A Hybrid Algorithm for the Initialization of Wavelet Neural Networks: Application in Epileptic Seizure Classification

Zarita Zainuddin, Kee Huong Lai, and Pauline Ong

Abstract—**Wavelet neural networks (WNNs) are a variant of artificial neural networks (ANNs), which are powerful mathematical modeling techniques that are used to model and study a variety of complex real-life problems. During the unsupervised learning stage, the translation vectors of the hidden nodes need to be determined. The conventional k-means and fuzzy c-means clustering algorithms have been used for this purpose. Nevertheless, these methods are prone and sensitive to initial cluster centers that have been randomly chosen. In this paper, a new hybrid clustering algorithm is presented. The type-2 fuzzy c-means clustering algorithm is hybridized with the metaheuristic harmony search algorithm. The new algorithm is then used to determine the translation vectors of WNNs. By incorporating the evolutionary harmony search algorithm in the clustering algorithm, the hybrid algorithm is able to escape from local minima while searching for the global minimum. To validate the effectiveness of the proposed algorithm, a real world problem of epileptic seizure classification problem is studied. Simulation results showed that the hybridized algorithm outperformed the stand-alone clustering algorithms.**

Index Terms—**Wavelet neural networks, k-means clustering, fuzzy c-means clustering, harmony search algorithm, epileptic seizure classification.**

I. Introduction

Artificial neural networks (ANNs) are powerful mathematical models that are inspired from biological neural networks. Both networks are concerned with how the interconnecting neurons process a massive amount of data at any given time. Some attractive features offered by ANNs include fault tolerance, massive parallel processing ability, and adaptive learning. Due to these properties, ANNs have been used in a wide range of application, such as classification, function approximation, and pattern recognition.

The feedforward multilayer perceptrons (MLPs) are one of the many different models of ANNs. Nonetheless, MLPs suffer from several disadvantages, such as the tendency to get trapped at local minima during the training stage and the failure to converge when high nonlinearities exist [1].

Zhang and Benveniste [2] introduced a new type of ANNs, called wavelet neural networks (WNNs). It was shown that this particular type of ANNs possesses the universal approximation property. An explicit link between the neural network coefficients and wavelet transforms was also developed.

In general, the design of WNNs includes several key areas. First, a learning algorithm needs to be specified to adjust the weights iteratively during the supervised learning process. Second, a suitable activation function needs to be chosen. It has been shown that some functions perform better than the other for certain application problems [1]. Third, an appropriate initialization of translation vectors and dilation parameter is also vital, as it will lead to simpler network architecture, and subsequently, shorter training time and higher accuracy.

This paper concerns the technique used for the selection of the translation vectors of WNNs. In [2], the translation vectors are chosen from the points located on the interval of the domain of the function used, whereas in [3], a dyadic selection scheme was used together with the conventional k-means clustering algorithm. In [4], the translation vectors were derived from the new input data. Meanwhile, an explicit formula was derived and proposed to compute the translation vectors of composite function WNNs [5]. In [6], an enhanced fuzzy c-means clustering algorithm, termed modified point symmetry-based fuzzy c-means algorithm, was proposed to initialize the translation vectors. The different aforementioned clustering algorithms aim at simpler algorithm complexity, as well as higher classification accuracy given by the WNNs.

The harmony search algorithm has been used to train ANNs, where an algorithm termed self-adaptive global best harmony search algorithm was used in the supervised training stage of ANNs [7]. In [8], the harmony search algorithm was hybridized with the k-means clustering algorithm. The new algorithm was then used in the task of documents clustering. In this paper, a hybridization of the type-2 fuzzy c-means clustering algorithm and the metaheuristic harmony search algorithm was considered. To demonstrate the effectiveness of the new hybrid algorithm, a real world problem, namely the task of epileptic seizure classification from electroencephalography (EEG) signals, was studied. For comparison purposes, other types of clustering algorithms, such as k-means and fuzzy c-means clustering algorithms were also considered.

The remaining of the paper is organized as follows. Section II presents the network architecture and learning

Manuscript received June 14, 2013; revised September 4, 2013. This work was supported by the Ministry of Higher Education, Malaysia under the FRGS research grant scheme and Universiti Sains Malaysia under the USM Fellowship Scheme.

Zarita Zainuddin and Kee Huong Lai are with the School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia (e-mail: zarita@cs.usm.my, laikeehuong1986@yahoo.com).

Pauline Ong is with Faculty of Mechanical and Manufacturing Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia (e-mail: ongp@uthm.edu.my).

