

# An Enhanced Wavelet Neural Network Model with Metaheuristic Harmony Search Algorithm for Epileptic Seizure Prediction

Zarita Zainuddin, Kee Huong Lai, and Pauline Ong

**Abstract**—The task of epileptic seizure prediction aims at differentiating between two classes of electroencephalography (EEG) signals, namely interictal and pre-ictal signals. The development of an automated classifier that is capable of performing such task with high sensitivity and low false positive rate is of paramount importance, as such classifier will improve the quality of life of patients diagnosed with epilepsy. In this paper, an enhanced wavelet neural network (WNN) model is proposed by incorporating the metaheuristic harmony search (HS) algorithm. The enhancement is accomplished via two modifications to the standard WNN model. First, a binary version of the HS algorithm is employed in the stage of feature selection, which aims at selecting the most optimal subset of input features for the WNN model during the preprocessing stage. Second, the HS algorithm is used to determine the best translation vectors for the hidden nodes of the WNN model. The simulation performed on the benchmark Freiburg dataset reported an average sensitivity of 85.55% and an average false positive rate of 0.22 per hour. It was found that the WNN model that gave the best performance is the one that employs the HS algorithm, in both feature selection and clustering stages. The satisfactory values of sensitivity and false positive rate obtained demonstrate the effectiveness of the proposed model for predicting the occurrence of impending seizures.

**Index Terms**—Epileptic seizure prediction, harmony search, feature selection, clustering, wavelet neural networks.

## I. INTRODUCTION

Epilepsy is a very common neurological disorder. Approximately 1% of the world population suffer from this chronic disease [1]. This illness, which is caused by excessive neuronal firing in the brain region, is characterized by the occurrence of recurrent seizures. Undoubtedly, epilepsy has serious economic and social implications on epileptic patients. They might encounter economic-related issues such as employment, insurance, and health-care needs. On the other hand, the social impacts on epileptic patients are no less pleasant. They live in constant fear of embarrassment. Besides, they tend to avoid social functions because of the discrimination and social stigma cast on them. To a greater

extent, epilepsy patients suffer from a reduced life quality due to the side effect of anticonvulsant drugs, as well as many other seizure-related injuries.

Given the astronomical number of patients diagnosed with epilepsy, and the serious life-threatening injuries that they might encounter, an alarm or a warning system that is capable of predicting an impending seizure attack is of great importance. An expert system that is able to perform such task would path the way for the realization of clinical epileptic seizure closed-loop intervention via drug administration and seizure warning devices such as vagus nerve stimulators. In other words, an impending seizure occurrence could be suppressed before its manifestation.

Electroencephalogram (EEG), which was first developed by German neuropsychiatrist Han Berger in 1924, is a very useful clinical tool that is used to monitor brain's activity. Apart from aiding syndrome classification, EEG also supports the general diagnosis of epilepsy. Furthermore, epileptologists are able to gain a better understanding of the mechanisms of ictogenesis [2].

A typical EEG signal can be further categorized into four different segments, namely interictal (seizure-free period), pre-ictal (before the occurrence of seizure), ictal (during the occurrence of seizure), and post-ictal (after the occurrence of seizure). These four distinct EEG segments are illustrated in Fig. 1.

In the task of epileptic seizure prediction, an automated classifier will be trained to perform a binary classification problem, that is, to distinguish between interictal and pre-ictal EEG signals. The existence of pre-ictal periods has been reported in [3], where the researchers found a decreased dynamical similarity minutes just before seizures from intracranial EEG recordings.

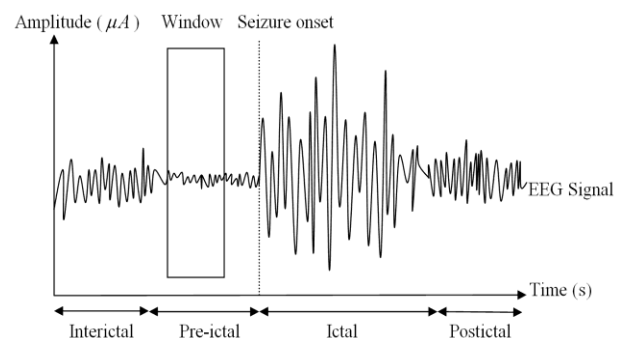


Fig. 1. Four different EEG segments.

