International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:3, No:11, 2009

Improved Wavelet Neural Networks for Early Cancer Diagnosis Using Clustering Algorithms

Zarita Zainuddin and Ong Pauline

***Abstract*—**Wavelet neural networks (WNNs) have emerged as a vital alternative to the vastly studied multilayer perceptrons (MLPs) since its first implementation. In this paper, we applied various clustering algorithms, namely, K-means (KM), Fuzzy C-means (FCM), symmetry-based K-means (SBKM), symmetry-based Fuzzy C-means (SBFCM) and modified point symmetry-based K-means (MPKM) clustering algorithms in choosing the translation parameter of a WNN. These modified WNNs are further applied to the heterogeneous cancer classification using benchmark microarray data and were compared against the conventional WNN with random initialization method. Experimental results showed that a WNN classifier with the MPKM algorithm is more precise than the conventional WNN as well as the WNNs with other clustering algorithms.

***Keywords*—**Clustering, microarray, symmetry, wavelet neural networks.

I. Introduction

ARTIFICIAL neural networks (ANNs), consist of a large number of interconnecting artificial neurons, which employ mathematical or computational models for information processing. Due to its fascinating characteristics of robustness, fault tolerance, adaptive learning and massively parallel capabilities, ANNs have been applied widely in vast applications, such as time series prediction, pattern recognition and system identification [1]-[3].

Multilayer perceptrons (MLPs), along with the back-propagation learning algorithm, are the most popular type of ANNs among all in practical situations [4]-[5]. Nevertheless, shortcomings of an MLP: difficulties in reaching the global minimum in a complex search space, time-consuming and failure to converge when high nonlinearities exist [6], have deteriorated the accuracy of its application.

To overcome the deficiencies of an MLP, a wavelet neural network (WNN) has been introduced as a vital alternative to the MLP [7]-[9]. By integrating wavelet families as the activation function in the hidden layer of WNNs, there are various issues that are concerned with WNNs, varying from different learning algorithms, network architecture, type of activation functions used in hidden layer and also the parameter initialization. In this paper, the issue of choosing the translation parameter will be addressed.

A proper initialization of the network parameter is crucial to achieve faster convergence rate and higher accuracy rate. Methods of using an explicit expression, hierarchical clustering, support vector machine, genetic algorithm and K-means clustering are among the approaches that have been implemented in the parameter initialization [10]-[14]. In this paper, we applied various clustering algorithms, namely, K-means (KM), Fuzzy C-means (FCM), symmetry-based K-means (SBKM), symmetry-based Fuzzy C-means (SBFCM) and modified point symmetry-based K-means (MPKM) clustering algorithms in initializing the WNN translation parameter. By integrating various clustering algorithms into the WNN, these proposed systems will be applied in a real world application, where the classification problem of heterogeneous cancer using the microarray data is our main concern.

This paper is organized as follows: In section II, a brief introduction of WNNs will be given, followed by the description of various clustering algorithms in section III. Next, a short review of a microarray experiment, materials and methodology are given in section IV. In section V, the experimental results, as well as the performance comparison between these modified WNNs and the conventional WNN with random parameter initialization are given. Finally, some conclusions and future directions are drawn in section VI.

II. Wavelet Neural Networks

A. Wavelet Neural Network Architecture

Since the first implementation of WNNs by Zhang and Benveniste [15], WNNs have received tremendous attention from other researchers [16]-[19], due to its great improvement over the weaknesses of MLPs.

A schematic diagram of a WNN, with d input nodes, m hidden nodes and L output nodes is shown in Fig. 1. The input layer will receive the input variable $\vec{x} = (x_1, ..., x_d)$ and transmit the accepted input variables to the next layer. The second layer is a hidden layer with a mother wavelet ψ in each hidden node. The nodes in this layer are given by the product of the mother wavelet as

This work was supported in part by the Ministry of Science Technology and Innovation (MOSTI) e-Science Grant under Grant 01-01-05-SF0243.

Zarita Zainuddin is with the School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia (phone: (6)04-6533940; fax: (6)04-6570910; e-mail: zarita@cs.usm.my).

