

# A New Quantum Radial Wavelet Neural Network Model Applied to Analysis and Classification of EEG Signals

Saleem M. R. Taha, Ph.D.  
Department of Electrical Engineering  
College of Engineering, University of Baghdad  
Jadiryah, Baghdad, Iraq

Abbas K. Nawar, M.Sc.  
Department of Electrical Engineering  
College of Engineering, University of Baghdad  
Jadiryah, Baghdad, Iraq

## ABSTRACT

In this paper, a new model of multi-level transfer function radial wavelet neural network using quantum computing is achieved. This model is applied to analyze and classify the electroencephalographic (EEG) signals. The independent component analysis (ICA) is used as processing after normalization of these signals. Some features are extracted from the data using the clustering technique (CT). A new factor that combines the accuracy and the time of classification is suggested to evaluate the performance of the proposed model with other previous models. This factor represents the accuracy to time ratio (ATR). The average accuracy of the proposed quantum radial wavelet neural network (QRWNN) model is 90.5619125% at 50 minutes. The ATR value is 1.8112, which shows the superiority of the proposed model.

## General Terms

Pattern Recognition, Theory, Algorithms

## Keywords

EEG Signals, Neural Networks, Quantum Computing, Radial Basis Functions, Wavelet Transforms, Radial Wavelet Neural Networks

## 1. INTRODUCTION

In the 1990s the concept of quantum neural network (QNN) was developed. It is based on quantum mechanics to overcome the shortcomings and inadequacies of traditional neural network model. Since then, many researchers in different fields made their QNN model. It is based on multi-incentive function making use of quantum superposition state in quantum theory. This QNN model has been successfully used in many applications giving high classification results in pattern recognition with uncertainty and cross-data.

The quantum wavelet neural network (QWNN) model has been introduced [1]. It is a neural network with multi-resolution based on the wavelet analysis theory and the theory of quantum superposition. The incentive function of the hidden layer in QWNN model uses non-linear wavelet instead of the non-linear sigmoid function in QNN. The superposition of selected wavelet function is called multi-wavelet incentive function. The QWNN model has advantage in accuracy and generalization ability for pattern classification.

This paper proposed a new quantum radial wavelet neural network (QRWNN) model. It combines Gaussian wavelet basis function with radial basis function and quantum neural network. Then it is applied to a very important application of analysis and classification of electroencephalographic (EEG) signals. The main advantage of the proposed QRWNN model is the capability to independently reveal the having of

uncertainty in the sample data and adaptively learn to quantify the existing uncertainty and reducing the computation processes during the training process, thus reducing the time of classification.

The EEG examination is a standard procedure in the study of all forms of cerebral disease. Its use in the study of patients with seizures and those suspected of having seizures becomes essential [2]. The EEG consists of a set of multichannel signals. The pattern of changes in signals indicates brain activities. There are five categories of these signals: delta, theta, alpha, beta, and gamma. Each of which is associated with a certain activity of the brain.

## 2. THE PROPOSED QUANTUM RADIAL WAVELET NEURAL NETWORK MODEL

Radial basis functions (RBFs) are embedded into a two-layer feed forward neural network (FFNN), such a network is characterized by a set of outputs. Each unit of the hidden layer implements a radial activated function. The output units implement a weighted sum of hidden unit outputs. In pattern classification application the input represents feature vectors, and the output corresponds to a class. QRWNN is a new field which combines the QNN with radial wavelet (RW). A motivation for using QRWNN is that the concept (quantized) is used to develop the hidden units of these networks. According to the idea of quantum state superposition, transfer function of neurons in the hidden layer is expressed as linear superposition multi-level transfer function leading to reduce the quantum interval. Hence, the classification has more freedom. Another advantage is that the input weights are cancelled, as well as the number of neurons in the hidden layer is reduced. Therefore the number of output weights becomes small.