DOI: 10.7763/IJAPM.2013.V3.235

algorithm of WNNs. Section III discusses the new hybrid clustering algorithm. Section IV studies the real world problem of epileptic seizure detection using the proposed algorithm. Section V gives the experimental results and discussion. Section VI concludes the paper.

II. Wavelet Neural Networks

A. Network Architecture

WNNs are feedforward neural networks with three layers, namely the input layer, the hidden layer, and the output layer. The input layer receives input and transmits the values received to the single hidden layer. Activation functions are embedded in the hidden nodes of hidden layer. These functions, such as Gaussian wavelet, Mexican Hat wavelet, and Morlet wavelet, perform the nonlinear mapping from the hidden layer to the output layer. Mathematically, a typical WNN is modeled by the following equation:

$$y(\mathbf{x}) = \sum_{i=1}^{p} w_{ij}\psi\left(\frac{\|\mathbf{x}-\mathbf{t}_i\|}{d}\right)+b \tag{1}$$

In (1), y is the desired output, $\mathbf{x}$ is the input vector, p is the number of hidden nodes, w_{ij} is the weight matrix, ψ is the activation function, $\mathbf{t}$ is the translation vector, d is the dilation parameter, and b is the bias term. The network architecture is shown in Fig. 1.

B. Learning Algorithm

After the type of activation function is defined, and both the translation vectors and dilation parameter are initialized, the wavelet activation function will compute the output values from the hidden layer. These values will then be sent to the output layer. During the supervised learning stage, the values of the weight matrix will be determined by solving a linear system.

Equation (1) can be rewritten in a more compact form $\mathbf{Y = GW}$, where $\mathbf{Y}$ is the output, $\mathbf{W}$ is the matrix that stores the weight values, and $\mathbf{G}$ is defined as follows:

$$\mathbf{G} = \begin{bmatrix} G(x_1,d_1,t_1) & G(x_1,d_2,t_2) & \cdots & G(x_1,d_p,t_p) \\ G(x_2,d_1,t_1) & G(x_2,d_2,t_2) & \cdots & G(x_2,d_p,t_p) \\ \vdots & \vdots & \vdots & \vdots \\ G(x_d,d_1,t_1) & G(x_d,d_2,t_2) & \cdots & G(x_d,d_p,t_p) \end{bmatrix}. \tag{2}$$

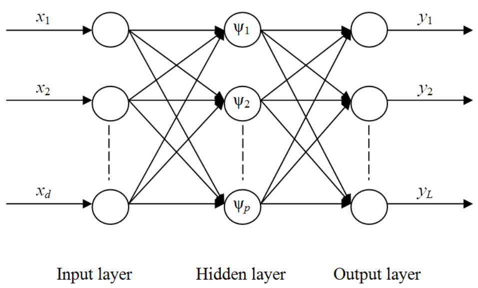

Fig. 1. Network architecture of a WNN.

In (2), the matrix $\mathbf{G}$ is usually a non-square matrix. Therefore, to solve for the weight matrix $\mathbf{W}$, the equation $\mathbf{W = G^{*}D}$ is used, where $\mathbf{G^*}$ is the pseudo-inverse, given by the formula $\mathbf{G}^* = \left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}$, where $\mathbf{G}^{\mathrm{T}}$ is the transpose of matrix $\mathbf{G}$. During the supervised training stage, the network will store the values of the weight values. The network is then ready for the testing stage.

III. Hybrid Algorithm

A. Fuzzy C-Means Clustering Algorithm

Bezdek et al. introduced the notion of fuzzy c-means (FCM) clustering in [9]. This algorithm aims to partition a set of data into several sets. Each datum belongs to all the clusters, each with its own membership, designated by the value stored in the partition matrix $U=[u_{ij}]$. The FCM algorithm minimizes the following objective function:

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}\left\|x_j - v_i\right\|^2. \tag{3}$$

In (3), J is the cost function, $m>1$ is the fuzzifier parameter whose value is normally set to be 2, c is the number of cluster centers, n is the number of data, and u_{ij} corresponds to the membership value for the data x_j and center v_i. The membership values are obtained as follows:

$$U=[u_{ij}] = \left[\left(\sum_{k=1}^{c}\left(\frac{\|x_j - c_i\|}{\|x_j - c_k\|}\right)^{\frac{2}{m-1}}\right)^{-1}\right]. \tag{4}$$

The cluster centers v_i are updated iteratively using the following equation:

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}},\ i=1,2,\ldots,c. \tag{5}$$

The FCM algorithm proceeds as follows:

1) Fix the number of cluster center, c.
2) Initialize the location of the centers randomly.
3) Compute the membership values using (4).
4) Update cluster centers using (5).
5) Repeat steps 3-4 until the locations of the centers stabilize.