Various methods have been reported in the literature of epileptic seizure prediction. In [4], a neural mass model,

Manuscript received September 29, 2014; revised December 12, 2014. 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, laikehuong1986@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).

which consists of pyramidal cells, excitatory and inhibitory interneurons, is introduced to model the dynamics of intracranial EEG data. It was reported that changes in the spatio-temporal parameter, which represent patients' preictal signatures, could aid in the task of epileptic seizure prediction. The results also suggested the existence of a pre-ictal period that precedes the onset of seizures. In [5], a feature space was derived from EEG signals by combining 22 linear univariate features. Different preprocessing and normalization methods were employed to study the biomedical signals. The nonlinear classification problem was accomplished using support vector machines (SVMs), where an average sensitivity of 73.9% and an average false positive rate of 0.15 per hour were reported. In [6], a probabilistic neural network (PNN) model was employed as the classifier. The problem of feature selection was addressed by means of genetic algorithm. The research reported a significance finding about the heterogeneity nature of epileptic patients. In other words, no two patients showed exactly identical EEG patterns, but in general, seizures of individual patients exhibited similar patterns.

Although a number of different classifiers and various preprocessing techniques have been used, there is still room for improvement in the area of epileptic seizure prediction. For instance, careful consideration should be given to the feature selection stage, because a good subset of features could not only get rid of irrelevant features, but it could also decrease the training time for classifiers. Moreover, some fine-tunings could be performed on the classifiers' architecture in order to achieve higher classification accuracy. In this work, the evolutionary harmony search (HS) algorithm is incorporated in the wavelet neural network (WNN) model in two aspects. First, a binary version of the HS algorithm is used to select the most optimal subset of features during the feature selection step. Second, the HS algorithm is employed to find the cluster centers or translation vectors of the hidden nodes of WNN. Both of these modifications aim at increasing the performance of the classifier.

The remaining of the paper is organized as follows. In Sect. II, the network architecture of WNN, as well as its learning algorithm are presented. In Section III, the concept of harmony search algorithm is examined. In Section IV, the two enhancements, namely in the steps of feature selection and the initialization of translation vectors, are discussed. In Section V, the problem of epileptic seizure prediction is studied by using the proposed enhanced WNN model. Section VI details the experimental results and discussion. Lastly, Section VII concludes the paper.

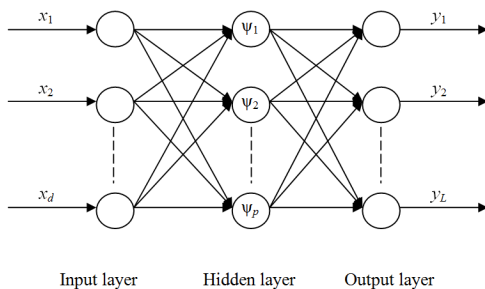


Fig. 2. Network architecture of a WNN model.

## II. WAVELET NEURAL NETWORKS

### A. Network Architecture

Introduced by Zhang and Benveniste [7], WNNs are a class of feedforward neural networks. WNNs are popular as learning models as they exhibit fast learning and are able to escape local minima. Unlike the conventional multilayer perceptrons (MLPs) that utilize global sigmoidal activation functions, WNNs employ localized wavelet activation functions. WNNs have been applied in various biomedical and industrial problems such as multiclass cancer classification of microarray gene expression profiles [8], air pollution [9], and pulping of the oil palm fronds [10]. This particular neural network model consists of three layers, namely the input layer, the hidden layer, and the output layer. The input layer receives input and propagates the values to a single hidden layer, where nonlinear mappings are performed by means of activation functions. The values obtained will then be summed up and sent to the final output layer. Mathematically speaking, a WNN is modeled by the following equation:

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

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,  $\psi$  is the activation function,  $\mathbf{t}$  is the translation vector,  $d$  is the dilation parameter, and  $b$  is the bias vector. The network architecture of a typical WNN model is shown in Fig. 2.

### B. Learning Algorithm

WNN employs a supervised learning algorithm where the values stored in the weight matrix  $w_{ij}$  will be determined by solving a system of linear equations. Observe that equation (1) can be rewritten in a more compact form  $\mathbf{Y} = \mathbf{G}\mathbf{W}$ , where  $\mathbf{Y}$  is the output,  $\mathbf{W}$  is the matrix that stores the weight values, and  $\mathbf{G}$  is defined as follows:

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

In (2), when matrix  $\mathbf{G}$  is a non-square matrix, the equation  $\mathbf{W} = \mathbf{G}^* \mathbf{D}$  is used to solve for the weight matrix  $\mathbf{W}$ .  $\mathbf{G}^*$  is the pseudo-inverse, where its formula is given by  $\mathbf{G}^* = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T$ , where  $\mathbf{G}^T$  is the transpose of matrix  $\mathbf{G}$ .

## III. HARMONY SEARCH ALGORITHM

Inspired from the improvisation process of musicians, the metaheuristic harmony search (HS) algorithm was introduced by Geem in 2001 [11]. The word harmony refers to the harmony in the field of music theory, where it is analogous to a solution vector in optimization problems. Each musical instrument is analogous to a decision variable; the musical instrument's pitch range 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 best harmony is analogous to the best solution that maximizes or minimizes a particular optimization problem [12]. The algorithm has been used widely in many engineering optimization problems because of its advantages over other metaheuristic algorithms. HS requires fewer mathematical requirements and derivative information is not necessary as it uses stochastic random searches [13]. For HS, the new candidate solutions are generated via all the previous solutions stored in the HM matrix. On the contrary, the new candidate solutions derived via the genetic algorithm (GA) only inherit information from two parents.

