (ECG). ECG is the important source for doctor to detect the any cardiac abnormality or heart disease. It works as a detector of any electrical activity on the surface of a body that generated by the heart [5]. The measurement that is being recorded by ECG is collected by placing several skin electrodes at some specific area of a body.

However, there are known pitfalls in electrocardiography [3]. In particular, an ECG may appear normal despite the presence of serious heart problems. In fact, because the ECG is just a record at one point in time like snapping a photograph, a patient with a normal ECG may nonetheless (rarely) develop a major change indicative of a heart attack 10 minutes after the test. That is because while the patient may have a history of cardiac symptoms, those symptoms might not have been present when the test was performed. As a result, the test may appear normal in a patient with severe heart disease as it cannot identify a narrowing within an artery unless it is evaluated during physical stress or during a heart attack. Similarly, though the ECG is the definitive test in patients with a rhythm disturbance, if the patient is not in the midst of an arrhythmia when the test is performed, it just may appear normal despite the existence of a serious problem.

Hence, by extracting some features from the ECG signal such as amplitude and duration of P, QRS and T wave, it can be used as an input vectors for the MLP network to determine the heart abnormality activity accurately. Numbers of activation function are used in order to activate the MLP network.

1.1.Literature Review

Artificial Neural Network (ANN) is one of the studies of Artificial Intelligence and is a new computing technology in the field of computer science study. ANN also considers the integration of neural networks with another computing method such as fuzzy logic to

maximize the interpretation ability of data.

Neural networks mostly used for problem solving in pattern recognition, data analysis, control and clustering [4]. ANN has various features including high processing speeds and the ability to learn the solution to a problem from a set of data or examples. Multilayer Perceptron (MLP) is the most suitable and referred neural networks in the pattern recognition detection.

This network can be trained to form various decision surfaces in the input space [4, 6-7].

Artificial Intelligence (AI) is the computer science study about making a machine or developing software that has the ability like human intelligent [3]. It is an integration of the computer science study and physiology. AI is mainly about making computers behave like humans. AI tries to solve the complex problems in more human like fashion and in minimum time compared what a human takes. That is why it is called as Artificial Intelligence.

Intelligence can be defined as the ability to think and imagine, creating, memorizing and understanding, recognizing patterns, making choices adapting to change and learn from experience. The major objective for this field of study is creating intelligence, either in trying to act like human-intelligent or not. AI is also known as “the study and design of intelligent agents”, where an intelligent agent is a system that observe the environment and act to achieve the maximum rate.

ANN has various features including high processing speeds and the ability to learn the solution to a problem from a set of data or examples. Multilayer Perceptron (MLP) [22] and Radial Basis Function (RBF) networks are the most suitable and referred neural networks in the pattern recognition detection. The MLP network consists of different kind of Multilayer Feed Forward neural network. This network can be trained to form various decision surfaces in the input space. However, MLP training processes usually consume large computation time and local minima problem often occur [4].