Ong Pauline is with is with the School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia (e-mail: ong.pauline@hotmail.com).

$$\psi_j(\zeta) = \psi(\| D_j(X - t_j) \|), j = 1,...,m \quad (1)$$

where D_j and t_j are the scaling and translation vectors respectively.

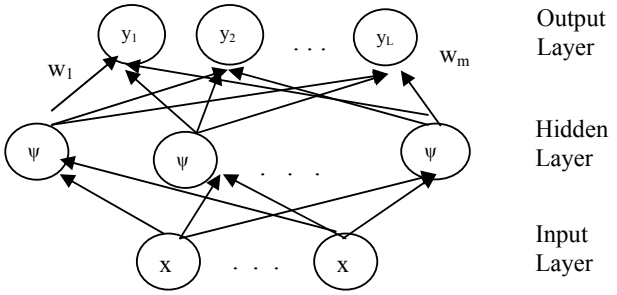

Fig. 1 A schematic diagram of wavelet neural network with d input nodes, m hidden nodes and L output nodes

The sigmoid function (logistic and hyperbolic tangent) is the commonly used basis function in an MLP. Compared to the basis functions in the hidden layers of the MLP, the mother wavelet used in the hidden nodes of a WNN is a localized activation function. Hence, the connection weight associated with the hidden nodes can be viewed as locally piecewise constant models, which leads to learning efficiency and structure transparency.

The third layer is the output layer. The output will be the linear combination of the weighted sum of the hidden layer, which is given by (2).

$$y_k(x) = \sum_{j=1}^{m} w_j \psi_j(\| D_j(X - t_j \|) + \Theta, k = 1,2,...,L \quad (2)$$

where w_j and Θ are the weight vector and bias term between hidden layer and output layer respectively. Obviously, all the neurons in any layer are fully connected to the preceding and also the succeeding layer, but no connections between the neurons within the same layer are allowed.

B. Learning Algorithm of Wavelet Neural Networks

Various learning algorithms can be applied to the training of a WNN; in this paper, the learning of the WNN is by the method of solving the pseudo-inverse with fixed parameter initialization. Therefore, only the weight matrix W needs to be adjusted during the training of WNN, in order to map the underlying relationship between the input and output space.

Before we begin to describe the learning algorithm of the WNN, let us define the cost function as in (3):

$$E(f(n)) = \frac{1}{2}(y_d(n) - y(n))^2 \quad (3)$$

where y_d is the desired output value and $y(n)$ is the output value from the WNN. Hence, the training of the WNN is based on the minimization of the cost function.

There are two stages involved in the learning of the WNN. Firstly, the scaling parameter is fixed. Next, the translation vector is chosen from the input vectors using either, KM, FCM, SBKM, SBFCM or MPKM. The details of these clustering algorithms will be discussed in next section.

Let us represent (2) as $Y = \psi W$, where

$$\psi = \begin{pmatrix} \psi(x_1, D_1, t_1) & \psi(x_1, D_2, t_2) & ... & \psi(x_1, D_m, t_m) \\ \psi(x_2, D_1, t_1) & \psi(x_2, D_2, t_2) & ... & \psi(x_2, D_m, t_m) \\ \vdots & \vdots & \vdots & \vdots \\ \psi(x_d, D_1, t_1) & \psi(x_d, D_2, t_2) & ... & \psi(x_d, D_m, t_m) \end{pmatrix} \quad (4)$$

is the output of wavelet families ψ, and $\psi(x_i, D_i, t_i) = \psi(\| D_i(x_i - t_i) \|)$.

Therefore, in order to solve the weight matrix W, $W = \psi^+ Y$ is computed. ψ^+ is the pseudo-inverse defined as $\psi^+ = (\psi^T \psi)^{-1} \psi^T$.