The five steps of the HS algorithm are detailed as follows: [13], [14].

A. Step 1: Initialize Parameters

In this step, five parameters (HMS, HMCR, PAR, BW, NI) are initialized. Harmony memory size (HMS) refers to the number of sets of solutions that will be stored in harmony memory (HM). On the other hand, harmony memory consideration rate (HMCR) and pitch adjusting rate (PAR) are two important parameters that are used to improvise a new candidate solution via global search and local search, respectively. Distance bandwidth (BW) is the value of 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 decision variable, respectively. The number of improvisations (NI) is used as the stopping criterion.

B. Step 2: Initialize Harmony Memory

The harmony memory (HM) is first stored with a pre-determined number of sets (usually ten) of solutions. These initial solutions are generated randomly. The final column of the matrix stores the value of the cost function.

$$HM = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^N & | & f(x_1) \\ x_2^1 & x_2^2 & \dots & x_2^N & | & f(x_2) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{HMS}^1 & x_{HMS}^2 & \dots & x_{HMS}^N & | & f(x_{HMS}) \end{bmatrix} \quad (3)$$

In (3),  $x_j^i$  is the value of the  $j$ th solution of the  $i$ th decision variable. The function  $f$  is the cost function to be maximized or minimized.

C. Step 3: Improvise a New Harmony

A new candidate solution, or a new harmony, is then generated based on the values stored in the existing HM. For each of the decision variable, a random number  $0 < r < 1$  is generated. Then, a new value,  $x'$  is derived using the following three rules:

- 1) If  $r < HMCR$ ,  $x'$  will take the value of one of the values stored in the same column in HM.
- 2) If  $r < PAR$ , 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.

D. Step 4: Update Harmony Memory

The quality of each of the newly generated candidate solution will be evaluated using the cost function given. The

value of the cost function will be stored in the last column and it will be compared with all the fitness values stored in HM. If the new solution is found to be a better solution, it will replace the worst solution in the existing HM.

E. 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 least or greatest value of the cost function.

IV. ENHANCEMENT OF WNN MODEL

A. Feature Selection

Feature selection is a well-known NP-hard combinatorial problem. It entails the search of an optimal subset of input features so as to maximize the overall classification accuracy of classifiers. By reducing the dimensions of input features, the redundant and irrelevant features can be eliminated and this saves computational cost. In general, feature selection algorithms falls into three main categories, namely wrapper, filter, and hybrid methods. The wrapper method is a classifier-dependent approach. This method is able to find a good subset of input features because it uses the classification accuracy reported by the classifiers. Nonetheless, for each selected feature subset, the classifier needs to be retrained. This implies that this approach incurs a very high computational cost because of the repetitive evaluation process.

The second method, which is called the filter approach, is much simpler and it uses a pre-determined evaluation metric to evaluate the quality or goodness of the selected features. This method is not computationally intensive and it can be performed very quickly because it does not require feedback from the classifier. Nevertheless, since this approach is independent of classifier, a feature subset of high quality does not necessarily give high classification accuracy when the input features are trained and tested using classifiers.

The third method, which combines the advantages of both the wrapper and filter approaches, is termed the hybrid approach. It is more computationally expensive but this is a small price to pay for its ability to be able to escape from local minima. From a mathematical point of view, feature selection can be formulated as optimization problems, where each feature can be included or excluded from the feature subset, with the classifier's classification accuracy being the function to be maximized.

An exhaustive approach could be used to find the optimal subset of input features but the technique is not plausible as it is too computationally expensive and time consuming. The number of possible subset grows exponentially with each added dimension. For an input with  $q$  features, there are a total of  $2^q - 1$  different combinations, provided that a minimum of one feature must be selected. Hence, the complication necessitates the use of metaheuristic methods to find reasonably good sub-optimal solutions without having to explore the entire solution space.

Most of the metaheuristic and evolutionary algorithms were first proposed and formulated for optimization problems that involve real variables. These algorithms have

been modified so that they can be used for problems that deal with integers and binary values. For instance, [15] reports a new version of ant colony optimization (ACO) algorithm that modifies the previously traversed paths to adjust the pheromone values. The proposed algorithm considers features with different priority and it achieves excellent performance in the task of feature selection. In [16], an improved binary particle swarm optimization (IBPSO) algorithm that utilizes the concept of pbest and gbest is proposed to study the feature selection problem for gene expression data.