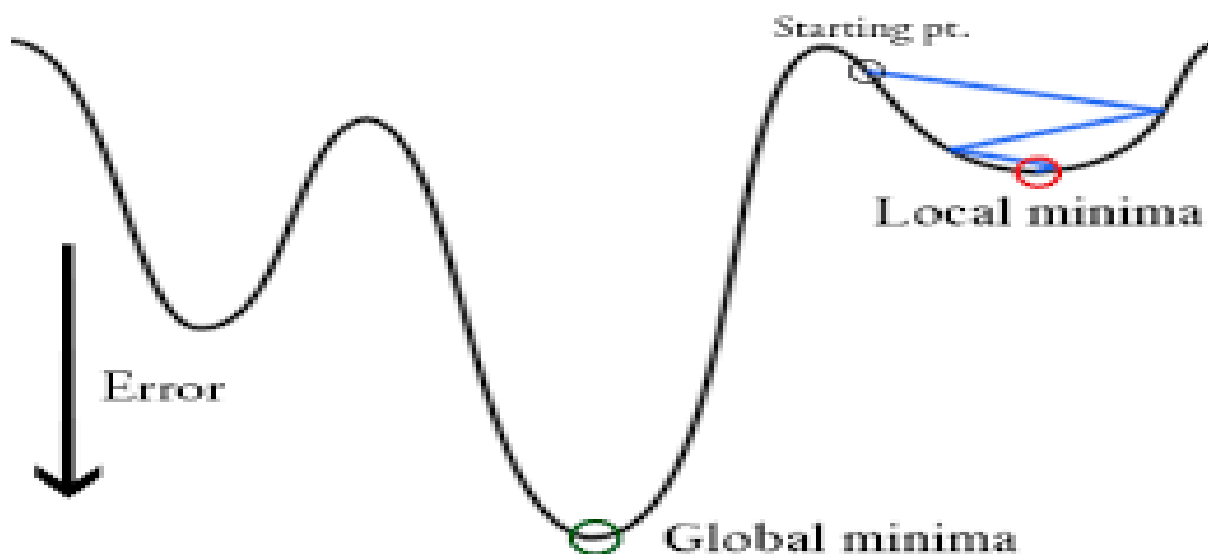


Fig.1.Local minima

1.1.1. Multilayer Perceptron (MLP)

An MLP network is a feed-forward artificial neural network that maps a set of input onto a set of appropriate output [8]. The network comprises of an input layer, one or more hidden layer, and an output layer. The number of nodes in the input and output layers is dependent on the number of input and output variables, respectively [9]. It has been proven that one hidden layer is sufficient for approximation of continuous function up to a certain acceptable accuracy [10]. A transfer function which also known as the system function or network function is a mathematical representation in terms of spatial or temporal frequency of the relation between the input and output. The transfer functions usually have a sigmoid shape but they may also take the form of other non-linear functions, piecewise linear functions or step functions [11].The input vector x_i is converted to a vector of hidden variables u_j using activation function φ_1 . The output u_j of the j^{th} node in the hidden layer can be expressed as

$$u_j = \varphi_1 + \left(\sum_{i=1}^n w_{ij} + \theta_j \right) \tag{1}$$

The weight and bias of the connection between the j^{th} node in the hidden layer and the i^{th} input node are represented by w_{ij} and θ_j , respectively. Fig. 2 shows the standard connection for an MLP network (Mat Isa et al, 2008).

