Jurnal Kejuruteraan SI 4(2) 2021: 147-152
https://doi.org/10.17576/jkukm-2021-si4(2)-22

Cardiac Abnormality Prediction using Tansig Based Multilayer Perceptron

(Peramalan Keabnormalan Kardiak Menggunakan Logsig Berasaskan Perceptron Berbilang Lapisan)

Sibani Priyadarshini Mohanty[a], Syahrull Hi-Fi Syam Ahmad Jamil[b], Jailani Abdul Kadir[c], Mohd Salman Mohd Sabri[d] & Fakroul Ridzuan Hashim[d,*]

[a]*Department of Biomedical Engineering, University of Strathclyde, United Kingdom*
[b]*Department of Mathematics, Science & Computer, Polytechnic of Tuanku Syed Sirajuddin, Perlis, Malaysia*
[c]*Department of Electrical Engineering, Polytechnic of Tuanku Syed Sirajuddin, Perlis, Malaysia*
[d]*Faculty of Engineering, National Defence University of Malaysia, Kuala Lumpur, Sg. Besi Camp. Kuala Lumpur, Malaysia*

**Corresponding author: fakroul@upnm.edu.my*

Received 14 March 2021, Received in revised form 13 April 2021
Accepted 30 September 2021, Available online 31 October 2021

ABSTRACT

An artificial neural network (ANN) is a network designed with adaptation to a computer system. The developed computer system will perform functions oriented to the way the brain works (neuron concept). This study is an extension to the study of the suitability of ANN to be applied in numbers of applications, especially in the field of medical engineering. ANN has been widely being used in medicine, ANN is widely applied in education, research, and even decision making. In this study, ANN will be trained for pre-testing to predict the cardiac abnormalities symptom based on selected reference parameters. This reference parameter is better known as the input parameter to the ANN to detect cardiac abnormalities, among which are the of the height of peak/wave (amplitude) and time occurrence of peak/wave (duration of time) extracted from the electrocardiogram (ECG) signal. A complete ECG complex contains a P peak, a QRS wave, and a T peak. For each P peak, QRS wave, and T peak, amplitude height and duration will be measured to serve as input parameters. This makes six parameters defined as inputs to the ANN. This study has used a Multilayer Perceptron (MLP) network as ANN structure by being trained using three different training algorithms namely Backpropagation (BP), Lavenberg Marquardt (LM) and Bayesian Regularization (BR). At the end of the study, it showed the MLP network which by BR training algorithm gave the highest accuracy prediction (94.04%), followed by LM (92.95%) and BP (88.77%). In this study all MLP networks were activated using the Tansig activation function.

Keywords: MLP network; ECG signal; cardiac abnormality; amplitude; duration; tansig activation function

ABSTRAK

Rangkaian neural buatan (ANN) merupakan rangkaian yang direkabentuk dengan adaptasi kepada sistem komputer. Sistem komputer yang dibangunkan akan melakukan fungsi berorientasikan cara otak bekerja (konsep neuron). Kajian ini merupakan lanjutan kepada kajian kesesuaian ANN untutk diterapkan dalam pelbagai jenis masalah, terutama dalam bidang kejuruteraan bioperubatan. ANN telah banyak digunakan dalam bidang perubatan, ANN banyak diterapkan dalam pendidikan, penyelidikan dan juga membuat keputusan. Dalam kajian ini, ANN akan dilatih bagi pra-ujian untuk melakukan pengesanan keabnormalan jantung berdasarkan beberapa parameter rujukan. Parameter rujukan ini lebih dikenali sebagai parameter masukan kepada ANN bagi mengesan keabnormalan jantung antaranya adalah amplitud dan tempoh masa yang diekstrak dari denyut isyarat Elektrokardiogram (ECG). Satu kompleks ECG yang lengkap mengandungi puncak P, gelombang QRS dan puncak T. Untuk setiap puncak P, gelombang QRS, dan puncak T, tinggi amplitud dan durasi akan diukur untuk dijadikan sebagai perameter masukan. Perkara ini menjadikan enam parameter ditetapkan sebagai masukan kepada ANN. Kajian ini telah menggunakan rangkaian Perceptron Berbilang Lapisan (MLP) sebagai struktur ANN dengan dilatih menggunakan tiga algoritma latihan yang berbeza iaitu Bayesian Regularization (BR), Lavenberg Marquardt (LM) dan Backpropagation (BP). Dihujung kajian ini menunjukkan rangkaian MLP yang oleh algoritma latihan BR memberikan ramalan berketepatan tertinggi (94.04%), diikuti oleh LM

(92.95%) dan BP (88.77%). Dalam kajian ini kesemua rangkaian MLP diaktifkan menggunakan fungsi pengaktifan Tansig.