The metaheuristic HS algorithm has also been considered to be used in the task of feature selection. In [17], the HS algorithm is combined with optimum-path forest classifier to study the problem of non-technical losses detection. In [18], the HS algorithm is used in conjunction with the concept of fuzzy-rough.

In this work, a novel variant of HS algorithm is proposed. Each input feature will be coded as either 0 (feature is excluded) or 1 (feature is selected). The harmony or a candidate solution is hence, represented by a binary string. The search for the optimal input subset is thus, equivalent to the search of the best combination of bits, through the Boolean hypercube. It is noted that the features are coded using only two distinct values and there exists no relation between each feature and its neighbors. The BW parameter is omitted in this HS-based approach.

Two major modifications are proposed to prepare the HS algorithm better suited for the task of feature selection. First, instead of randomly initializing only 10 solutions in HM, a total of 100 solutions is first generated, and the best 10 solutions with the highest fitness values (classification accuracy) are then included in the HM. This is based on the observation that once the HM is filled with 10 initial solutions, the newly generated candidate solutions will be derived from only these 10 solutions. Therefore, to increase the diversification or exploration of the search process, more solutions need to be generated initially, and only after that, the 10 solutions of the best quality will be included in the HM.

The second modification is regarding the use of the HMCR and PAR parameters. Because each decision variable can take only two values (0 or 1), the HMCR parameter will be used to determine whether the decision variable will be assigned the same value as the majority of the values located in the same column of HM. On the other hand, the PAR parameter will act as a bit-flip operator. The rules of step 3 of the HS algorithm are revised as follows, where  $x$  is the old value and  $x'$  is the new value:

- 1) If  $r < \text{HMCR}$ ,  $x' = x^{\text{major}}$ , where  $x^{\text{major}} = 0$  if there are more 0's in the same column, or  $x^{\text{major}} = 1$  if there are more 1's in the same column.
- 2) If  $r > \text{HMCR}$ ,  $x' = x^{\text{minor}}$ , where  $x^{\text{minor}} = 0$  if there are less 0's in the same column, or  $x^{\text{minor}} = 1$  if there are less 1's in the same column.
- 3) If  $r < \text{PAR}$ ,  $x' = x^{\text{flip}}$ , where  $x^{\text{flip}} = 0$  if  $x = 1$ , or  $x^{\text{flip}} = 1$  if  $x = 0$ .
- 4) If  $r > \text{PAR}$ ,  $x' = x$ , where  $x$  is the value of the decision variable after step (a) and (b) (value of  $x$  remains

unchanged).

### B. Initialization of Translation Vectors with HS Algorithm

The translation vectors refer to the coordinates or locations of the hidden nodes of WNN. A set of good translation vectors will ensure a better learning process, as well as faster convergence for the neural network model under consideration. From the literature, several approaches have been used to initialize the translation vectors, which include random points from the interval of the domain used [7], dyadic selection scheme with  $k$ -means clustering algorithm [19], and a novel fuzzy  $c$ -means clustering algorithm, named modified point symmetry-based fuzzy  $c$ -means algorithm [8].

In [20], the type-2 fuzzy  $c$ -means clustering algorithm is hybridized with the metaheuristic HS algorithm. While the decision variables in the case of feature selection are coded in binary values, the decision variables in the case of determining the translation vectors take the form of real values. The number of cluster centers is determined as follows:

$$k = \left\lfloor \sqrt{\frac{n}{2}} \right\rfloor. \quad (4)$$

In (4),  $k$  is the number of cluster centers, and  $n$  is the number of data points. Each candidate solution takes the form of (5).

$$\mathbf{x} = [x_1^1, x_2^1, x_3^1, \dots, x_d^1, x_1^2, x_2^2, x_3^2, \dots, x_d^2, \dots, x_1^k, x_2^k, x_3^k, \dots, x_d^k]. \quad (5)$$

In (5), where  $x_i^j$  denotes the  $i$ th value of the  $j$ th centers. In [20], it is shown empirically that the WNN model that is initialized with the hybrid clustering algorithm gives the best performance compared to other clustering algorithms. The hybrid algorithm adds a second layer of fuzziness by assigning a membership function to the original membership value. In addition, the incorporation of the HS algorithm is able to help the solution vectors escape local minima. The HMCR and PAR parameters used ensure that the entire solution space can be explored more thoroughly, both locally and globally. Readers are referred to paper [20] for the detail of the clustering algorithm.