The proposed QRWNN consists of  $n_i$  inputs, one layer of  $n_h$  multilevel hidden nodes, and  $n_o$  output nodes. Let the only synaptic weight connecting the  $i$ th output nodes to the  $j$ th hidden nodes be  $w_{ij}$ . Let  $x^k = [x_1^k, x_2^k, x_3^k, \dots, x_{n_i}^k]^T \forall k = 1, \dots, m$ , where  $m$  is the number of the feature vectors of data set  $X$ . Suppose a multilevel hidden unit has  $n_s$  discrete quantum levels. Its activation function can be written as a superposition of  $n_s$  activation functions, each shifted by  $\theta^s$ . Gaussian wavelet has been chosen to serve as an adaptation basis function to network's hidden layer.

$$X_\theta = \beta(x - \theta^s) \quad (1)$$

$$\Psi(x) = \frac{1}{n_s} \sum_{s=1}^{n_s} \Psi_o(\|X_\theta - T_1\|/\lambda) \quad (2)$$

Where  $\Psi_o(x)$  is activation function,  $\beta$  is a slope factor, and  $\theta^s$  is the jump positions in the activation function. The step

widths of the multi-level activation function, called the quantum intervals, are determined by jump-positions  $\theta^s$ . The input to the  $j$ th hidden unit from  $k$ th feature vector is  $\hat{h}_j^k = \sum_{l=1}^{n_i} w_{lj} x_l^k$ . Therefore response of the  $j$ th multi-level hidden node to the  $k$ th feature vector  $x^k$  can be written as

$$x_i^k = [x_1^k, x_2^k, \dots, x_{n_i}^k]^T$$

$$x_{j,\theta}^k = \beta(x_j^k - \theta^s) \quad (3)$$

$$r_j = \frac{\|x_{j,\theta}^k - T_1\|}{\lambda_1}$$

Therefore response of the  $j$ th multi-level hidden node to the  $k$ th feature vector  $x^k$  can be written as

$$H_j^k = \frac{1}{n_s} \sum_{s=1}^{n_s} \Psi(\|x_{j,\theta}^k - T_1\|/\lambda) \quad \text{where } s = 1, 2, \dots, n_s, \\ j = 1, 2, \dots, n_h \quad (4)$$

The output vector of output layer is

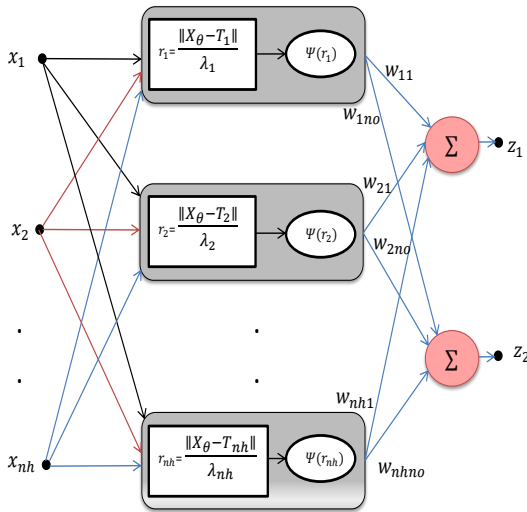
$$z_i^k = \sum_{j=1}^{n_h} w_{ji} H_j^k \quad i = 1, 2, \dots, n_o \quad (5)$$

The general model of the QRWNN is

$$z_i^k = \sum_{j=1}^{n_h} w_{ji} \left( \frac{1}{n_s} \sum_{s=1}^{n_s} \Psi \left( \frac{\sqrt{\sum_{l=1}^{n_i} (B(x_l - \theta_j^s) - T_j)^2}}{\lambda_j} \right) \right) \quad (6)$$

Where  $k$  is sample number.

The linear partition generated by an additional hidden unit has all the degree of freedom to align itself along any direction on the feature space. On other hand, the activation function within multi-level hidden nodes transfer function can only “spread-out” or “collapse-in” parallel to each other [3]. Figure 1 shows the structure of the QRWNN.



**Fig 1: Quantum radial wavelet neural network structure.**

### 3. A GRADIENT-DESCENT-BASED LEARNING ALGORITHM FOR THE QRWNN

The assumed underlying algorithm is easily calculated, best for large data sets. Convergence between these learning algorithms based on rate  $\alpha$ . If they are too large (such as

constant) oscillation may occur. If they are too small, may not move far enough to reach a local minimum [4]. This can be done in two ways: on-line learning or example-by-example, in which the weights out  $w_{ji}$  are adjusted after every training pattern; and batch or off-line learning, in which learning (weight adjustment) occurs after all of the training examples have been presented to the network once [5].