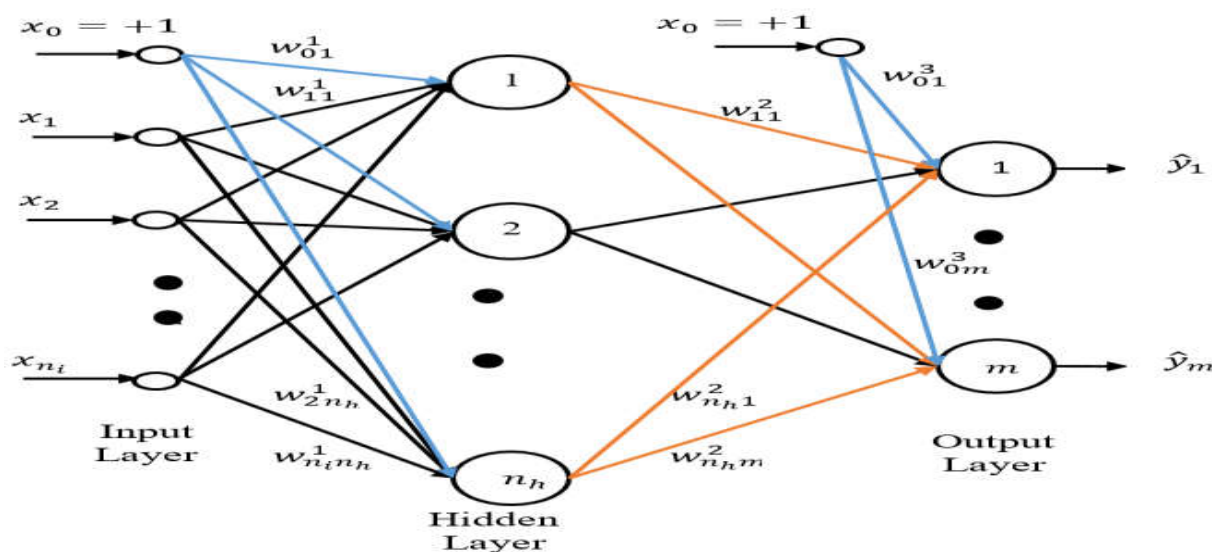


Fig.2. Structure of MLP network

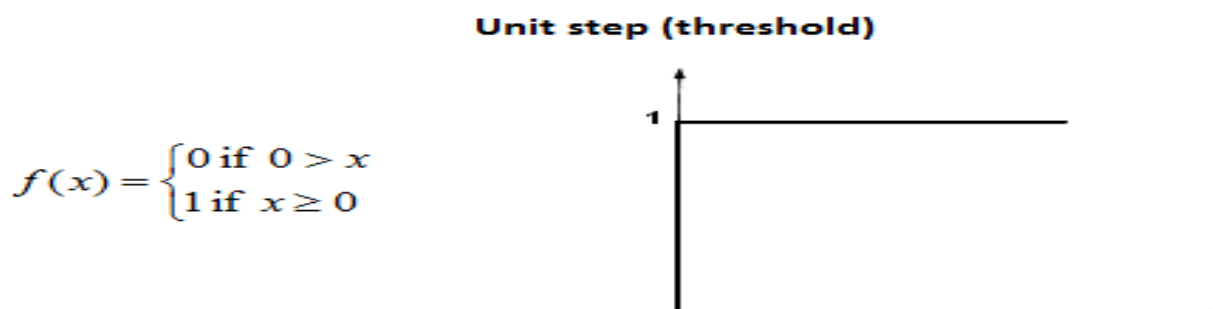
The network allows the inputs to be connected directly to the output nodes via weighted connections to form a linear model, which is in parallel with the conventional non-linear MLP model [12]. Therefore, the k th output from the MLP network can be expressed by

$$output = \varphi_2 + \left(\sum_{j=1}^h w_{jk} + \theta_k \right) \tag{2}$$

1.1.2. Activation Function

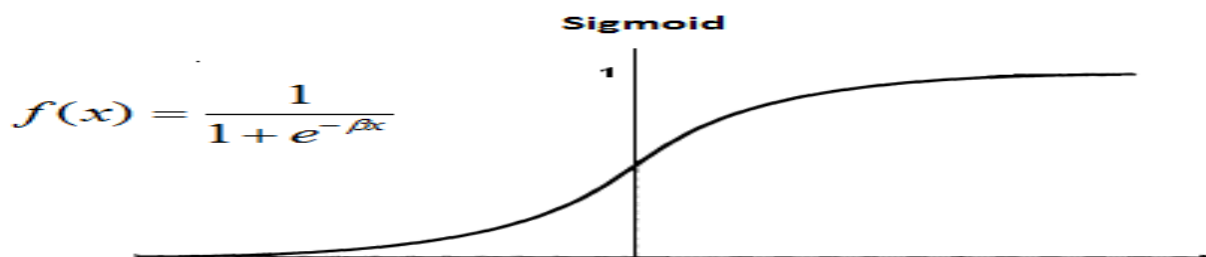
One of important unit in the structure of neural network is their net inputs, using a scalar-to-scalar function called as activation function or threshold function or transfer function. The output of the result value called as unit's activation. An activation function act as the limiter to the amplitude of the output of a neuron. Enabling in a limited range of functions is called squashing functions [13-14]. It squashes the permissible amplitude range of the output signal to some finite value. Some of the most commonly used activation functions are to solve non-linear problems. These functions are:

1) Unit Step Function (US): Activation function US function is given as follows:



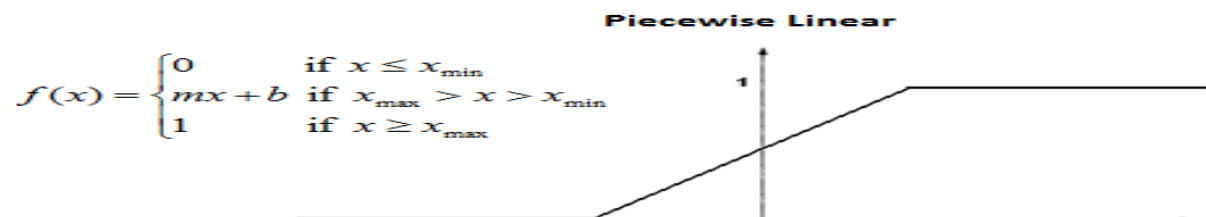
Based on the Heaviside step function or the function always denoted with H or θ (but sometimes u (signal processing), 1 or $\mathbb{1}$). The function is a discontinuous type which the value is zero for negative statement while one for positive statement. It is an example of the general class of step functions, all of function can be represented as a linear combinations of translations.