## V. EPILEPTIC SEIZURE PREDICTION

### A. Data Acquisition

The EEG signal used in this study is obtained from a publicly available dataset recorded at the Epilepsy Center of the University Hospital of Freiburg, Germany [21]. The database consists of invasive EEG recordings of 21 epileptic patients. The EEG signals were recorded using a Neurofile NT digital video EEG system with 128 channels, sampled at 256 Hz. For each patient, the EEG signals were obtained from six electrodes; three of them were focal electrodes, selected from the seizure onset area, whereas the remaining three were extrafocal electrodes. A minimum of 24 hours of interictal EEG signals was recorded for each patient. In addition, each patient file contains at least 50 minutes of pre-ictal EEG recordings. The number of seizures

experienced by the patients ranged from 2 to 5. Out of these 21 patients, 10 patients, each with 5 seizures were selected for analysis. The summary of these 10 patients is shown in Table I.

### B. Feature Extraction and Selection

The technique of discrete wavelet transform (DWT) was used to extract the useful information embedded in the EEG signals. In this study, the Daubechies wavelet of order 4 (db4) was used [5], [22].

A total of five decomposition levels was employed. After the feature extraction stage, the six groups of wavelet coefficients that were obtained, which correspond to different frequency subbands.

TABLE I: INFORMATION OF THE 10 PATIENTS FROM THE DATABASE

| Patient | Sex | Age | Seizure origin     | Seizure analyzed | Interictal duration |
|---------|-----|-----|--------------------|------------------|---------------------|
| 3       | M   | 14  | Frontal            | 5                | 24h                 |
| 4       | F   | 26  | Temporal           | 5                | 24h                 |
| 5       | F   | 16  | Frontal            | 5                | 24h                 |
| 9       | M   | 44  | Temporal/Occipital | 5                | 24h                 |
| 10      | M   | 47  | Temporal           | 5                | 24h                 |
| 16      | F   | 50  | Temporal           | 5                | 24h                 |
| 17      | M   | 28  | Temporal           | 5                | 24h                 |
| 18      | F   | 25  | Frontal            | 5                | 25h                 |
| 20      | M   | 33  | Temporal/Parietal  | 5                | 26h                 |
| 21      | M   | 13  | Temporal           | 5                | 24h                 |

The six frequency subbands are  $d_1$  (64-128Hz),  $d_2$  (32-64Hz),  $d_3$  (16-32Hz),  $d_4$  (8-16Hz),  $d_5$  (4-8Hz), and  $a_5$  (0-4Hz), where  $d$  and  $a$  stand for detail coefficients and approximation coefficients, respectively.

For each signal, a total of 8 different summary statistics were derived from each of the six subbands, which are: (i) maximum, (ii) minimum, (iii) 90th percentile, (iv) 10th percentile, (v) mean, (vi) standard deviation, (vii) skewness, and (viii) kurtosis of the wavelet coefficients. Therefore, each EEG signal is represented or characterized by a vector with 48 values. The feature extraction stage is then followed by feature selection stage, which is described in Section IV. A. The values of the parameters used are:  $HM = 10$ ,  $HMCR = 0.7$ ,  $PAR = 0.3$ ,  $NI = 1000$ .

### C. Classification Using WNN

The WNN used to perform the classification task is initialized using the method described in Section IV. The Morlet wavelet, as shown in Fig. 3, is used as the activation function. The equation of the function is given as follows:

$$\psi(x) = \cos(5x) \times \exp\left(-\frac{x^2}{2}\right). \quad (6)$$

### D. Performance Evaluation

The results were reported in two performance metrics, namely sensitivity and false positive rate. In the survey reported in [23], the optimal time frame for a prediction time window, called seizure occurrence period (SOP) was investigated. It was reported that most patients preferred an SOP less than 1 hour. Three different SOP values are

considered in this work, namely 10 minutes, 20 minutes, and 30 minutes. The methodology is summarized in the block diagram, as shown in Fig. 4.

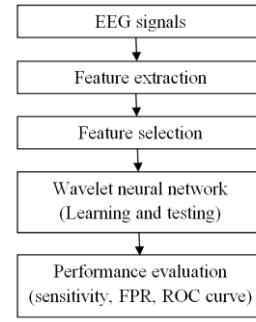


Fig. 4. Block diagram for epileptic seizure prediction.

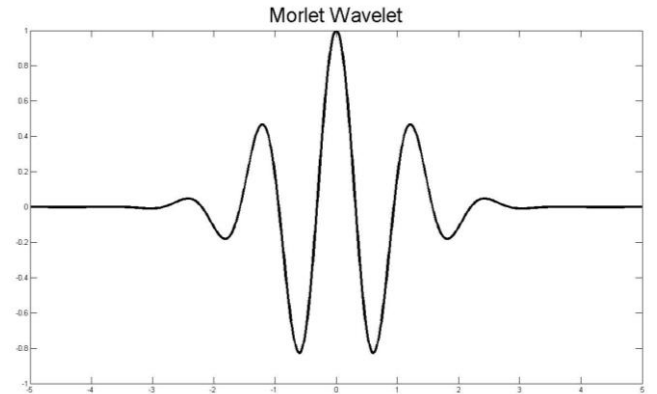


Fig. 5. The Morlet wavelet function.