B. Type-2 Fuzzy C-Means Clustering Algorithm

In [10], Rhee and Hwang proposed an extension to the conventional FCM clustering algorithm by introducing the notion of other membership grades to type-1 membership values. The new algorithm is named type-2 fuzzy c-means (T2FCM) clustering algorithm. The idea came from the fact that the conventional FCM clustering might result in undesirable clustering when noise exists in the input data. As such, a triangular membership is proposed, as shown in the following equation:

$$a_{ij} = u_{ij} - \left(\frac{1 - u_{ij}}{2} \right). \tag{6}$$

In (6), u_{ij} and a_{ij} are the type-1 and type-2 membership values, respectively. The type-1 membership value is the value obtained from the FCM algorithm, whereas the type-2 membership value is calculated using (6). In T2FCM clustering algorithm, the cluster centers are obtained from a new formula that has been modified accordingly, as shown in the following equation:

$$\hat{v}_i = \frac{\sum_{j=1}^{n} a_{ij}^m x_j}{\sum_{j=1}^{n} a_{ij}^m}. \tag{7}$$

C. Harmony Search Algorithm

The metaheuristic harmony search algorithm was inspired from the improvisation process of musicians [11]. The analogies of this algorithm to the actual usage in the field of music theory are described as follows [7]. A harmony in music is analogous to a solution vector; each musical instrument that produces a musical note is analogous to a decision variable; the pitch range of the musical instrument is analogous to the range of the decision variable; audience's aesthetics is analogous to the local and global search used during the optimization process. The algorithm requires fewer mathematical requirements [12]. Moreover, derivative information is not necessary as the algorithm uses stochastic random searches. In addition, since a total of 10 sets of solutions are stored in the harmony memory, it means that alternative solutions can also be generated for each of the optimization problem.

In general, a typical harmony search algorithm consists of five main steps [11], [12].

1) Step 1: Initialize parameters

In the first step, five important parameters are initialized. Harmony memory size (HMS) is the number of sets of solutions desired that will be stored in harmony memory (HM). Harmony memory consideration rate (HMCR) and pitch adjusting rate (PAR) are two parameters that are employed to find a new solution vector by searching the entire solution space both locally and globally. The best values of HMCR and PAR, which are between 0 and 1, are determined empirically [13]. Distance bandwidth (BW) refers to the step or increment size, which is found by multiplying a small value, usually 0.001, with the range xU-xL, where xU and xL are the upper bound and lower bound of a variable, respectively. The number of improvisations (NI) is used as the stopping criterion. The algorithm will terminate once it reaches the final iteration.

2) Step 2: Initialize harmony memory

The harmony memory (HM) is first stored with ten sets of possible solutions, which are generated randomly. The last column of the matrix stores the value of the cost function.

$$HM = \left[\begin{array}{cccc|c} x_1^1 & x_1^2 & \cdots & x_1^N & f(\mathbf{x}_1) \\ x_2^1 & x_2^2 & \cdots & x_2^N & f(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{\mathrm{HMS}}^1 & x_{\mathrm{HMS}}^2 & \cdots & x_{\mathrm{HMS}}^N & f(\mathbf{x}_{\mathrm{HMS}}) \end{array} \right]. \tag{8}$$

In (8), x_j^i denotes the value of the *j*th solution which corresponds to the *i*th decision variable. The function *f* is the cost function to be minimized.

3) Step 3: Improvise a new harmony

A new harmony, or a new possible solution, is generated based on the existing HM by utilizing either local search or global search. For each of the decision variable, a random number $0 < r < 1$ is generated, and then a new value, x' is proposed using the following three methods:

1) If $r < \mathrm{HMCR}$, x' will take the value of one of the value in the same column in HM.
2) If $r < \mathrm{PAR}$, a small increment size will be added to x'.
3) Else, x' will be chosen randomly from all the possible values in the entire solution space.

4) Step 4: Update harmony memory

For each of the newly generated solution vector, the value of the corresponding cost function will be calculated. The value obtained will be compared with all the existing values in the last column of HM. If the new solution is found to have a lower value (better solution), then it will replace the worst solution in the HM.

5) Step 5: Stopping criterion

The algorithm terminates when it reaches the maximum number of specified NI. The best solution is the one that gives the smallest value of the cost function. All the remaining possible solutions that are stored in HM are also near-optima, and hence can be used as alternative solutions.

D. The Hybridized Algorithm

In [14], the FCM algorithm is hybridized with the harmony search algorithm to find the optimal parameters in determining the best zone structures of aquifers. A similar algorithm is also considered in [15], where the new algorithm is used in the task of medical image segmentation.