The learning of QRWNN parameters is considered in two steps. The synaptic weights out need to be update first in order to train the QRWNN to consistently partition the feature space of the given data set. Simultaneously, the uncertainty present in the feature space must be learned through the adaptation of the parameters  $\theta_j^s$ .

#### 3.1 Updating the Weights out of QRWNN

Let  $d_i^k = [d_1^k d_2^k d_3^k \dots d_{n_o}^k]^T$  be the desired output vector for the  $k$ th input feature vector  $x_k$ . Let  $z_i^k = [z_1^k z_2^k z_3^k \dots z_{n_o}^k]^T$  be the actual output. A gradient-descent-based algorithm for learning the synaptic weights of the QRWNN can be derived by minimizing the quadratic error function sequentially for each  $k$ .

$$E^k = \frac{1}{2} \sum_{i=1}^{n_o} (d_i^k - z_i^k)^2 \quad \text{with } k = 1, 2, \dots, m.$$

Where  $z_i^k$  is the actual output of the training pattern  $k$  and output  $i$ .  $d_i^k$  is desired output of the training pattern  $k$  and output  $i$ .  $m$  is the total number of the training pattern.  $E$  are mean square error functions.

The weights out  $w_{ji}$  are adjusted after each feature vector  $x^k$  so that  $E^k$  is minimized, this method is called example-by-example or on-line learning. The advantages of this method are [4]:

- It requires less computation per step.
- Randomization may help escape poor local minima.

The parameters  $w_{ji}$  are updated by minimizing objective function  $E^k$ .

$$\frac{\partial E^k}{\partial w_{ji}} = e_i^k \hat{H}_j^k \quad (7)$$

$$r_j^i = \frac{\sqrt{\sum_{l=1}^{n_i} (B(x_l - \theta_j^s) - T_j)^2}}{\lambda_j}$$

Where  $e_i^k = z_i^k - d_i^k$ .

At any epoch  $r$ , adjustment of parameters  $w_{ji}$  are performed.

#### 3.2 Updating the Quantum Intervals

The quantum intervals of quantum neurons in the hidden layer can be learned by minimizing the class-conditional variances at the output of hidden nodes. On other hand quantum intervals can be learned by a minimal output variety of hidden layer neuron based on the same data category of sample data. The output of variance for class  $c_i$  is [5]

$$\sigma_{j,i}^2 = \sum_{x^k \in c_i} (\langle \hat{H}_j^{c_i} \rangle - \hat{H}_j^k)^2 \quad (8)$$

Where  $\langle \hat{H}_j^{c_i} \rangle = \frac{1}{C_i} \sum_{x^k \in c_i} \hat{H}_j^k$

and it represents the sum of outputs of hidden neurons for all the inputs that belong to class  $c_i$  divided by the number of samples in that class.

The adjustment of quantum intervals  $\theta_j^s$  will be realized by minimizing objective function G.

$$G = \frac{1}{2} \sum_{j=1}^{n_h} \sum_{i=1}^{n_o} \sigma_{j,i}^2 = \frac{1}{2} \sum_{j=1}^{n_h} \sum_{i=1}^{n_o} \sum_{x^k \in c_i} (\langle \hat{H}_j^{c_i} \rangle - \hat{H}_j^k)^2 \quad (9)$$

The update equation for  $\theta_j^s$  can be obtained by setting the change in  $\theta_j^s$ , say  $\Delta\theta_j^s$  proportional to the gradient of G with respect to  $\theta_j^s$  as

$$\frac{\partial G}{\partial \theta_j^s} = \sum_{i=1}^{n_o} \sum_{x^k \in c_i} (\langle \hat{H}_j^{c_i} \rangle - \hat{H}_j^k) \left[ \frac{\partial \langle \hat{H}_j^{c_i} \rangle}{\partial \theta_j^s} - \frac{\partial \hat{H}_j^k}{\partial \theta_j^s} \right] \quad (10)$$

$$\frac{\partial \langle \hat{H}_j^{c_i} \rangle}{\partial \theta_j^s} = \frac{1}{c_j} \sum_{x^k \in c_i} \frac{\partial H_j^k}{\partial \theta_j^s} \quad (11)$$

Where  $\frac{\partial H_j^k}{\partial \theta_j^s} = \frac{1}{n_s} \frac{\partial \Psi}{\partial \theta_j^s} = -\frac{1}{n_s} \frac{\partial \Psi}{\partial r_j^s} \beta$ .