Kata kunci: Keabnormalan ECG; amplitud; tempoh; rangkaian MLP; fungsi pengaktifan tansig

INTRODUCTION

Cardiovascular disease has shown an increasing indicator which lead causes of death among Malaysian, followed by depression. The severity of this disease can be minimized if early detection, such as pre-diagnosis or pre-monitoring can be performed as soon as possible. Early detection of abnormalities due to cardiac activity known as arrhythmia may provide useful information for physicians to determine the level of treatment needed (Barker et al. 2003; Custer & Rue 2008; Mark 1998). Monitoring and prevention of heart attack episodes is important on the basis of information provided by the electrocardiogram signal (ECG). The P, QRS, and T peaks are included in the ECG signal for interpretation. Atrial depolarization results in a P peak ventricular depolarization produces a complex QRS and ventricular repolarization marks a T peak. All of these peaks are results of cardiac electrical activities (Barker et al. 2003; Custer & Rue 2008).

Analysis of ECG signals is a common technique used in clinical trials to detect any activity in the heart. The cardiac rate monitoring industry started at the end of the 1950s. Literally, figures for each year indicate a rise in the amount of documents reviewed. Work on automated ECG signal processing since then has intensified and paved way for an area of innovation known as biomedical engineering. To date, different forms of ECG signal analysis algorithms have been developed and implemented during the 80's. Most of the works concentrate on the QRS complex identification (Mark 1998). However, there are many parameters in the ECG signal recording that can provide valuable details for showing patient status (Houghton & Gray 2012; Plesnik et al. 2010; Hashim et al. 2019).

The ANNs are models which are designed on the basis of how neurons function and interact in the brain (Haykin 1994). They are equipped with functionality that can address linear and non-linear problems, data input-output, and analogies to neurobiology. ANN has been commonly used in a variety of areas including science and technology, banking, and schooling. ANN is widely used in engineering for high precision pattern recognition, data classification, system identification, image processing, and more.

This study uses information from ECG signals extracted using MLP networks. This study uses three different training algorithms namely Bayesian Regularization (BR), Lavenberg Marquardt (LM), and Backpropagation (BP) (Payal et al. 2013; Kaensar 2013). The system is architectured by selecting algorithms in order to identify the patient's condition either based on amplitude and duration parameters extracted from P, QRS, and T peaks under normal conditions or otherwise. Here, the MLP network is activated by the selected Tansig activation function.

MULTILAYER PERCEPTRON NETWORK

ANN is a model that is constructed from simple function units that mimic the behaviour of a single neuron; the tiny cells that are the building blocks of our brain. These computer analogy of biological neurons, called artificial neurons, are combined to form networks that inherit the nature and architecture of the human brain. Such ANNs can be used as statistical models for non-parametric grading or regression, data classification and non-linear function estimation. ANN is also used as a framework for the behavior of human neuronal biological models. ANNs are taught to deliver high-quality predictive results. ANNs tend to work with the ability to suit the output of the human brain, however it should not be assumed to be an alternative to brain activity. The developed model can therefore, be an alternative to the system. One of the most widely used software for this specific purpose is the MLP network (Gas 2003; Sivaram & Hermansky 2011; Mat Isa et al. 2006).

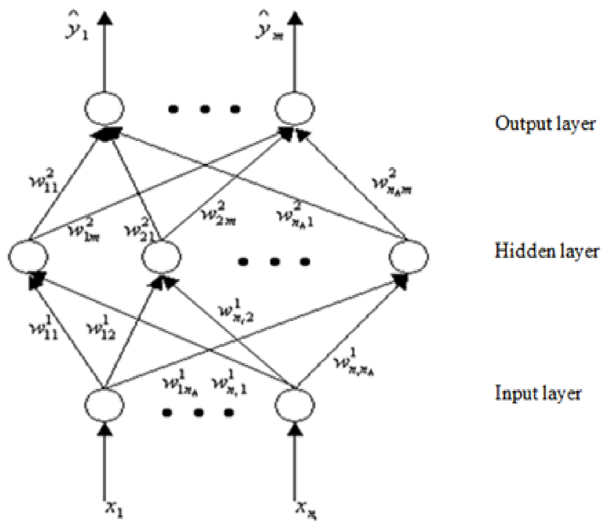

FIGURE 1. MLP network architecture design
Source: Ramli et al. (2005)