Motivated by the advantages and attractive features offered by T2FCM and harmony search algorithms, a new hybrid algorithm, which combines these two algorithms, is proposed. The new algorithm, termed type-2 fuzzy c-means harmony search algorithm proceeds in a similar way as the five main steps for the conventional harmony search algorithm; however, some adjustments are made accordingly.

In the first step of the algorithm, the optimization problem is formalized. In other words, equation (3) is used as the cost function. The five parameters are initialized accordingly. In the second step, the harmony memory is generated, with a total of 10 solutions (cluster centers) stored in it. Next, a new harmony is generated. Based on the values of the newly generated candidate solution, the membership grade of each datum to each center is calculated using equation (4). Then, a new membership value will be recalculated by using the triangular membership function from the type-2 fuzzy c-means clustering algorithm, which is shown in equation (6).

The next step involves updating the cluster centers by using equation (7). This is followed by the computation of the value of the cost function. HM is updated if the new solution is found to have a lower value of cost function. The algorithm will terminate once it reaches its NI.

The mains steps of the hybrid algorithm are given as follows:

Step 1: Formalize optimization problem
- Minimize equation (3), the cost function
- Initialize the 5 parameters - HMS, HM, HMCR,
- PAR, BW

Step 2: Generate harmony memory
- Randomly generate 10 candidate solutions
- Compute the value of cost function and store in the last column of HM

Step 3: Improvise a new harmony
- Generate a new candidate solution based on *r*, HMCR, and PAR

Step 4: Reassign membership value
- Compute the membership grade of each datum using equation (4)
- Reassign a new membership grade using equation (6)

Step 5: Update cluster centers
- Find the new cluster centers using equation (7)

Step 6: Compute value of cost function
- Compute the value of cost function for the new candidate solution
- Store the value of cost function in the last column of HM

Step 7: Update harmony memory
- Compare the new solution in step 3 with those in HM
- Replace the worst solution with the new solution if the new solution has a lower value of cost function

Step 8: Terminate algorithm
- Repeat steps 3-7 until the stopping criterion is met

IV. Epileptic Seizure Classification

A. Epileptic Seizure

An estimated 50 million people worldwide suffer from the neurological disease of epilepsy seizure [16]. Electroencephalogram (EEG) is used as a vital tool in diagnosing this chronic disorder. The EEG signals will normally be recorded a few days before a patient undergoes a surgery. The biomedical signals will be generated and monitored. Nevertheless, scrutinizing the EEG signals is a very time consuming process. Therefore, there is a great need of the development of an expert system that can assist epileptologists in this matter. An automated system that is capable of differentiating normal EEG signals (interictal) from epileptic EEG signals (ictal) with high accuracy is highly desirable. It will not only save medical expenditure, but it also speeds up the pre-surgical evaluation process.

A number of classifiers have been proposed in the literature to study this problem. They include the use of radial basis function networks [17], recurrent neural networks [18], and probabilistic neural networks [19]. In this paper, WNNs will be used to study the binary classification problem of epileptic seizure. The WNNs model is chosen because of its compact topology and fast learning speed. Moreover, WNNs utilize localized wavelet activation functions instead of the global sigmoidal activation function in multilayer perceptrons. In this work, Tte translation vectors of the WNNs will be initialized using the new hybrid algorithm which combines the T2FCM and the harmony search algorithms.

B. Methodology

The EEG signal used in this study is obtained from a publicly available benchmark dataset [20]. The dataset consists of five different sets of data (A to E). Sets A to D are normal EEG signal recorded under different conditions, whereas set E consists of only epileptic EEG signal. Set A was recorded from healthy subjects with their eyes open, whereas set B with their eyes closed. On the other hand, set C was recorded from epileptic patients, from the hippocampal formation of the opposite hemisphere of the brain. Set D was also recorded from epileptic patients, but the area was within the epileptogenic zone. Set E was obtained from epileptic patients when they were experiencing seizure. Each set of data consists of 100 segments; with each segment is a time series with 4097 data points. Each segment was recorded for 23.6 s at a sampling rate of 173.61 Hz.

To extract the information embedded in the EEG signals, the technique of discrete wavelet transform (DWT) was used. In wavelet transform, a signal of interest is convolved with a chosen mother wavelet, $\psi(x)$. From the specified mother wavelet, a series of daughter wavelets can be generated via dilation and translation. A family of wavelets, $\psi_{a,b}(x)$, can be generated as shown in the following equation:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right), a \in \mathbb{R}_+, b \in \mathbb{R}, \tag{9}$$

where a and b are the dilation and translation parameters, respectively.