According to Equations (10) and (11), the adjustment formula of  $\theta_j^s$  is:

$$\Delta\theta_j^s = -\alpha_\theta \frac{\partial G}{\partial \theta_j^s} \quad (12)$$

Where  $\alpha_\theta \in (0,1)$  is the learning ratio of  $\theta_j^s$

---

*Algorithm: Training The QRWNN*

The 1<sup>st</sup> step in the QRWNN classifier algorithm is to initialize all the parameters according to the above equations.

Update the synaptic output weights:

For  $k=1,2,\dots,m$

For  $j=1,2,\dots,n_h$

$$x_j^k = [x_1^k, x_2^k, x_3^k, \dots, x_{n_i}^k]^T$$

For  $s=1,2,\dots,n_s$

$$x_{j,\theta}^k = \beta(x_j^k - \theta_j^s)$$

$$r_j = \frac{\|x_{j,\theta}^k - T_1\|}{\lambda_1}$$

$$H_j^k = \frac{1}{n_s} \sum_{s=1}^{n_s} \Psi(r_j^s)$$

For  $i=1,2,\dots,n_o$

$$z_i^k = \sum_{j=1}^{n_h} w_{ji} H_j^k$$

For  $i=1,2,\dots,n_o$

$$dw_i^k = d_i^k - z_i^k$$

Update the quantum intervals

For  $k=1,2,\dots,m$

For  $k=1,2,\dots,n_h$

$$x_j^k = [x_1^k, x_2^k, x_3^k, \dots, x_{n_i}^k]^T$$

For  $s=1,2,\dots,n_s$

$$x_{j,\theta}^k = \beta(x_j^k - \theta_j^s)$$

$$r_j = \frac{\|x_{j,\theta}^k - T_1\|}{\lambda_1}$$

$$H_j^k = \frac{1}{n_s} \sum_{s=1}^{n_s} \Psi(r_j^s)$$

$$DH_j^{k,s} = \frac{\partial \Psi(r_j^s)}{\partial r_j^s}$$

For  $j=1,2,\dots,n_h$

$$\langle H_j^{c_i} \rangle = \frac{1}{c_i} \sum_{x^k \in c_i} H_j^k$$

$$\langle DH_j^{c_i,s} \rangle = \frac{1}{c_i} \sum_{x^k \in c_i} DH_j^{k,s}$$

For  $k=1,2,\dots,m$

For  $j=1,2,\dots,n_h$

For  $s=1,2,\dots,n_s$

$$\theta_j^s = \theta_j^s + \alpha_\theta \frac{\beta}{n_s} \sum_{i=1}^{n_o} \sum_{x^k \in c_i} (\langle H_j^{c_i} \rangle - H_j^k) (\langle DH_j^{c_i,s} \rangle - DH_j^{k,s})$$

#### 4. DATA SELECTION

From the data available at [6], the whole database consists of five EEG data set (denoted A-E), each containing 100 single channel EEG signals of 23.6 seconds from five separate classes. Sets A and B consisted of signals taken from surface EEG recordings of five healthy volunteers with eye open and eye closed, respectively. Signals in set C and set D were recorded in seizure-free intervals from five epileptic patients from the hippocampal function formation of the opposite hemisphere of the brain and from within the epileptic zone, respectively. Set E contains the records of five epileptic patients during seizure activity. All EEG recordings were made with the same 128-channel amplifiers system, using an average common reference. The recorded data was digitised at 173.61 samples per second using 12-bit resolution. Band-pass filter setting were 0.53-40 Hz (12dB/oct). The amplitude of EEG recordings are given in micro volt [7].

#### 5. CLASSIFICATION OF EEG SIGNAL

The classification of EEG can be divided into three parts. Part one is to process raw EEG signals by removing all noise and artifacts, but preserving all the characteristics of the original signal. As well as cleaning the signal from the influence of the reference electrode, if one is used [8]. Independent Component Analysis (ICA) is the best method to accomplish all of these, since the EEG data are non-Gaussian, non-linear, and non-stationary [9].