In 1958, Rosenblatt developed a model of artificial neurons in his study called the Perceptron model (Haykin 1994). He positioned multiple neurons in sequence of layers creating a network as shown in Figure 1. In other parts of this paper, ANNs are referred to as MLP networks. The MLP network structure, referred to in Figure 1, consists of an input layer, the hidden layer, with n_h neurons, and an output layer with m neurons, respectively. The output is mathematically expressed as

$$\sum_{j=1}^{n_h} w_{jk}^2\, \partial\left(\sum_{i=1}^{n_i} w_{ij}^1 x_i^0(t) + w_{k0}^1 b_j^1\right) \tag{1}$$

$$for\ 1 \le j \le n_h \text{ and } 1 \le k \le m$$

In Equation 1, w_{ij}^1 is the weight that links the input layer to the hidden layer, w_{jk}^2 represents the weight that connects the hidden layer to the output layer. The b_j^1 in Equation (1) represents the hidden node threshold, whereas the input parameter x_i^0 feeds the MLP network through the input layer. Throughout this study, the Tansig activation function has been chosen to activate the MLP network, which in Equation 1 denotes as $\partial(\cdot)$. Referring to Equation (1), the values of w_{ij}^1 , w_{jk}^2 and b_j^1 are calculated on the basis of a suitable algorithm and set to a minimum vector, but is allowed to recur at each iteration. To train the MLP network, three separate training algorithms such as BP, LM, and BR algorithms are selected. Tansig's transfer function is associated with a bipolar sigmoid. The output of the Tan-Sigmoid activation function ranges from -1 to +1. Figure 2 demonstrates Tansig's activation function for network activation by MLP. The threshold must be between-1 and +1 at any given point.

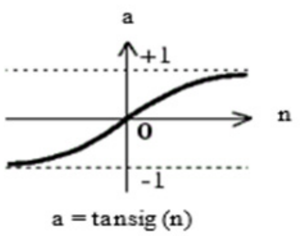

FIGURE 2. Tansig activation function
Source: Mitra et al. (2000)

BACK PROPAGATION (BP) ALGORITHM

ECGs are measurement of electrical impulses from electrical stimulation or electrical signals in the heart. The ECG signals are measured, recorded, and trained using the BP algorithm. The algorithm normally trains the network using the steepest descent technique. The BP algorithm is capable of measuring the value of derivatives while making effective weight changes depending on the parameter / protocol defined as the learning process (Kaensar 2013). BP connecting the weight between the j^{th} neuron of the hidden layer directly in order to enter the i^{th}-layer is updated as:

$$w_{ji}(t) = w_{ji}(t-1) + \Delta w_{ji}(t) \tag{2}$$
$$b_j(t) = b_j(t-1) + \Delta b_j(t)$$

The updated weight, Δw_{ji}(t) and Δb_j(t) given by:

$$\Delta w_{ji}(t) = \eta_w \rho_j(t)\, x_i(t) + \alpha_w \Delta w_{ji}(t-1) \tag{3}$$
$$b_j(t) = \eta_b \rho_j(t) + \alpha_b \Delta b_j(t-1)$$

where the w represents the weight and b represents the threshold, respectively. The α_w and α_b are the constant momentum values to determine that control the influence of the past value on the current. The η_w and η_b are noted as training/learning rates while ρ_j (t) is the signal error of the j^{th} neuron (of the hidden layer), propagated back to the network.

The output node error is shown in Equation 4, with the activation function being treated as a linear parameter:

$$\rho(t) = y_k(t) - \hat{y}_k(t) \tag{4}$$

where y_k (t) is the predicted output and is$\hat{y}_k(t)$esired output. The output at the hidden layer is

$$\rho_j(t) = F'(x_i(t)) \sum_j \rho_j^k(t)\, w_{jk}^2(t-1) \tag{5}$$

or simply $F(x_i$ (t)) is derived from $F(x_i$ (t)) with respect to x_i (t). The BP algorithm uses the gradient descent method to find the minimum of the error function is by taking small steps down the gradient. By making small updates to the weights, the algorithm slowly converges to a minimum. However, as the function approximated by the ANN is often non-convex, the algorithm can get stuck in one of the many local minima and does not guarantee a global minimum. This also makes it sensitive to the selected parameters of the user (Payal et al. 2013).