2) Unipolar Sigmoid Function (SF): Activation function of PS function is given by:



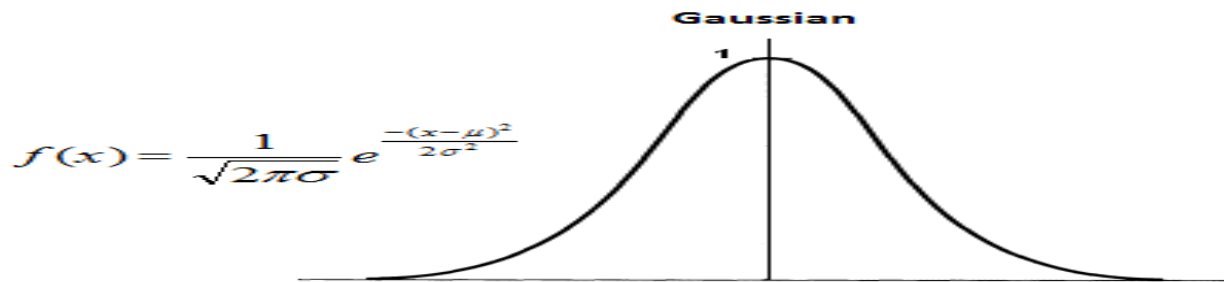
One of the most commonly used functions is the Unipolar Sigmoid transfer function. This transfer function takes the input (which may have any value between plus and minus infinity) and squashes the output into the range 0 to 1. The sigmoid transfer function is commonly used in multilayer networks that are trained using the back propagation algorithm, in part because this function is differentiable.

3) Piecewise Linear (PF): In mathematics, a piecewise linear function is a real-valued function defined on the real numbers. The graph is composed of straight-line sections.



In a piecewise-defined function, each of whose pieces is an affine function. Sometimes the function is assumed to be continuous, in that situation the graph is a polygonal curve.

4) Gaussian Function (GF): The Gaussian graph is a characteristic symmetric "bell curve" shape or normal tabulation.



The function is widely used in statistics to describe the normal distributions while in signal processing to define Gaussian filters. The function also applied to the in image processing application where two-dimensional Gaussians are used for Gaussian blurs.

1.1.3. Electrocardiogram (ECG)

The ECG is a medical tool that used to assess and recorded the electrical activity and muscular functions of the heart. Although it is a simple test to perform, a lot amount of training is required for the interpretation of the ECG tracing. The heart has two stage electrical pumps and the heart's electrical activity can be measured by placing electrodes on the skin [15]. The electrocardiogram can measure the rate and rhythm of the heartbeat, and also provide indirect evidence of blood flow to the heart muscle. The interpretation of ECG waves of normal ECG readings, it consists of waveform components which indicate electrical events during one heartbeat. These waveforms are labelled P, Q, R, S, T and U [16]. P wave is the first short upward movement of the ECG tracing. It indicates that the atria are contracting, pumping blood into the ventricles. The QRS wave normally beginning with a downward deflection, Q; a larger upwards deflection, a peak (R) and then a downwards S wave. The QRS complex represents ventricular depolarization and contraction. The PR interval indicates the transit time for the electrical signal to travel from the sinus node to the ventricles. T wave is normally a modest upwards waveform, representing ventricular repolarization.

1.1.4. ECG Signal Morphology

A healthy individual with a normal heart cycle is capable of displaying an ECG complex that consists of three main waves P, QRS and T. Put simply, the P wave corresponds to atrial contraction, the QRS wave corresponds to the electrical signal travelling through the ventricular muscle which causes ventricles contraction and the T wave corresponds to repolarization of the ventricles. In AF, the electrical actually through the atria is disorganized. Therefore, the two atria do not contract in unison. Therefore, there is no P wave on the ECG.