A summary of the learning algorithm of WNN is given as below:

i. Initialize the values for D_i and t_i.
ii. Feed in the input vector X into the WNN.
iii. Calculate the product of the hidden layer by using eqn. (1).
iv. Solve for the weight matrix W by using the pseudo-inverse method.
v. Obtain the output value of WNN, $y(n)$ from step (iv).
vi. Compare $y(n)$ with the desired output value, y_d.
vii. Calculate the cost function as in eqn. (3).
viii. Repeat steps (ii) to (vii) until it meets the stopping criterion.

III. The Clustering Algorithm

A. The Conventional K-means and Fuzzy C-means

The action of grouping together the patterns into dissimilar clusters, with respect to a similarity measure is referred as clustering, in such a way that the patterns within the same cluster have higher similarity measure than the patterns in the other clusters.

The two main approaches of cluster analysis are the crisp/hard clustering and fuzzy clustering, where the KM and FCM are the most widely used algorithms for each approach respectively. For the former approach, each pattern can only be categorized into one cluster, whereas for the latter, each pattern can belong to more than one cluster, with a degree of similarity which is specified by a membership function. The details of KM and FCM algorithms can be obtained in [20].

However, both the KM and FCM use Euclidean distance as its similarity measure which tends to detect the hyper-spherical shaped cluster of equal size, which is unfavorable when the real world data consisted of various sizes and

shapes. Thus, these conventional clustering algorithms need enhancements.

B. The Symmetry-Based K-means and Fuzzy C-means

If we notice our surrounding carefully, it can be found that the concept of symmetry exists in almost every single area in our daily life. A circle, a car, a ladder or even the crystal chemical substances of the snow are exhibiting the characteristic of symmetry in its nature. Thus, a novel similarity measure, namely the point symmetry distance (PSD) which is adopted in the conventional clustering algorithms is proposed by Su and Zhou [21], in such a way that the patterns are assigned to a cluster if they have a symmetrical sense with respect to the cluster center.

Given a set of patterns $p_i, i = 1,..., N,$ and a cluster center $c,$ to measure the degree of symmetry of a data point with respect to a cluster center, PSD is defined as:

$$d_s(p_j, c) = \min_{i=1,...,N, i \neq j} \frac{\| (p_j - c) + (p_i - c) \|}{\| (p_j - c) \| + \| (p_i - c) \|} \quad (5)$$

Pursuant to this, Su, Zhou and Hsieh [22] had combined the PSD with K-means and Fuzzy C-means clustering algorithms, where, the resulting algorithms have been proven to work efficiently in handling the data with different geometric properties. We name the resulting clustering algorithm by Su and Zhou as Symmetry Based K-Means (SBKM) and Symmetry Based Fuzzy C Means (SBFCM) respectively.

The details of SBKM and SBFCM can be obtained from [21] and [22].

C. The Modified Point Symmetry-based K-means

Upon analyzing the SBKM and SBFCM, Chung and Lin [23] pointed out that there are potential problems that existed in the PSD, including the disadvantages of lack of distance difference symmetry property, unsatisfactory in handling the symmetrical inter-clusters, and lack of closure property. Pursuing in solving these shortcomings, Chung and Lin have proposed a new operator, namely the symmetry similarity level (SSL), which is now defined as:

$$SSL(p_i, c_k) = \max_{p_j \in c_k} \sqrt{\frac{DSL^2(p_i, c_k, p_j) + OSL^2(p_i, c_k, p_j)}{2}} \quad (6)$$

for $1 \le k \le K$ and $1 \le i \le N$ where K is number of clusters and N is number of data points.

As observed from (6), the SSL operator consists of two components, namely the distance similarity level (DSL) and orientation similarity level (OSL), where DSL and OSL are defined as follows:

$$DSL(p_i, c_k, p_j) = \begin{cases} 1 - \dfrac{| d_i - d_j |}{d_i}, & if \quad 0 \le d_j / d_i \le 2 \\ 0 & otherwise \end{cases} \quad (7)$$

where $d_i = \overline{p_i c_k}$ and $d_j = \overline{p_j c_k}$,

$$OSL(p_i, c_k, p_j) = \frac{v_i . v_j}{2 \| v_i \| \| v_j \|} + 0.5 \quad (8)$$

where $v_i = (c_k - p_i)$ and $v_j = (p_j - c_k)$.