## VI. RESULTS AND DISCUSSION

In this work, two enhancements on the architecture and configuration of the WNN model are proposed, which are feature selection and cluster initialization. For each of the ten patients selected from the Freiburg database, preictal and interictal EEG signals are extracted accordingly. A time window of 16 seconds is used. With 256 Hz sampling frequency, it means that each EEG segment contains 4096 data points. Each of the 10 selected patients experienced 5 seizures. Hence, a 5-fold cross validation is employed. The first four sets of data are considered the testing sets, whereas the last set of data is the testing set. The process is repeated 5 times so that each of the 5 seizures is tested.

It should be noted that during the feature selection stage, the WNN model was initialized randomly. Only after an optimal subset of input features is identified, the selected featured will be fed into the WNN model that is initialized using the proposed clustering algorithm. This is done to fine-tune the architecture of the WNN model so as to increase the classification accuracy. If both the feature selection and cluster initialization stages are performed simultaneously, it would be impractical as it would incur a very high computational cost.

For comparison purposes, different WNN models are also considered. The 4 models are described as follows: Model A does not employ any enhancements in both aspects. Model B only implements the improvement in the feature selection stage without the enhancement in clustering. Model C skips the enhancement in the feature selection stage and only

considers the improvement in the clustering stage. Model D utilizes both the enhancements in feature selection and cluster initialization.

#### A. Results

The results are reported in two performance metrics, namely sensitivity and false positive rates (FPR), across 3 different seizure occurrence periods (SOP), as shown in Table II.

The formulas for FPR and specificity are defined as follows:

$$\text{FPR} = \frac{\text{FP}}{\text{TN}+\text{FP}} \times 100, \quad (7)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}} \times 100, \quad (8)$$

In (7) and (8), the abbreviations FP and TN stand for false positive and true negative, respectively.

It should be noted that the relationship between specificity and FPR is given by:

$$\text{FPR} = 1 - \text{specificity}. \quad (9)$$

The ROC curve of the best classifier, namely model D, which utilizes the harmony search algorithm in both the feature selection and cluster initialization stages, is given in Fig. 5.

#### B. Discussion

From Table II, it is observed that the WNN model that gives the best performance (value highlighted in boldface) in terms of sensitivity is model D, which utilizes the WNN model that incorporates the HS algorithm in both the feature selection and cluster initialization stages. Model D, with 30 minutes of SOP, gives an average sensitivity value of 85.55% and an average FPR of 0.22 per hour.

In terms of the lowest FPR, model B (with enhancement only in the aspect of feature selection), with 10 minutes of SOP, reports an average FPR of 0.17 per hour. Despite this low value of FDR, this model also yields a relatively low sensitivity, at 71.23%.

It is evident from Table II that a longer time window of SOP increases the value of sensitivity. However, at the same time, it will increase the value of FPR as well.

While a high value of sensitivity is desirable, a high FPR is not acceptable because the patients will receive too many false alarms, which will affect their daily routines and hence, weaken their trust in the seizure prediction device. Therefore, there is a trade-off between these two parameters.

It is also noticed that model B gives better results compared to model C. This implies that feature selection plays a more crucial role compared to cluster initialization for the WNN model. By applying the additional feature selection step, irrelevant features that could degrade the performance of classifiers can be discarded. Furthermore, a good subset of input features that can best discriminate the two classes of data (pre-ictal and interictal) can be identified.

#### C. Performance Comparison

For performance comparison purposes, the results obtained in this work are compared with those reported by

other researchers in the literature using the same benchmark Freiburg dataset. Different classifiers have been proposed in the task of epileptic seizure prediction. They include neural mass model [4], support vector machines [5], and phase synchronization-based model [24].

In [4], a neural mass model is developed to study the problem of epileptic seizure prediction. Using different values of SOP, the values of sensitivity reported are 87.07% and 82.90%. The main finding of the work is that the spatio-temporal changes in the parameters derived from the EEG signals provide valuable information that characterizes the nature of the EEG signals during the pre-ictal stage.

In [5], 22 univariate features are generated from each of the 6 EEG signals. Therefore, each patient is characterized by a 132 dimensional feature space. The study investigates the effects of preprocessing and normalization of the input features. The work reports that the best configuration to be used is the one that utilizes the smoothing technique, combined with normalization by the maximum value.