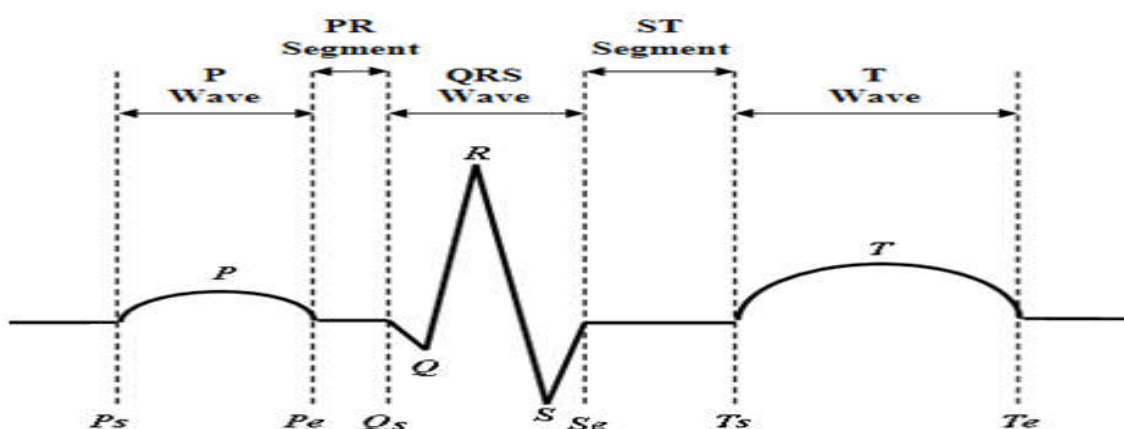
Table1 shows the ECG signal morphology that occurs in the heart.

Table 1. ECG signal complex

Wave	Start	Peak	End
P	Ps: Start point of P wave	P: Peak of P wave	Pe: End point of P wave
QRS	Qs: Start point of Q wave	R: Peak of R wave	Se: End point of S wave
T	Ts: Start point of T wave	T: Peak of T wave	Te: End point of T wave

1.1.5.Fiducial Point Detection

A fiducial point is a point corresponding to the peak or a location of the three main waves in an ECG complex, which can be used as a reference point. In an ideal complex, there are at least nine fiducial points which can be identified. However, the boundaries of each wave cannot be distinguished by the human eye and therefore no clear definition can be given regarding the beginning or end of each of the main waves. In [17] proposed a technique to detect the locations of the three major waves in the ECG signal by calculating the signal amplitude and width as well as the slope. In [18] also proposed tracking the local maximum and minimum radius of curvature. The combination of these two methods is able to improve the sensitivity and accuracy of ECG wave detection. Many manufacturers of ECG recording equipment apply a combination of these techniques into their devices.



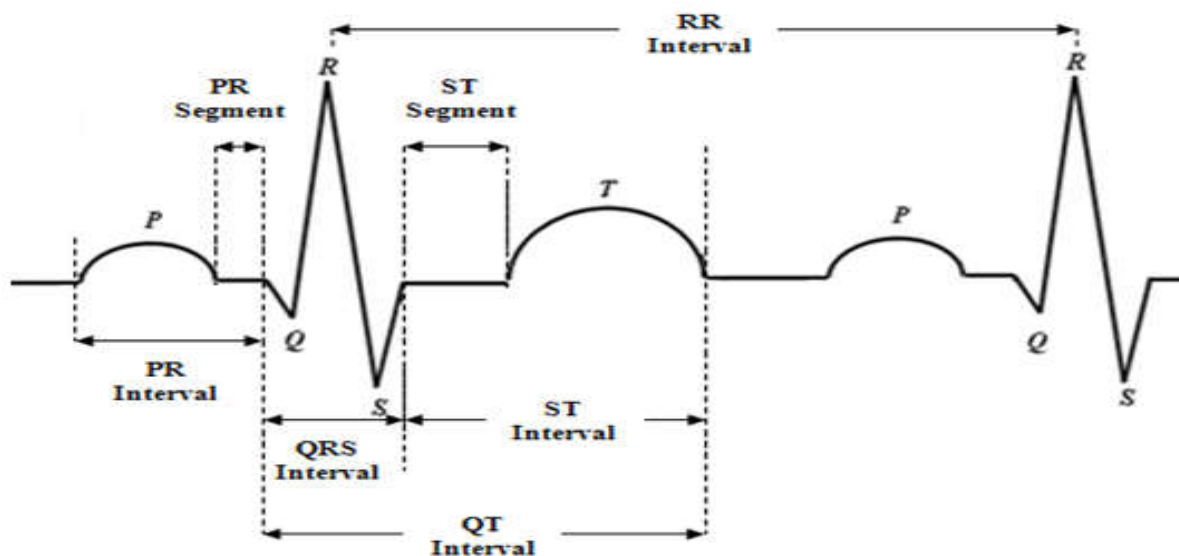


Fig.3. The fiducial points taken from a complex ECG signal

Fig. 3 shows the identified fiducial points of a complex of ECG signal. These fiducial points are often used in classifying cardiac abnormalities including PR interval, i.e. the distance measured from beginning of point P to peak R. The PR segment is measured starting at endpoint of P to the starting point of Q. The QRS interval is measured starting at point Q up to end of point S. Next, the QT interval and the ST intervals are measured based on the distances from beginning of point Q to end of point T and from end of point S to end of point T respectively. The ST segment is measured from the endpoint of S to the start point of T. The RR interval is measured from the R peak to the next R peak. In addition, the amplitudes of the P, QRS and T peaks also contribute to the fiducial points.