In continuous wavelet transform, the values of the dilation and translation parameters are varied continuously over the $\mathbb{R}_+ \times \mathbb{R}$ half-plane. Nevertheless, the continuous nature of the parameters yields an abundance of data, which are redundant. On the contrary, in DWT, the values of the two parameters are sampled discretely. In general, a dyadic scale is used. By choosing $a_0 = 2$ and $b_0 = 1$, a discrete subset of values $a = a_0^j = 2^j$ and $b = ka_0^j b_0 = k2^j \,\forall j,k \in \mathbb{Z}$ is generated. Equation (9) can be rewritten as follows:

$$\psi_{j,k}(x) = a_0^{-\frac{j}{2}}\psi\left(a_0^{-j}x - kb_0\right), j,k \in \mathbb{Z}. \tag{10}$$

Wavelet transform examines the signal in different frequency components, where each component corresponds to a resolution matched to its scale. In this study, the Daubechies wavelet of order 4 (db4) was used [21]. A scheme of four decomposition levels was employed. In each decomposition level, a signal is decomposed into coarse approximation, *a*, and detail information, *d*. After the feature extraction stage, five groups of wavelet coefficients were obtained, which correspond to different frequency. The given groups of frequency were:

1) d_1 (43.4-86.8Hz),
2) d_2 (21.7-43.4Hz),
3) d_3 (10.8-21.7Hz),
4) d_4 (5.4-10.8Hz), and
5) a_4 (0-5.4Hz).

To reduce the dimension of the input data, a feature selection stage was employed. For each of the five groups of wavelet coefficients obtained, four summary statistics were derived. They are:

1) the 90th percentile of the absolute values,
2) the 10th percentile of the absolute values,
3) the means of the absolute values, and
4) the standard deviation of the wavelet coefficients.

The activation function used for the WNNs in this study, as shown in Fig. 2, is the Morlet wavelet function, given by the following equation:

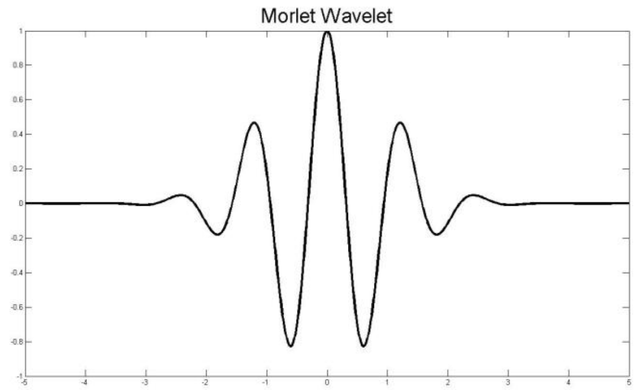

Fig. 2. The Morlet wavelet function.

$$\psi(x) = \cos(5x) \bullet \exp\left(-\frac{x^2}{2}\right). \tag{11}$$

During the training stage, a normal EEG signal and an epileptic EEG signal were indicated by the values 0 and 1, respectively. On the other hand, during the testing stage, a threshold value of 0.5 was chosen. In other words, any output equals to or greater than 0.5 will be reassigned a value of 1; otherwise, it will be reassigned a value of 0. As this paper investigates the effect of different initialization methods on the performance of wavelet neural networks, the number of hidden neurons was all set to the same number, using the rule of thumb $k = \sqrt{n/2}$, where k is the number of cluster centers, and n is the number of data [22].

A 10-fold cross validation was adopted in this study. The 100 signals from each of the five groups were partitioned into 10 groups. The first nine sets were used as training set whereas the last set was used as the testing set. The average of the ten performance metrics was computed and reported.

The performance of the WNNs was reported in the following three statistical measures:

$$\text{Sensitivity } = \frac{\text{TP}}{\text{TP}+\text{FN}} \times 100. \tag{12}$$

$$\text{Specificity } = \frac{\text{TN}}{\text{TN}+\text{FP}} \times 100. \tag{13}$$

$$\text{Overall accuracy } = \frac{\text{TN}+\text{TP}}{\text{TN}+\text{TP}+\text{FN}+\text{FP}} \times 100. \tag{14}$$

In equation (12), (13), and (14), TP, TN, FP, and FN stand for the case of true positive, true negative, false positive, and false negative, respectively. The simulation was carried out using the mathematical software MATLAB® version 7.10 (R2010a).