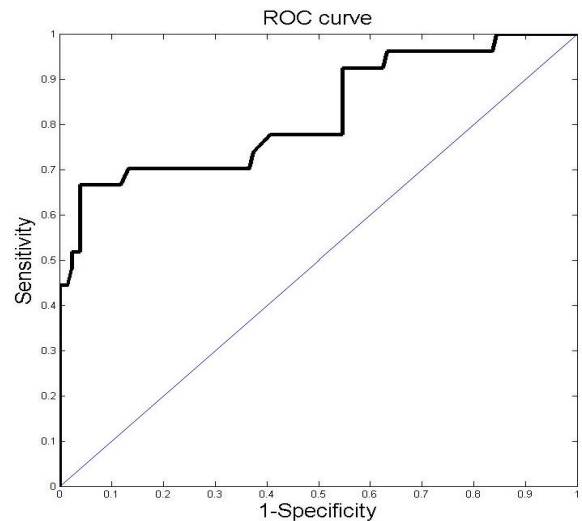


Fig. 6. ROC curve for model D.

Using a support vector machine (SVM) classifier, the technique reports an average sensitivity of 73.9% with an FPR of 0.15 per hour. The high value of sensitivity obtained demonstrates the importance of input feature preprocessing in the task of epileptic seizure prediction.

In [24], it is found that prior to seizure onset, both the increase and decrease in phase synchronization are reported. An expert system is developed based on the information obtained from this phase synchronization technique. The values of SOP and FPR are fixed at 10 min and 0.15 per hour, respectively. However, the average sensitivity value is not reported explicitly.

As shown in Table III, the value of sensitivity obtained by the model proposed in this work, namely 85.55%, is comparable to those reported from previous works.

In fact, the enhanced wavelet neural network model used in this work outperformed most of the other methods, except the model proposed in [4]. The minute difference in the values of sensitivity is most probably due to the fact that the EEG signals can be segmented from different frames.

To illustrate, there are at least 24 hours of interictal EEG recordings available for each patient, and there are many

different ways that can be employed to extract a small sample from this long segment of biomedical signal.

In terms of the value of false positive rate (FPR), all the models in Table III reported values ranging from 0.15-0.22 per hour. Some classifiers performed only marginally better than the others.

TABLE II: PERFORMANCE METRICS

| WNN Model                    | SOP (minutes) | Sensitivity (%) | FPR ( $h^{-1}$ ) |
|------------------------------|---------------|-----------------|------------------|
| Model A                      | 10            | 65.32           | 0.18             |
| (no enhancement              | 20            | 65.28           | 0.18             |
| in both aspects)             | 30            | 69.19           | 0.25             |
| Model B                      | 10            | 71.23           | <b>0.17</b>      |
| (enhancement in              | 20            | 76.32           | 0.19             |
| feature selection only)      | 30            | 78.74           | 0.21             |
| Model C                      | 10            | 68.76           | 0.24             |
| (enhancement in              | 20            | 69.97           | 0.24             |
| cluster initialization only) | 30            | 74.43           | 0.30             |
| Model D                      | 10            | 77.76           | 0.19             |
| (enhancement in both         | 20            | 78.83           | 0.24             |
| feature selection and        | 30            | <b>85.55</b>    | 0.22             |
| cluster initialization)      |               |                 |                  |

TABLE III: PERFORMANCE COMPARISON

| Method                      | Sensitivity (%) | FPR ( $h^{-1}$ ) |
|-----------------------------|-----------------|------------------|
| Neural mass model with      |                 | 0.20             |
| excitatory and inhibitory   | 87.07           |                  |
| interneurons [4]            | 82.90           | 0.16             |
| Support vector machines     |                 | 0.15             |
| with features preprocessing | 73.90           |                  |
| and normalization [5]       |                 |                  |
| Phase synchronization with  |                 | 0.15             |
| bivariate empirical mode    | -               |                  |
| decomposition [24]          |                 |                  |
| Wavelet neural network      |                 | 0.22             |
| with enhancement in feature | 85.55           |                  |
| selection (this work)       |                 |                  |

## VII. CONCLUSION

In this paper, two improvements are proposed and applied to the standard WNN model. Both enhancements are accomplished by incorporating the metaheuristic HS algorithm. A binary version of the HS algorithm is used in the feature selection stage. This is followed by the use of the standard HS algorithm in the clustering stage to find the best coordinates of the translation vectors of hidden nodes of WNN. The enhanced WNN is then tested on the real world problem of epileptic seizure prediction. Simulation results show that the WNN model that yields the greatest sensitivity and lowest false positive rate is the one that uses both the enhancements in the feature selection and cluster initialization stages. There is also a trade-off between the values of sensitivity and false positive rates across different seizure occurrence period. The satisfactory values of sensitivity and false positive rates obtained from the simulation demonstrate that the proposed enhanced WNN model could be used to aid in the task of epileptic seizure prediction to improve the quality of life of epileptic patients. For future works, different training algorithms could also be used and tested. The proposed classifier could also be tested

on the more comprehensive European epilepsy database (EPILEPSIAE) to demonstrate its robustness.