1.2. Modified Recursive Prediction Error

Recursive prediction error algorithm (RPE) is originate derive by [19-20] modify the algorithm in order to train the MLP networks. RPE is a Gauss-Newton (GN) type algorithm capable to produce better performance than a steepest gradient descent (SGD) type algorithm such as back propagation algorithm. However, the GN type algorithm is time consuming than SGD type algorithm since the computational complexity of GN. In this study, the optimization on momentum and learning rate of the RPE algorithm has been further improved the convergence rate. Differ from constant in [20], in [21] did some modification of the RPE algorithm to minimize the cost function as shown below.

$$\begin{aligned}
 & J(\hat{\Theta}) \\
 &= \frac{1}{2N} \sum_{t=1}^n \varepsilon^T(t, \hat{\Theta}) \Lambda^{-1} \varepsilon(t, \hat{\Theta}) \tag{3}
 \end{aligned}$$

updates the estimated parameter vector, $\hat{\Theta}$ (consists of w's and b's), recursively using GN algorithm:

$$\hat{\Theta} = \hat{\Theta}(t - 1) + P(t)\Delta(t) \tag{4}$$

with

$$\Delta(t) = \alpha_m(t)\Delta(t - 1) + \alpha_g(t)\psi(t)\varepsilon(t) \tag{5}$$

Where $\varepsilon(t)$ is the prediction error while Λ is the symmetric positive definite $m \times m$ matrix respectively. m in the Equation (3) is the number of output nodes while $\alpha_m(t)$ is the momentum and $\alpha_g(t)$ is the learning rate, respectively. $\alpha_m(t)$ and $\alpha_g(t)$ are arbitrarily assigned with values range between 0 and 1 and the typical value of $\alpha_m(t)$ is closed to 1 while $\alpha_g(t)$ is closed to 0, respectively. In this work, $\alpha_m(t)$ and $\alpha_g(t)$ are varied to improve the convergence rate of the RPE algorithm according to:

$$\alpha_m(t) = \alpha_m(t - 1) + a \tag{6}$$

and

$$\alpha_g(t) = \alpha_m(t)(1 - \alpha_m(t)) \tag{7}$$

where $\alpha_m(t)$ is initialized $0 \leq \alpha_m(0) \leq 1$ with a is a small constant, typically $a = 0.01$. The gradient of the one step ahead predicted output represented by ψ by respect parameter of the network:

$$\psi(t, \Theta) = \left[\frac{d\hat{y}(t, \Theta)}{d\Theta} \right] \tag{8}$$

Matrix $P(t)$ in Equation (9) is updated recursively according to:

$$P(t) = \frac{1}{\lambda(t)} \left[P(t - 1) - \frac{P(t - 1)\psi(t)\psi^T(t)P(t - 1)}{\gamma} \right] \tag{9}$$

with

$$\gamma = \lambda(t)I + \psi^T(t)P(t - 1)\psi(t) \tag{10}$$

$\lambda(t)$ in Equation (9) is the forgetting factor, $0 \leq \lambda(t) \leq 1$, updated by using [19] scheme as

$$\lambda(t) = \lambda_0\lambda(t - 1) + (1 - \lambda_0) \tag{11}$$

where λ_0 and the initial forgetting factor $\lambda(0)$ are the design values of 1 and 0.95. Initial value of $P(t)$ matrix, $P(0)$ is normally set to βI where I is the identity matrix and β is a constant between 100 to 10000. Too small β value might be slower the learning process while estimated parameters unable to converge well if β is too large. Hence, $\beta = 1000$ is selected to compromise between the two points.

2. EXPERIMENTAL

In this paper, six extracted features are taken from each complex ECG will be used as input vectors of the MLP network. These features are amplitude of P wave, QRS segment and T wave and the duration of P wave, QRS segment and T wave. Based on the features, it will have developed a MLP network with six input nodes. The selection of all this six features are based on the normal and abnormal ECG reading. The total data are 200 used with each condition is 100, taken from Massachusetts Institute of Technology Beth Israel Hospital (MIT-BIH) database.