From (7), Chuang and Lin have integrated the SSL operator into the K-means algorithm (MPKM), where the resulting MPKM performed efficiently in detecting not only the symmetrical intra-clusters, but also the symmetrical inter-clusters.

The flow of the MPKM algorithm is as follows:

Step 1: **Initialization**

Initialize the cluster centers $c_i, i = 1,..., K$, by selecting K points randomly from all the data points

Step 2: **Coarse-Tuning**

Use the conventional K-Means clustering algorithm to update the *K* cluster centers.

Step 3: **Fine-Tuning 1**

Find out the set SB_{ik} of all candidate symmetrical data points p_j for each data point p_j, such that $DSL(p_i, c_k, p_j) \ge \alpha$ and $OSL(p_i, c_k, p_j) \ge \beta$, where $p_j \in C_k$.

Step 4: **Fine-Tuning 2**

Compute (6) for each data point p_i.

If the value for $SSL(p_i, c_k*, p_j)$ is the largest and the most symmetrical point p_j relative to c_k* belongs to SB_{ik}, assign data point p_i to cluster center c_k*.

Otherwise, assign data point p_i to the cluster center c_k** with the shortest distance.

Step 5: **Updating**

Compute the new center for each cluster by

$$c_k^{new} = \frac{1}{| C_k |} = \sum_{p_i i \in Ck} p_i$$

Where C_k is the set of the data points that have been assigned to the cluster center c_k, and $| C_k |$ is the number of data points in C_k.

Step 6: **Continuation**

If there are no patterns changes the clusters or the iterations reach a predefined maximum value, then stop. Otherwise, go to Step 3.

More details in the efficiency and effectiveness of the MPKM in detecting the symmetrical inter and intra-clusters can be obtained from [23].

IV. Microarray

A. Microarray Experiment

Classification of patient samples is a critical issue in cancer diagnosis and treatment, since a correct classification enables specific therapies to pathogenetically distinct tumor types, which enables the maximizing of efficacy and minimizing of toxicity in the therapy. However, heterogeneous cancers are hard to be distinguished into subtypes by clinical and histopathological means, since the traditional diagnostic method mainly depends on the morphological appearance of the tumor. Furthermore, tumors with similar morphological appearance might response diversely to the same treatment. Hence, heterogeneous cancer distinction only can be made in hindsight, based on the response of patients towards the treatments received, where differentiating heterogeneous cancers highly rely on the experience of the physician.

Fortunately, emergence of microarray experiment as a cutting edge technology in bioinformatics enables the monitoring of expression levels of thousands of genes simultaneously. Hence, gaining insight of the cellular mechanism and the pathway of the biological reaction are no longer an obstacle.

Microarray slides are slides with spatially ordered probes of cDNA sequence which are printed on a rectangular grid form on the slides [24]. To start a microarray experiment, firstly, mRNA are extracted from the control samples and experimental samples, which are then will be reverse transcribed into cDNA. Next, the cDNA for both control and experimental samples are labeled with fluorescent dyes, cy3 (green colour) and cy5 (red colour) respectively. After that, both labeled samples are mixed and poured onto the microarray slides, where the hybridization will take place. The cDNA only will bind to the specific probes on the array, by following the base-pair complementarities. Then, the slides are washed, to get rid those labeled samples which are not hybridized with the probes on the array.

After the hybridization, the microarray slides are scanned by a laser, where the fluorescent dye of the labeled samples which bound with the probes will be excited by the laser, where the emitted detectable light will be captured by a scanner. The so-called gene expression levels can be measured from the amount of the bound labeled samples, where it can be quantified by measuring the fluorescence intensities, since the probes with more bound labeled samples will fluoresce more intensely. By using image processing software, the fluorescence background of this scanned image is then subtracted, and the expression values for each probe on the rectangular grid are calculated. In the end, when all the data from all samples are collected, a gene expression matrix, which is the input for the classification system is obtained, where its rows correspond to the single gene and its column correspond to the single sample.