LAVENBERG MARQUARDT (LM) ALGORITHM

Levenberg Marquardt (LM) training algorithm operates on a deterministic optimization gradient. LM is a superior variant of the BP algorithm. The profitability of the LM algorithm is compared to the BP algorithm is seen during training of the MLP network as it provides a faster convergence rate than the BP algorithm while maintaining the relative stability (Payal et al. 2013). The LM training algorithm is constructed, based on the quasi-Newton method, to increase the second-order training speed and bypass the Hessian matrix. Since the LM function represents the sum of the squares, the Hessian matrix is calculated using:

$$H = J^T J \tag{6}$$

and the gradient of the matrix is calculated as:

$$g = J^T \rho \tag{7}$$

where J is a Jacobian matrix that determines network error based on network weight and bias. However, the matrix can be determined using standard BP techniques and is less complex than calculating the Hessian matrix (Payal et al. 2013). The LM algorithm applies Newton-like equation estimator to update the equations in the Hessian matrix as:

$$\Delta w = -[J^T J + \mu I]^{-1} J^T \rho \tag{8}$$

where updated weight, Δw is controlled by μ. The Hessian is small (approaches to zero) in order to update Newton-like equation currently represented as a scalar, μ. Newton estimator process has been proven to converge more than he BP algorithm for more accurate estimation and performance with a minimal error. The μ decreases at every successful step (minimizing the error), but as the μ increases with a tentative step function (increasing error), it increases the performance. At each iteration, however, the minimum error must be reduced (Payal et al. 2013).

PARTICIPANTS

In a study by Thomas Bayes, he introduced a training/learning algorithm known as D. Bayes rule. Given the data D, the Bayes rule can be implemented to obtain 'θ',which is known as the posterior probability [9]. Generally, the obtained posterior probabilities represents an entire distribution of possible values of θ. Bayes rule can be derived from

$$p(\theta|D) = \frac{p(D|\theta)\, p(\theta)}{p(D)} \tag{9}$$

with $p(\theta)$ being the initial probability of a parameter θ, continuing with $p(\theta|D)$ which is referred to as the likelihood of the data before the probability of data. This rule of thumb is then applied to MLP network which affects the probability distribution among the network weights. The w received from the training data, is used to calculate $p(w|D)$, which is the MLP network's weights. The posterior distribution thus, can be observed by,

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)}$$

$$= \frac{p(D|w)p(w)}{\int p(D|w)\, p(w)dw} \tag{10}$$

Referring to Bayesian equation, it shows that the BR training algorithm trains and optimizes the weights that may change our belief of the weights; from the initial $p(w)$ to the posterior $p(D|w)$ as shown in Figure 3.

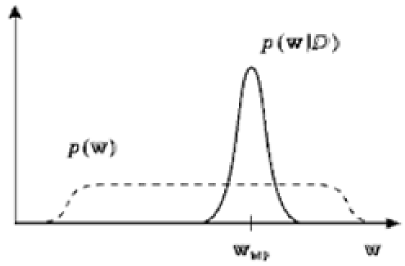

FIGURE 3. Change per weight to posterior/post weight

DATA SAMPLE

In this work, the ECG signal samples were obtained from the MIT-BIH Repository. Each signal sample was extracted into six distinct features, namely the amplitude and duration of the P-peak, the QRS complex, and the T-peak each (MIT-BIH). All these features were recorded as input parameters to the developed MLP network. The MLP network was then constructed using these six input parameters to chart the network input-output connection. A total of 1 000 datasets were used of which 800 datasets were used as training set and the remaining 200 as the testing set.

RESULTS AND DISCUSSION

This study implemented the same analysis as in the study (Adnan et al. 2018; Mitra et al. 2000). Analysis in this study focuses on two phases: prediction and classification. The first analysis is performed to obtain the optimal MLP network structure. The optimal structure of the MLP network is then fixed to 100 iterations in the first phase. In this phase, the analysis determined the optimal number of hidden nodes for the MLP network. Table 1 shows the analysis performance of the three types of BP, LM and BR training algorithms.

TABLE 1. MLP optimum structure

Training algorithm	Number of hidden nodes
BR	2
LM	4
BP	9

The second phase is a further analysis using the optimal structure derived from the first analysis results. The second analysis is performed to classify ECG signals based on whether the patient is in the normal condition or not. All three BP, LM, and BR algorithms are tested to determine the accuracy of the given predictions. Table 2 tabulates the performance on the prediction capability of the second analysis for both training and testing phases.