V. Results and Discussion

A. Results

The performance metrics obtained from WNNs with different methods of initialization is shown in Table 1. As shown in Table I, it can be observed that the WNNs that used the translation vectors generated by the k-means (KM) clustering method gave the poorest result, with an overall classification accuracy of 94.80%. When the harmony search algorithm was used together with the k-means algorithm, the accuracy increased to 97.20%. WNNs that used the conventional FCM clustering algorithm reported an overall classification accuracy of 97.15%. The hybridization of FCM with the harmony search algorithm improved the accuracy further to 97.50%. When T2FCM clustering was used, an accuracy of 98.87% was obtained. The best performance was obtained by the WNNs that employed the hybridized T2FCM and harmony search algorithms, which gave an overall classification accuracy of 99.15%. The receiver operating characteristics (ROC) curve of the best classifier that used the T2FCM and harmony search algorithm is shown in Fig. 3. The graph plots the values of sensitivity against 1 – specificity over a range of different threshold values.

B. Discussion

From Table I, it can be observed that all the clustering algorithms that have been hybridized with the evolutionary harmony search algorithm showed better performance than their stand-alone counterparts. A steady increase in the overall classification accuracy was noticeable when the KM algorithm was substituted with FCM, and subsequently T2FCM.

TABLE I: Performance Metrics

Initialization methods	Sensitivity	Specificity	Overall
K-means	85.00 ±1.28	97.30 ±0.18	94.80 ±0.55
K-means with harmony search	91.80 ±0.75	98.85 ±0.24	97.20 ±0.28
Fuzzy c-means	93.82 ±0.89	97.92 ±0.65	97.15 ±0.55
Fuzzy c-means with harmony search	92.00 ±0.93	98.97 ±0.25	97.50 ±0.32
Type-2 fuzzy c-means	94.96 ±0.28	99.43 ±0.17	98.87 ±0.15
Type-2 fuzzy c-means with harmony search	98.13 ±0.21	99.69 ±0.09	99.15 ±0.13

FCM outperformed the primitive KM algorithm because it uses soft clustering instead of crisp or hard clustering. In soft clustering, a datum is assigned to more than one cluster. On the contrary, KM that uses hard clustering assigns one datum to only one center, and this degrades the overall classification accuracy.

While FCM relies on one fuzzifier, T2FCM adds a second layer of fuzziness by assigning a membership function to the membership value obtained from the type-1 FCM membership values. Outliers or noise can be handled more efficiently and higher classification accuracy can be obtained via the introduction of this membership function. It is also worthwhile to point out that biomedical images and signals are usually contaminated by noise due to several factors such as human erroneous judgments and faulty machine calibration. Therefore, the use of fuzzifier parameters in this case is appropriate as they can handle the noise well.

The incorporation of harmony search algorithm in the conventional clustering algorithm is able to prevent the solution vectors from getting trapped at local minima. In addition, the two parameters used, namely, HMCR and PAR, can make sure that the entire solution space will be explored more thoroughly, both locally and globally.

C. Statistical Test

As shown in Table I, the values of three performance metrics, namely, sensitivity, specificity, and overall classification accuracy, were reported, together with the variance of the 10-fold cross validation methods. To further evaluate the statistical significance between two classifiers that employed different initialization methods, the McNemar's test was performed. The formula for the standardized normal test statistics is given as follows:

$$Z_{ij} = \frac{f_{ij} - f_{ji}}{\sqrt{f_{ij} + f_{ji}}}, \tag{15}$$

where Z_{ij} is the test statistics that measures the significance between the accuracies reported by classifier *i* and *j*, f_{ij} is the number of cases where the EEG signals were correctly classified by classifier *i*, but incorrectly classified by classifier *j*. The term f_{ji} is defined in a similar way. A 5% level of significance was chosen. In other words, the accuracies reported by two classifiers are said to differ significantly if $|Z_{ij}| > 1.96$. A positive value of Z_{ij} implies that classifier *i* outperformed classifier *j*.

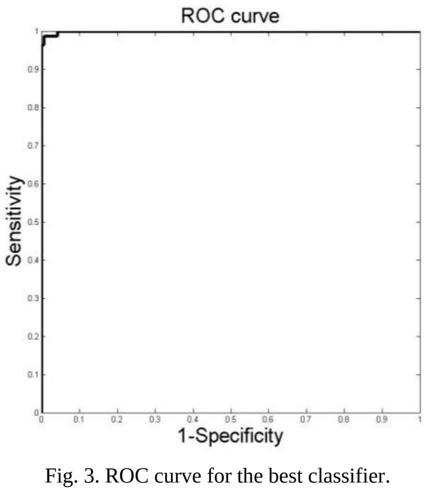

Fig. 3. ROC curve for the best classifier.