## ACKNOWLEDGMENT

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

## REFERENCES

- [1] L. D. Iasemidis, D. S. Shiau, W. Chaovalitwongse, J. C. Sackellares, P. M. Pardalos, J. C. Principe, P. R. Carney *et al.*, "Adaptive epileptic seizure prediction system," *IEEE Trans. on Biomed. Eng.*, vol. 50, pp. 616-627, 2003.
- [2] C. D. Binnie and H. Stefan, "Modern electroencephalography: Its role in epilepsy management," *Clin. Neurophysiol.*, vol. 110, pp. 1671-1697, 1999.
- [3] L. V. Quyen, J. Martinerie, M. Baulac, and F. Varela, "Anticipating epileptic seizure in real time by a nonlinear analysis of similarity between EEG recordings," *Neuroreport.*, vol. 10, pp. 2149-2155, 1999.
- [4] A. Aarabi and B. He, "Seizure prediction in hippocampal and neocortical epilepsy using a model-based approach," *Clin. Neurophysiol.*, vol. 125, pp. 930-940, 2014.
- [5] J. Rasekhi, M. R. K. Mollaei, M. Bandarabadi, C. A. Teixeira, and A. Dourado, "Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods," *J. Neurosci. Meth.*, vol. 217, pp. 9-16, 2013.
- [6] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echaz, and B. Litt, "Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients," *IEEE T. Biomed. Eng.*, vol. 50, pp. 603-615, 2003.
- [7] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. Neural Networ.*, vol. 3, pp. 889-989, 1992.
- [8] 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.
- [9] Z. Zainuddin and P. Ong, "Modified wavelet neural network in function approximation and its application in prediction of time-series pollution data," *Appl. Soft Comput.*, vol. 11, pp. 4866-4874, 2011.
- [10] Z. Zainuddin, W. R. W. Daud, P. Ong, and A. Shafie, "Wavelet neural networks applied to pulping of oil palm fronds," *Bioresour. Technol.*, vol. 102, pp. 10978-10986, 2011.
- [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. 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.
- [13] K. S. Lee and Z. W. Geem, "A new metaheuristic algorithm for continuous engineering optimization: harmony search theory and practice," *Comput. Method Appl. M.*, vol. 194, pp. 3902-3933, 2005.
- [14] 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.
- [15] R. Forsati, A. Moayedikia, R. Jensen, M. Shamsfard, and M. R. Meybodi, "Enriched ant colony optimization and its application in feature selection," *Neurocomputing*, vol. 142, pp. 354-371, 2014.
- [16] L. Chuang, H. Chang, C. Tu, and C. Yang, "Improved binary PSO for feature selection using gene expression data," *Comput. Biol. Chem.*, vol. 32, pp. 29-38, 2008.
- [17] C. C. O. Ramos, A. N. Souza, G. Chiachia, A. X. Falcao, and J. P. Papa, "A novel algorithm for feature selection using harmony search and its application for non-technical losses detection," *Comp. Elect. Eng.*, vol. 37, pp. 886-894, 2011.
- [18] R. Diao and Q. Shen, "Feature selection with harmony search," *IEEE T. Syst. Man. Cy. B*, vol. 37, pp. 663-684, 2005.
- [19] K. Hwang, S. Mandayam, S. S. Udpa, and L. Udpa, "Characterization of gas pipeline inspection signals using wavelet basis function neural networks," *NDT and E Int.*, vol. 33, pp. 531-545, 2000.
- [20] Z. Zainuddin, K. H. Lai, and P. Ong, "A hybrid algorithm for the initialization of wavelet neural networks: application in epileptic seizure classification," *Int. J. Appl. Phy. Maths.*, vol. 3, pp. 352-358, 2013.
- [21] T. Maiwaid, M. Winterhalder, R. A. Scheibe, H. U. Voss, A. S. Bonhage, and J. J. Timmer, "Comparison of three nonlinear seizure

prediction methods by means of the seizure prediction characteristic,” *Physica D*, vol. 194, pp. 357-368, 2004.

- [22] 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.
- [23] A. S. Bonhage, F. Sales, K. Wagner, R. Teotonio, A. Carius, A. Schelle, and M. Ihle, “Views of patients with epilepsy on seizure prediction devices,” *Epilepsy Behav.*, vol. 18, pp. 388-396, 2010.
- [24] Y. Zheng, G. Wang, K. Li, G. Bao, and J. Wang, “Epileptic seizure prediction using phase synchronization based on bivariate empirical mode decomposition,” *Clin. Neurophysiol.*, vol. 125, pp. 1104-1111, 2014.



**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. She was awarded an M.Sc. in applied mathematics by Ohio University, USA. Dr. Zarita received her Ph.D. from Universiti Sains Malaysia. Her research interests include neural networks 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.