TABLE 2. Accuracy performance analysis

Training algorithm	Accuracy performance (%)		
	Training	Testing	Overall
BR	94.10	93.98	94.04
LM	93.75	92.15	92.95
BP	89.88	87.65	88.77

From Tables 1 and 2, it can be seen that the BR algorithm is able to train the MLP network by creating the simplest network, requiring only two hidden nodes. The results shown in Table 1 shows that BR is able to train MLP networks better than using the LM or BP training algorithms. LM and BP algorithms require up to four and nine nodes respectively to train the MLP network, while BR algorithm does the same with just two nodes; thus, using the simplest network architecture. Also, from Table 2, the MLP network trained by the BR training algorithm gives the highest overall accuracy of 94.04% when compared to the LM and BP networks with 92.95% and 88.77%, respectively. The result clearly shows that the performance accuracy for training and testing phases using LM training algorithm is 93.75% and 92.15%, respectively. Additionally, the BP training algorithm performance accuracy lies among 89.88% and 87.65% for both training and testing; which is the lowest among all the three training algorithms.

CONCLUSION

From the study that have been done, it shows the artificial naural network is able to result high accuracy prediction performance. High prediction performance shown by using extracted amplitudes and durations of ECG signal as the input parameter to the MLP network. On the other hand, the result given by BR training algorithm is better suited to design the network with the simplest structure giving more accurate prediction results than by using the LM and BP training algorithms.

ACKNOWLEDGEMENT

The project is funded by KPT.JPP.PPPP.700-1.jld.19(94) and (UPNM/2019/GPJP/2/TK/6) research grants. All the authors are the team members of the research grant.

DECLARATION OF COMPETING INTEREST

None

REFERENCES

Adnan, J., Nik Daud, N.G., Ahmad, S., Mat, M.H., Ishak, M.T., Hashim, F.R. & Ibrahim, M.M. 2018. Heart abnormality activity detection using multilayer perceptron (MLP) network. *AIP Conference Proceedings* 1930(1): 020013.

Barker, R.L., Burton, B.J. & Zieve, P.D. 2003. *Principles of Ambulatory Medicine.* 6th edition. Philadelphia: Lippincott, Wilkins & Williams.

Custer, J.W. & Rue, R.E. 2008. *The Harriet Lane Handbook*. 18th edition. Philadelphia: Mosby Elsevier Inc.

Gas, B. 2013. Self-organizing multilayer perceptron. *IEEE Transactions on Neural Networks* 21(11): 1766-1779.

Hashim, F.R., Adnan, J., Ahmad, K.A., Hi-Fi Syam, A.J.S., Januar, Y. 2019. MLP based Tan-Sigmoid activation function for cardiac activity monitoring. *MATEC Web Conf.*

Haykin, S. 1994. *Neural Network: A Comprehensive Foundation.* New Jersey: Prentice Hall.

Houghton, A.R. & Gray, D. 2012. Making Sense of the ECG. 3rd Edition. London: Hodder Education.

Kaensar, C. 2013. Analysis on the parameter of back propagation algorithm with three weight adjustment structure for hand written digit recognition. *10th International Conference on Service Systems and Service Management (ICSSSM).*

Mark, J.B. 1998. Atlas Cardiovascular Monitoring. New York: Churchill Livingstone.

Mat Isa, N.A., Hashim F.R., Mei, F.W., Ramli, D.A., Wan Omar, W.M. & Zamli, K.Z. 2006. Predicting quality of river's water based on algae composition using artificial neural network. *Industrial Informatics, 2006 IEEE International Conference.*

MIT-BIH Arrhythmia Database – PhysioNet cited from www.physionet.org/physiobank

Mitra, P., Mitra, S. & Pal, S.K. 2000. Staging of cervical cancer with soft computing. *IEEE Transaction on Biomedical Engineering* 47(7): 934-940.

Payal, A., Rai, C.S. & Reddy, B.V.R. 2013. Comparative analysis of Bayesian regularization and Levenberg-Marquardt training algorithm for localization in wireless sensor network. *15th International Conference on Advanced Communication Technology (ICACT).*

Plesnik, E., Malgina, O., Tasic, J.F. & Zajc, M. 2010. ECG signal acquisition and analysis for telemonitoring. *15th IEEE Mediterranean Electrotechnical Conference (MELECON).*

Ramli, D.A., Mohamad Saleh, J., Hashim, F.R. & Mat Isa, N.A. 2005. Multilayered perceptron (MLP) network trained by recursive least squares algorithm. *Computers, Communications, & Signal Processing with Special Track on Biomedical Engineering* 288-291.

Sivaram, G.S.V.S. & Hermansky, H. 2011. Multilayer perceptron with sparse hidden outputs for phoneme recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).*