TABLE I
Information for the benchmark dataset used in the experimental simulations

Dataset	Number of Genes	Number of Samples	Number of Classes	Author
LEU	7129	72	2	Golob et al.
SRBCT	2308	63	4	Khan et al.
GLO	12625	50	2	Nutt et al.
CNS	7129	40	5	Pomeroy et al.

B. Microarray Benchmark Dataset

1) Leukemia (LEU): Golub et al. [25] were the pioneers in distinguishing between subtypes of heterogeneous cancers based on the genes expression signatures, where the proposed approach is applied to the human acute leukemia dataset with two subtypes, namely ALL and AML. The dataset comes from a gene expression study of 47 ALL and 25 AML tissue samples and there are a total of 7129 genes used in this microarray experiment.

2) Small Round Blue-Cell Tumor (SRBCT): The SRBCT dataset used by Khan et al. [26] composed of 2308 gene expressions in four classes: Ewing's sarcoma (EWS), Burkitt's lymphoma (BL), neuroblastoma (NB) and rhabdomyosarcoma (RMS). This dataset comes from 63 tissue samples, with 23 samples of EWS, 8 samples of BL, 12 samples of NB and 20 samples of RMS.

3) Glioma (GLO): Glioma is a type of cancer that starts in the brain or spine. The Glioma dataset used by Nutt et al. [27] consists of 12,625 genes of 50 samples, with 28 glioblastomas and 22 anaplastic oligodendrogliomas. The glioblastomas and anaplastic oligodendrogliomas are classified further into classic (CG) and non-classic gliomas (NG).

4) Central Nervous System Embryonal Tumor (CNS): Embryonal tumors of the CNS are a group of heterogeneous tumors, which consists of 5 subclasses: medulloblastoma (MED), malignant glioma (MG), atypical teratoid/rhabdoid tumors (AT/RT), normal cerebellum (NC) and primitive neuroectodermal (PNET) are studied by Pomeroy et al. [28] A total of 40 tissue samples with 7129 genes of CNS are used, with 10 samples of MED, 10 samples of MG, 10 samples of AT/RT, 4 samples of NC and 6 samples of PNET.

A summary of the benchmark dataset is given in Table I.

C. Microarray Data Preprocessing

Microarray data consists of an overwhelming number of genes relative to the number of samples. However, the majority of such genes are probably irrelevant in discriminating between the subclass of the heterogeneous cancers. Hence, gene selection is a crucial aspect in microarray data analysis.

Before the process of gene selection, preprocessing of microarray data by using logarithmic transformation and normalization are required

Let x_{gi} be the element in the gene expression matrix, which denotes the spot intensity measurement of gene g-th $(g = 1,...,G)$ with respect to sample i-th $(i = 1,...I)$. Applying

the logarithmic transformation, the transformed spot intensity measurement x_{gi} is calculated by

$$x_{gi} = log(x_{gi}) \quad (9)$$

Subsequently, quantile normalization is applied on the transformed spot intensity measurement x_{gi}. Let the median mock array of x_{gi} be

$$M_g = median\{X_{g1}, X_{g2}, ..., X_{gI}\} \quad (10)$$

Followed from that, the percentiles $(Q_{i0}, Q_{i1}, ..., Q_{i100})$ of the *i*-th sample and the percentiles $(Q_{M0}, Q_{M1}, ..., Q_{M100})$ of the median mock array are calculated. Based on the percentiles $(Q_{i0}, Q_{i1}, ..., Q_{i100})$, for any value of x_{gi}, the interval $[Q_{ih}, Q_{i(h+1)}]$ for which it belongs to is obtained. Thus, the normalized value, X'_{gi}, can be found by finding the linear interpolation between the points (Q_{mh}, Q_{ih}) and $(Q_{M(h+1)}, Q_{i(h+1)})$.