All the data are injected to the developed network randomly in order to generalize the network. The total number of data must follow the rule of thumb, whilst data used must ten times greater than number of nodes. The extracted features from ECG signal that are set as an input vector, the neural network is used to classify the heart abnormality, whether it is normal or abnormal. The classification of the heart condition into normal and abnormal are represented as two output nodes as shown on Fig. 4. Each output node functioned as a classifier for the heart condition. The MLP network is trained by MRPE training algorithm.

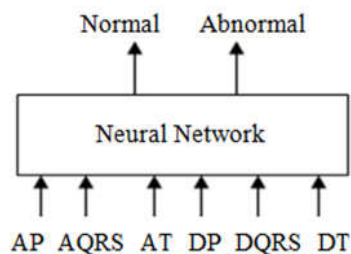


Fig.4. Heart abnormal detection block by using MLP network

Based on Fig. 4, the input vectors are represented by AP = Amplitude of P wave, AQRS = Amplitude of QRS complex, AT = Amplitude of T wave, DP = Duration of P wave, DQRS = Duration of QRS complex and DT = Duration of T wave.

3. RESULTS AND DISCUSSION

The MLP network is tested for its performance through computational experiment on six input vectors that had been extracted from the ECG signal from [16]. The six input vectors are the amplitude of P wave, QRS segment, T wave and the duration of P wave, QRS segment and T wave. The total data selected is 200; 100 for training purpose and 100 for testing sample. Moreover, the results are compared with MLP network activated by four different activation functions (US, SF, PF and GF).

The research applies two different analyses. The optimum structure analysis becomes the first analysis, which is implemented to determine the optimum number of hidden nodes that can produce the best heart abnormality classification. From the results, the MLP network activated by four different activation functions (US, SF, PF and GF) are shown in Table 2.

Table 2. Performance of Optimum Structure Analysis of MLP Network

Neural Network	Optimum Number of Hidden Nodes
MLP activated by US	8
MLP activated by SF	6
MLP activated by PF	16
MLP activated by GF	10

The result shows in Table 2 indices MLP network with SF activation function becomes the simplest network with six hidden nodes. In the other hand, the network capable to reach the optimum convergence after the 6th hidden node. MLP network activated by US reached the optimum performance with eight hidden nodes, follows by GF and PF with ten and 16 hidden nodes respectively. Although the MLP network with SF activation function gives the simplest network, but sometime the network trapped at the local minima. Increment number of hidden nodes may increase the complexity of the network. Furthermore, there are time consuming for the network with high complexity.

In order to ensure the performance of the networks, another analysis need to be done before selection of the best activation function is made. The second analysis, (i.e. the performance comparison analysis) is then applied based on the optimum structure of the MLP network (number of hidden node/nodes) obtained in the first analysis. The analysis is done based on

the accuracy of corrected classifications of heart abnormality. Table 3 shows the result of the analysis during the training and the testing phase.

Table 3. Performance of accuracy analysis of MLP network

Neural Network	Accuracy Percentage, (%)		
	Training	Testing	Overall
MLP activated by US	79.75	77.50	78.63
MLP activated by SF	86.92	85.73	86.33
MLP activated by PF	83.50	82.25	82.88
MLP activated by GF	83.88	82.50	83.19

The results tabulated in Table 3 shows performance of MLP network activated with different activation function. From the table, it shows the MLP network activated by SF give the highest accuracy with 86.92% and 85.73% for training and testing phase respectively. MLP with GF activation function come second with 83.88% and 82.50% accuracy for both training and testing phases. MLP network activated by PF and US also give significant results but no better than SF and GF.

The results show in Table 2 and Table 3 show the MLP network activated by the SF forms the simplest network architecture as it only requires six hidden nodes as compared to other activation functions (US, PF and GF) and capable to give a high accuracy results. MLP network with GF activation function give second best accuracy results although the network inherits a higher complexity network to be compared with MLP network activated by US. The MLP network activated by PF done a better classification process than MLP network with US as the activation function. Here, the MLP network with US activation function may be trapped in the local minima

4. CONCLUSION

This research proves the capability of neural network to classify the heart abnormality data. This research also shows that the MLP network is capable and reliable to classify the heart abnormality based on both amplitude and duration of the P, QRS and T peaks of ECG complex by using four different activation functions. This research also notes that the

selection of appropriate activation function of the MLP network may improvement the performance of the network during the optimum structure analysis and the accuracy analysis.