Part two is feature extraction by cluster technique (CT). Feature extraction is the process of extracting useful information from the signal. Features are characteristics of a signal that are able to distinguish between normal and abnormal cases of EEG signals. The requirements for feature extraction are [8]:

- Reducing the size of the data by selecting appropriate features.
- The selected features should be minimally redundant and the expected results should maximally depend on these features.

- Preserving all information from the signal that is needed for classification.

The CT method is proposed for feature extraction from the original EEG database. It is conducted in three steps as follows:

Step 1: Each EEG data is divided into  $n$  groups, which are called clusters with specific time interval.

Step 2: Each cluster is partitioned into  $m$  sub-clusters with a specific duration.

Step 3: Eight statistical features are extracted from each sub-cluster data point. The statistical features are minimum, maximum, mean, median, first quartile range, third quartile range, inter-quartile range and standard deviation of the EEG data.

Part three is EEG signal classification by QRWNN model. Where the obtained features are employed as the input to the proposed QRWNN model.

## 6. EXPERIMENTS ON EEG SIGNALS

The epileptic EEG data has five sets, Set A to Set E, and each set contains 100 channel of data. Every channel consists of 4069 data points with 23.6 seconds. Each channel data is normalized and then divided into 16 groups where each group is called cluster and each cluster consists of 256 data points of 1.475 seconds. Then every cluster is again partitioned into sub-clusters. Each sub-cluster contains 64 observations of 0.3688 second. Eight statistical features are calculated from each sub-cluster. These features are the input for each of five different classification techniques, namely: Feed Forward Neural Network (FFNN), Quantum Neural Network (QNN), Wavelet Neural Network (WNN), Quantum Wavelet Neural Network (QWNN), and finally the Quantum Radial Wavelet Neural Network (QRWNN).

The number of elements of each pair has been divided into two parts:

- First part is used for training network called training samples. These samples usually take 70% from data. The data sample is sent through the network to find and then calculate error and accuracy of training data.
- Second part is used for testing network called testing samples. These samples take the remaining 30% of data. The testing samples are tested by the above methods and the error and accuracy of testing data are calculated.

The performance of a particular run of the program, or a particular reading by an expert was evaluated in terms of the accuracy of the classification. This accuracy represents the per cent of the ratio of the total number of correct classifications to the total number of testing samples.

## 7. RESULTS

In this work the patterns for training and testing processes are chosen randomly. Training is conducted until the standard of error fell below 0.0001. The mean square error denotes the error limit to stop QRWNN training.

A three layer structure of QRWNN is applied. It consists of 32 inputs, 2 outputs, 10 hidden layer nodes and 2 quantum intervals. The slope factor of multi-level transfer function  $\beta_h$  is chosen to be 0.4 by trial and error. The learning rate for weight adjusting is set to 0.5.

For QRWNN the average accuracy of 90.561% is obtained after 500 iterations. The average Mean Square Error (MSE) is 0.16 and time of the training process is 50 minutes. Table 1 shows the results of the five classification techniques.

In order to evaluate the performance of the proposed QRWNN model with the other models (FFNN, QNN, WNN, and QWNN), a comparison factor is proposed. It represents a trade-off between the time taken for classification and the accuracy of it. This factor represents the classification accuracy to classification time ratio and is denoted as (ATR). As the value of ATR becomes high that refers to a better performance. The ATR value of QRWNN classifier is 1.811, while its values for FFNN, QNN, WNN and QWNN are 0.104, 0.105, 0.122 and 0.108, respectively. That shows the superiority of the QRWNN model.

**TABLE 1. Comparison of EEG classification results**

Classifier	No. of $n_h$	No. of $n_s$	MSE for training	MSE for testing	Accuracy %	Time (min)	ATR
FFNN	128	-	0.17	0.3	75.076	750	0.104
QNN	64	7	0.16	0.3	83.713	790	0.105
WNN	64	-	0.04	0.4	88.003	720	0.122
QWNN	64	10	0.07	0.3	94.187	870	0.108
QRWNN	10	2	0.16	0.2	90.561	50	1.811

## 8. DISCUSSION

An epileptic EEG data is used in this work to test the performance of the proposed QRWNN model. All calculations are performed using MATLAB (version 7.12.0.635). 2240 vectors are used for training and 960 vectors for testing.