D. Microarray Data Feature Selection

Next, we employ predictive gene selection by using conditional *t* (CT) method which has been proposed by Amaratunga and Cabrera [24], where, this novel method has been proved that it outperforms the ordinary *t* test, and also a popular regularization method, SAM. The details of the algorithm of CT can be found in [24]. There are still thousands of genes selected when the commonly used p-value cut-off of 0.05 is chosen for CT. Thus, a total number of 30 top ranked genes from each subtype of cancer are selected as the input variables into the WNN. It is reasonable to choose 30 top ranked genes in order to avoid the occurrence of under-fitting or over-fitting during the training of WNN, since 30 is not extremely big or small.

V. Experimental Simulations

A. Cross Validation

Excessive training will force the ANNs to memorize the input vectors and insufficient training will cause the ANNs unable to learn from the input vectors presented to it, where it will lead to poor generalization when new inputs are presented to the ANNs. Therefore, in order to avoid these problems, multifold cross validation is used. The samples are divided into k groups, where $k > 1$.

Firstly, one group from the samples is left out, where the training of the neural network involves the remaining of the samples. Next, the validation error is measured by testing it on the group left out. The process is repeated for k times, by each different group respectively as the testing set. The average of the validation error is calculated. In this study, we use a 10-fold cross validation.

B. Performance Index

The performance of the WNN based classifier is calculated by

$$classification\_rate = \frac{N_t}{N_{all}} \times 100\% \quad (11)$$

where N_t is the number of testing samples that been classified correctly and N_{all} is the total number of testing samples.

C. Performance Assessment and Discussion

Predictive success of the conventional WNN and the WNN with translation parameter initialization using KM, FCM, SBKM, SBFCM and MPKM in classifying the microarray benchmark dataset is shown in Table II.

As shown in Table II, the predictive competence of all WNN models is considerably superior in categorizing the heterogeneous cancer microarray dataset that we used. Ranging from 94.44% to 98.66% (LEU dataset), 100% (SRBCT dataset), 86% to 88% (GLO dataset) and 92.5% to 95% (CNS dataset), apparently all WNN models worked efficiently in separating the heterogeneous cancer into its subtypes, which is beneficial for cancer treatment, diagnosis and therapy.

By varying the initialization methods that we used in choosing the translation parameter of the WNN, observably,

TABLE II
Comparison of the Classification Capability of WNN with Different Initialzation Algorithm

Dataset	No. of Classes	Initialization Algorithm	Classification Rate
LEU	2	Random	94.44
		KM	95.83
		FCM	97.22
		SBKM	97.22
		SBFCM	97.22
		MPKM	98.61
SRBCT	4	Random	100
		KM	100
		FCM	100
		SBKM	100
		SBFCM	100
		MPKM	100
GLO	2	Random	86
		KM	88
		FCM	88
		SBKM	88
		SBFCM	88
		MPKM	88
CNS	5	Random	92.5
		KM	92.5
		FCM	92.5
		SBKM	95
		SBFCM	95
		MPKM	95

the conventional WNN with random parameter initialization yielded the lowest classification rate, specifically, 94.44%, 86% and 92.5% for the LEU, GLO and CNS datasets respectively.

Adopting the clustering algorithms in translation parameter initialization improves the classification rate slightly. Besides that, integrating KM and FCM with PSD had improved the predictive capability of WNN mildly, which can be observed in the classification results of the LEU and CNS datasets.

The efficacy and efficiency of the WNN is further enhanced by applying the SSL operator with the KM in the network initialization, which can be seen in the LEU dataset. The classification rate increased from 94.44% (conventional WNN) to 98.61% when the MPKM algorithm is employed.

In general, the highest classification for all microarray benchmark datasets is achieved when MPKM is applied in the parameter initialization. This verifies again the promising effectiveness of MPKM in choosing the proper initialization of the translation parameter for achieving faster convergence rate and higher accuracy rate.

By using the proposed methods, there are still some misclassification samples, which might be probably due to the process of gene selection. We know that tissue purity is important in the microarray experiment, yet biopsies of cancer tissues usually contain cancerous tissue and are unavoidable, a small portion of normal tissue. Hence, the genes expression values derived from the microarray experiment, in fact are from a composition of normal and cancerous tissues. An improved technique for the extraction of RNA from the heterogeneous cancers samples might help in enhancing the gene selection.