TABLE II: VALUE OF TEST STATISTICS FOR MCNEMAR'S TEST

Classifier	A	B	C	D	E	F
A	-	**2.71**	**2.31**	**3.21**	**3.44**	**3.41**
B		-	1.00	1.41	**2.53**	**2.89**
C			-	1.00	**2.33**	**3.00**
D				-	1.89	**2.65**
E					-	0.58
F						-

Keys: A: K-means (KM)
B: K-means with harmony search (KM-HS)
C: Fuzzy c-means (FCM)
D: Fuzzy c-means with harmony search (FCM-HS)
E: Type-2 fuzzy c-means (T2FCM)
F: Type-2 fuzzy c-means with harmony search (T2FCM-HS)

In Table II, the row refers to classifier *i*, whereas the column refers to classifier *j*. Table II is an anti-symmetric matrix, with the property $m_{ij} = -m_{ji}$. Therefore, the values of $M = [m_{ij} \mid i > j]$ have been omitted. The values that report significant difference between the two selected classifiers are highlighted in boldface. From Table 2, it is observed that in terms of statistical significance, classifier F outperformed classifier A to D, except classifier E. In other words, the performance of the wavelet neural networks models that was initialized using the T2FCM clustering algorithm was superior to those that were initialized using KM, KM-HS, FCM, and FCM-HS algorithms. In addition, the performance of classifier F was comparable to that of classifier E. Even though the reported values of overall classification accuracy of classifier F was higher than that of classifier E, further analysis revealed that the difference was not statistically significant.

D. Performance Comparison

The same dataset has been used for the same classification problem and reported in the literature. For performance comparison purposes, the results obtained in this study were compared with those reported by other researchers using different types of classifiers.

TABLE III: PERFORMANCE COMPARISON

Classifier	Accuracy	Ref.
Multilayer perceptrons with K-means clustering	99.60	[23]
Artificial neural network with time frequency analysis	97.73	[24]
Multilayer perceptrons with line length feature	97.77	[25]
Wavelet neural network with type-2 fuzzy c-means and harmony search	99.15	This work

As shown in Table III, multilayer perceptrons were used to perform the same classification task, namely, differentiating between the normal (set A to D) and epileptic (set E) EEG signals. Different feature selection techniques, such as the incorporation of k-means clustering method, time frequency analysis, and line length feature were considered. The overall classification accuracy obtained by using wavelet neuralnetworks, namely 99.15%, is comparable to those reported in [23]-[25].

In [23], the wavelet coefficients obtained from discrete wavelet transform were clustered using k-means algorithm for each of the frequency sub-band. On the other hand, in

[24], different values of frequency resolution, time windows, and frequency sub-bands were considered in order to extract the most suitable features. In [25], the line length feature, which represents the variation of the EEG signals' amplitude and frequency, was used in conjunction with multilayer perceptrons for the same classification problem.

Apart from the initialization methods used to initialize the translation vectors of the hidden nodes, feature selection is another important aspect to be considered to improve the overall classification accuracy. In this work, the features were obtained from the summary statistics of the wavelet coefficients derived from the method of discrete wavelet transform. For future work, different feature selection and dimensionality reduction methods will be considered.

VI. Conclusion

In this paper, a novel clustering algorithm, which combines the T2FCM and harmony search algorithms, was presented. The hybrid algorithm was used to locate the translation vectors of WNNs. The task of epileptic seizure classification was investigated to validate the robustness of the enhanced WNNs. Result showed that the new hybrid algorithm outperformed the KM and FCM algorithms. This suggests the feasible implementation of this automated classifier in the task of epileptic seizure detection to assist neurologists in their decision making process.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Ministry of Higher Education, Malaysia under the FRGS research grant scheme (203/PMATHS/6711323) and Universiti Sains Malaysia under the USM Fellowship Scheme.

References

[1] Z. Zainduddin and P. Ong, "Function approximation using artificial neural networks," *Int. J. Syst. App., Eng. & Dev.*, vol. 4, pp. 173-178, 2007.

[2] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. on Neural Networks*, vol. 3, pp. 889-898, 1992.

[3] K. Hwang, S. Mandayam, S. S. Udpa, and L. Udpa, "Characterization of gas pipeline inspection signals using wavelet basis function neural networks," *NDT & E Int.*, vol. 33, pp. 531-545, 2000.

[4] C. Lin, "Nonlinear systems control using self-constructing wavelet neural networks," *Appl. Soft Comput*., vol. 9, pp.71-79, 2009.

[5] J. Cao, Z. Lin, and G. B. Huang, "Composite function wavelet neural networks with extreme learning machine," *Neurocomputing*, vol. 73, pp. 1405-1416, 2010.