The advantage of QRWNN is that it overcomes the problems of slow speed and low accuracy of convergence and the shortcomings of generalization ability for pattern recognition of the traditional neural network. The combination of QNN with radial wavelet making the network more clear and greatly reduces the size of the network and improve the learning speed to become very fast in comparison with other systems. Also, if the eigenvectors of samples are located in the overlap edges of two modes, the QRWNN will distribute them to all the relevant categories in proportion. Gaussian wavelet is often used for time-frequency analysis of EEG signals, because it can capture EEG events well. The other key advantage of Gaussian wavelet is that it will manage to detect the great changes in EEG signals. In addition, it is a high frequency signal and is suitable to be a mother wavelet. Hence, Gaussian mother wavelet that is used as activation

function sense provides considerable flexibility in designing the QRWNN model. The other advantage of QRWNN over QWNN is the reduction of the number of neurons in the hidden layer which could lead to a little computation, 10 hidden nodes for QRWNN but 128 hidden nodes for FFNN and 64 hidden nodes for QNN, WNN and QWNN. This means reducing the total number of parameters (input weights, output weights, and jump position) to be learned.

## 9. CONCLUSION

In this paper a QRWNN model with Gaussian wavelet as an incentive function is proposed for classification of EEG signals. Because EEG signals possess a combination of slow variations over long periods, with sharp transient variations over short periods, wavelet neural networks are more natural choice than other mainstream neural networks for EEG analysis. Thus, the main contribution of this work is the development of a new QRWNN model for the analysis and classification of EEG signals. It has been found that using quantum neurons instead of traditional neurons and radial features are often able to learn faster. The other advantage of QRWNN is the reduction of the hidden layer nodes leading to a little computation. The results of the experiments demonstrated that the proposed QRWNN model has the best performance by having the highest value of ATR compared to FFNN, QNN, WNN and QWNN EEG classifiers. The ATR factor represents a compromise between classification accuracy and time of classification. Thus, its value provides a performance indication of different techniques. The ATR value of QRWNN is greater than its values of FFNN, QNN, WNN and QWNN by 17.41, 17.24, 14.84 and 16.76 times, respectively. In addition, we can choose the suitable basis function for the data in QRWNN making it more flexible than the rest.

## 10. REFERENCES

- [1] Taha, Z. K. 2012 Quantum Neural Network Model with Wavelet Theory. M. Sc. Thesis. Electrical Engineering Dept., University of Baghdad.
- [2] Ropper, A. H. and Brown, R. H. 2011 Adams and Victor's Principles of Neurology, 8<sup>th</sup> edition.
- [3] Karayiannis, N. B. and Purushothaman, G. 1997. Quantum neural networks (QNN's): inherently fuzzy feed forward neural networks, IEEE Transactions on Neural Networks, vol. 8, no. 3.
- [4] More on Regression Gradient Descent Classification. COMP-652 Lecture 2, September 6, 2005. Available: <http://www.facweb.litkgp.ernet.in/~sudeshna/courses/>
- [5] Mahdi, A. A. H. 2010 The Application of Neural Network in Financial Time Series Analysis and Prediction Using Immune System. M. Sc. Thesis. School of Computing and Mathematical Sciences, Liverpool John Moores University.
- [6] EEG Time Series (epileptic data). 2005. Available: [http://www.meb.Unibonn.de/epileptologie/science/physi\\_kleegdata.html](http://www.meb.Unibonn.de/epileptologie/science/physi_kleegdata.html)
- [7] Siuly, Y. L. and Wen, P. 2010. Analysis and classification of EEG signals using a hybrid cluster technique. In Proceedings of the IEE/ICME International Conference on Complex Medical Engineering, Gold Coast, Australia, July 13-15, pp. 34 – 39.
- [8] Horlings, R. 2008. Emotion Recognition Using Brain Activity. Man-machine Interaction Group, Delft University of Technology, Faculty of Electrical Engineering, Mathematics, and Computer Science.
- [9] Principal Component Analysis (PCA) and Independent Component Analysis (ICA). Available: <http://www.cis.hut.fi/projects/ica/fastica/>