VI. Conclusion

A proper initialization method for WNN model is vital in improving the predictive competence and classification capability of WNN. In this paper, translation parameter initialization based on KM, FCM, SBKM, SBFCM and MPKM are developed and its application in microarray data cancer classification is studied.

The experimental results showed that these improved WNN models performed excellently in the cancer classification, where the highest correct classification percentage ranged from 88% to 100%, and they outperformed the WNN with random initialization. The improved WNN model with MPKM achieved the highest classification rate in all microarray benchmark datasets. Thus, the beneficial potential of this initialization method is promising.

Application of the MPKM with other neural network models and further real world application can be our future direction. Enhancing the MPKM with intelligent agents like fuzzy logic, particle swarm optimization will be an interesting topic to be pursued.

References

[1] E. Gómez-Ramírez, K. Najim and E. Ikonen, "Forecasting time series with a new architecture for polynomial artificial neural network," *Applied Soft Computin*, vol.7, pp.1209-1216, 2007.

[2] H. Zhang, B. Zhang, W. Huang and Q. Tian, "Gabor wavelet associative memory for face recognition," *IEEE Transactions on Neural Networks*, vol. 16, pp. 275-278, 2005.

[3] S. Srivastava, M. Singh, M. Hanmandlu and A.N. Jha, "New fuzzy wavelet neural networks for system identification and control," *Applied Soft Computing*, vol. 6 ,pp. 1-17, 2005.

[4] Y. Shirvany, M. Hayati and R. Moradian, " Multilayer perceptron neural networks with novel unsupervised training method for numerical solution of the partial differential equations," *Applied Soft Computing*, vol. 9, pp. 20-29, 2009.

[5] Z. Zainuddin and Evans, "Human face recognition using accelerated multilayer perceptrons," *International Journal of Computer Mathematics*, vol. 80, pp. 535-558, 2001.

[6] Z. Zainuddin and P. Ong, "Function approximation using artificial neural networks," *International Journal of Systems Applications, Engineering & Development*, vol. 1, pp. 173-178, 2007.

[7] A. Banakar and M. Fazle Azeem, "Artificial wavelet neural network and its application in neuro-fuzzy models," *Applied Soft Computing*, vol. 8, pp. 1463-1485, 2008.

[8] B. Biswal, P.K. Dash, B.K. Panigrahi and J.B.V. Reddy, "Power signal classification using dynamic wavelet network," *Applied Soft Computing*, vol. 9, pp. 118-125, 2009.

[9] C.J. Lin, "Nonlinear systems control using self-constructing wavelet networks," *Applied Soft Computing, vol*. 9, pp. 71-79, 2009.

[10] W. Maoan, J. Shijiu, W. Likun and Z. Van, "Defect characteristic prediction of pipeline by means of wavelet neural network based on the hierarchical clustering algorithm (Published Conference Proceedings style)," *Proceedings of Biennial International Pipeline Conference*, Alberta, 2004, pp. 921-924.

[11] Z. Xiao-Guang, K. Ying-Zhi, G. Dao-Hua and W. Xing Biao, "Fuzzy wavelet neural networks based on SVM," *Journal of East China University of Science and Technology*, vol. 32, no. 11, pp. 1351-1354, 2006.

[12] K. Seong-Ju, K. Yong-Taek, S. Jae-Yong and J. Hong-Tae, "Design of the scaling-wavelet neural network using genetic algorithm (Published Conference Proceedings style)," *Proceedings of the International Joint Conference on Neural Networks*, Honolulu , 2002, pp. 2174-2179.

[13] Y. Oussar and G. Dreyfus, "Initialization by selection for wavelet network training," *Neurocomputing*, vol. 34, pp. 131-143, 2000.

[14] Z. Dahai, B. Yanqiu, Y.B. Bi and Y. Sun, "Design and initialization algorithm based on modulus maxima of wavelet transform for wavelet neural network (Published Conference Proceedings style)," *International Conference on Power System Technology*, Singapore, 2004, pp. 897-901.