[6] Z. Zainuddin and P. Ong, "Reliable multiclass cancer classification of microarray gene expression profiles using an improved wavelet neural network," *Expert Syst. Appl*., vol. 38, pp. 13711-13722, 2011.

[7] S. Kulluk, L. Ozbakir, and A. Baykasoglu, "Training neural networks with harmony search algorithms for classification problems," *Eng. Appl. Artif. Intel.*, vol. 25, pp. 11-19, 2012.

[8] M. Mahdawi and H. Abolhassani, "Harmony K-means algorithm for document clustering," *Data Min. Knowl. Disc*., vol. 18, pp. 370-391, 2009.

[9] J. C. Bezdek, R. Ehrliuch, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geoscienes*, vol. 10, pp. 191-203, 1984.

[10] C. Hwang and C. F. Rhee, "A type-2 fuzzy c-means clustering algorithm," in *Proc. IEEE FUZZ Conference*, pp. 1926-1929, vol. 4, Melbourne, Australia, 2001.

[11] Z. W. Geem, J. H. Kim, and G. V. Loganathan, "A new heuristic optimization algorithm: harmony search," *Simulation*, vol. 76, pp. 60-68, 2001.

[12] S. K. Lee and Z. W. Geem, "A new metaheuristic algorithm for continuous engineering optimization: harmony search theory and practice," *Comput. Methods Appl. M.*, vol. 194, pp. 3902-3933, 2005.

[13] K. S. Lee, Z.W. Geem, S. Lee, and K. Bae, "The harmony search heuristic algorithm for discrete structural optimization," *Eng. Optimiz.*, vol. 37, pp. 663-684, 2005.

[14] M. T. Ayvaz, "Simultaneous determination of aquifer parameters and zone structures with fuzzy c-means clustering and meta-heuristic harmony search algorithms," *Adv. Water Resour*., vol. 30, pp. 2326-2338, 2007.

[15] O. M. Alia, R. Mandava, and D. Ramachandram, "Harmony search-based cluster initialization for fuzzy c-means segmentation of MR images," in *Proc. TENCON-IEEE Region 10 Conference*, pp. 1-6, 2009.

[16] World Health Organization (WHO). (2009). Epilepsy fact sheet. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs999/en/

[17] K. Aslan, H. Bozdemir, C. Şahin, S. Oğulata, and R. Erol, "A radial basis function neural network model for classification of epilepsy using EEG signals," *J. Med. Syst*., vol. 32, pp. 403-408, 2008.

[18] E. D. Übeyli, "Analysis of EEG signals by implementing eigenvector methods/recurrent neural networks," *Digit. Signal Process*, vol. 19, pp. 134-143, 2009.

[19] E. D. Übeyli, "Lyapunov exponents/probabilistic neural networks for analysis of EEG signals," *Expert Syst. Appl*., vol. 37, pp. 985-992, 2010.

[20] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, vol. 64, 2001.

[21] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *J. Neurosci. Meth*., vol. 123, pp. 69-87, 2003.

[22] V. M. Kanti, J. T. Kent, and J. M. Bibby, *Multivariate analysis*, London: Academic Press, 1980.

[23] U. Orhan, M. Hekim, and M. Ozer M, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Sys. Appl.*, vol. 38, pp. 13475-13481, 2011.

[24] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Automatic seizure detection based on time-frequency analysis and artificial neural networks," *Comput. Intell. Neurosci.*, pp. 1-13, 2007.

[25] L. Guo, D. Rivero, J. Dorado, J. R. Rabuñal, and A. Pazos, " Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks," *J. Neurosci. Methods*, vol. 191, pp. 101-109, 2010.

Zarita Zainuddin is a professor at the School of Mathematical Sciences, Universiti Sains Malaysia. She obtained her B. Sc. in mathematics from Monmouth College, Illinois, USA, and subsequently received an M. Sc. in Applied Mathematics from Ohio University, USA. She later obtained her PhD from Universiti Sains Malaysia in 2001. Her research interests include neural network learning algorithms, mathematical modeling, and crowd dynamics.

Kee Huong Lai is a Ph. D. student at the School of Mathematical Sciences, Universiti Sains Malaysia. He received his B. Sc. (Edu.) with a major in mathematics. His research interests include neural networks, metaheuristic algorithm, and biomedical signal processing.

Pauline Ong received her Ph. D. in applied mathematics from Universiti Sains Malaysia, Penang, Malaysia in year 2011. She is currently a senior lecturer with Universiti Tun Hussein Onn Malaysia, Johor, Malaysia. Her current research interests include artificial neural networks, mathematical modeling, and metaheuristic algorithm.