5. ACKNOWLEDGEMENTS

This research work is supported by the UPNM Short Grant (UPNM/2014/GPJP/TK/07).

6. REFERENCES

- [1] National Institute of Health (NIH). What cause a heart attack? Maryland: NIH, 2013
- [2] Fakroul RH, John JS, Lykourgor P, Nik GND. EMG cancellation from ECG signals using modified NLMS adaptive filters. In IEEE Conference on Biomedical Engineering and Sciences, 2014, pp. 735-739
- [3] Myers R. The benefits and limitations of electrocardiography. Ontario: Medcan, 2013
- [4] Russell S., Norvig P. Artificial Intelligence: A modern approach. London: Pearson Education, 2010
- [5] Fakroul RH, John JS, Lykourgor P. Multi-classify hybrid multiyared perceptron (HMLP) network for pattern recognition applications. In International Conference on Artificial Intelligence Applications and Innovations, 2012, pp. 19-27
- [6] Dzati AR, Junita MS, Fakroul RH, Nor Ashidi MI. Multilayered perceptron (MLP) network trained by recursive least squares algorithm. In 1st International Conference on Computers, Communications, and Signal Processing with Special Track on Biomedical Engineering, 2005, pp. 288-291
- [7] Nor Ashidi MI, Fakroul RH, Fong WM, Dzati AR, Wan M WO, Kamal ZZ. Predicting quality of river's water based on algae composition using artificial neural network. In 4th IEEE International Conference on Industrial Informatics, 2006, pp. 1340-1345
- [8] Negnevitsky M. Artificial intelligence: A guide to intelligent systems. Essex: Addison Wesley, 2005
- [9] Funahashi KI. On the approximate realization of continuous mappings by neural networks. Neural Networks, 1989, 2(3):183-192
- [10] Mat Isa N A, Mashor M Y, Othman N H. An automated cervical pre-cancerous diagnostic

system. *Artificial Intelligence in Medicine*, 2008, 42(1):1-11

[11] Mat Isa NA, Wan M WMF. Clustered-hybrid multilayer perceptron network for pattern recognition application. *Applied Soft Computing*, 2011, 11(1):1457-1466

[12] Haykin S. *Neural networks and learning machines*. Ontario: Pearson, 2009

[13] Cybenko G. Approximation by superposition's of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 1989, 2(4):303-314

[14] Snyman J. *Practical mathematical optimization: An introduction to basic optimization theory and classical and new gradient-based algorithms*. Berlin: Springer Science and Business Media, 2005

[15] Rakesh K P, Thompson E G, George P. *Electrocardiogram*. Georgia: WebMD LLC, 2015

[16] *Practical Clinical Skills. ECG interpretation*. Pennsylvania: Medical Simulation and Training LLC, 2017

[17] Pan J, Tompkins J W. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 1985, 32(3):230-236

[18] Boulgouris N.V., Plataniotis K.N., Tzanakou E.M. *Biometrics: Theory, methods, and applications*. Ontario: John Wiley and Sons, 2009

[19] Ljung L., Soderstrom T. *Theory and practice of recursive identification*. Cambridge: MIT Press, 1983

[20] Chen S, Cowan C F N, Billings S A, Grant P M. A parallel recursive prediction error algorithm for training layered neural networks. *International Journal of Control* 1990, 51(6):1215-1228

[21] Mashor M Y. Modified recursive prediction error algorithm for training layered neural network. *International Journal of the Computer, the Internet and Management*, 2003, 11(2):24-36

[22] Ihsan M Y, Azlee Z, Rozita J, Megat S A M A, Rahimi B, Abu H A H, Zairi I R. Comparison between cascade forward and multi-layer perceptron neural networks for NARX functional electrical stimulation (FES)-based muscle model. *International Journal on Advanced Science, Engineering and Information Technology*, 2017, 7(1):215-221

[23] NorM NM, Jailani R, TahirN M, YassinI M, Zairi I R, Rahmat H. EMG signals analysis of

BF and RF muscles in autism spectrum disorder (ASD) during walking. *International Journal on Advanced Science, Engineering and Information Technology*, 2016, 6(5):793-798

How to cite this article:

Adnan J, Daud N G N, Mokhtar A S N, Hashim F R, Ahmad S, Rashidi A F, Rizman Z I. Multilayer perceptron based activation function on heart abnormality activity. *J. Fundam. Appl. Sci.*, 2017, 9(3S), 417-432.