[15] Q. Zhang and A. Beveniste, "Wavelet networks," *IEEE Transactions on Neural Networks*, vol. 3, pp. 889-898, 1992.

[16] C. Jiacong and L. Xingchun, "Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique," *Engineering Applications of Artificial Intelligence*, vol. 21, pp. 1255-1263, 2008.

[17] E. Avci and D. Avci, "The performance comparison of discrete wavelet neural network and discrete wavelet adaptive network based fuzzy inference system for digital modulation recognition," *Expert Systems with Application*, vol. 35, pp. 90-101, 2008.

[18] L. Jie, L. Dai-fei, D. Xue-ru, Z. Zhong and D. Feng-qi, "Prediction of $AI(OH)_3$ fluidized roasting temperature based on wavelet neural network," *Transactions of Nonferrous Metals Society of China*, vol. 17, pp. 1052-1056, 2007.

[19] K. Vinay Kumar, V. Ravi, Mahil Carr and N. Raj Kiran, "Software development cost estimation using wavelet neural networks," *The Journal of Systems and Software*, vol. 81, pp. 1853-1867, 2008.

[20] K. Hammouda, "A comparative study of data clustering techniques," unpublished

[21] S. Mu-Chun and C. Chien-Hsing, "A modified version of the K-means algorithm with a distance based on cluster symmetry", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.23, pp. 674-680, 2001.

[22] S. Mu-Chun, C. Chen-Hsing and H. Chen-Chiung, "Fuzzy C-means algorithm with a point symmetry distance," *International Journal of Fuzzy Systems*, vol. 7, no. 4, pp. 175-181, 2005.
[23] C. Kuo-Liang and L. Jhin-Sian, "Faster and more robust point symmetry-based K-means algprithm," *Pattern Recognition*, vol. 40, pp. 410-422, 2007.
[24] D. Amaratunga and J. Cabrera, *Exploration and analysis of DNA microarray and protein array data*. John Wiley & Sons: New Jersey, pp. 123-126, 2004.
[25] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield and E.S. Lander, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, pp. 531-537, 1999.
[26] J. Khan, S.J. Wei, M. Ringnér, L.H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C.R. Antonescu, C. Peterson and P. Meltzer, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural network," *Nature Med*, vol. 7, pp. 673-679, 2001.
[27] C.L. Nutt, D.R. Mani, R.A. Betensky, P. Tamayo, J.G. Cairncross, C. Ladd, U. Pohl, C. Hartmann, M.E. McLaughlin, T.T. Batchelor, P.M. Black, A.V. Deimling, A.L. Pomeroy, T.R. Golub and D.N. Louis, "Gene expression-based classification of malignant gliomas correlates better with survival than histological classification," *Cancer Res*., vol. 63, pp. 1602-1607, 2003.
[28] S.L. Pomeroy, P. Tamayo, M. Gaasenbeek, L.M. Sturla, M. Angelo, M.E. McLaughlin, J.Y. Kim, L.C. GoumnerovaC, P.M. Black, C. Lau, J.C. Allen, D. Zagzag, J.M. Olson, T. Curran, C. Wetmore, J.A. Biegel, T. Poggio, S.E.S. Lander and T.R. Golub, "Prediction of central nervous system embryonal tumor outcome based on gene expression," *Nature*, vol. 415, pp. 436-442, 2002.

Zarita Zainuddin received the B. S. degree in Mathematics from Monmouth College, USA in 1979, M. Sc. in Applied Mathematics from Ohio University, USA in 1981, M. Sc. in Mathematics (Control Theory) from UMIST, UK in 1986 and Ph.D from Universiti Sains Malaysia in 2001.

She is currently an Associate Professor at the School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia. Her current interests include neural networks, image processing, bioinformatics and crowd dynamics.

Ong Pauline received the B. S. degree in Mathematical Sciences from Universiti Sains Malaysia, Malaysia, in 2004.

She is currently a Ph.D student in Applied Mathematics at the School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia. Her research interests include neural networks, image processing and bioinformatics.