

Research on Brain Signals via Artificial Neural Network and Swarm Intelligence Algorithms

Sema YILDIRIM*¹, Hasan Erdinc KOCER², A. Hakan EKMEKCI³

Received: 10/26/2018 Accepted : 05/16/2019

Abstract: Artificial Neural Networks (ANNs) that are the ability to learn from their environment in order to improve their performance are widely used in numerous applications. The Backpropagation (BP) Algorithm is one of the most popular and effective model of ANNs. However, since it uses gradient descent algorithm, which attempts to minimize the error of the network by moving gradient of the error curve, easily get trapped at local minima. In order to avoid this problem and to obtain a better classifier, we proposed an ANNs and Swarm Intelligence (SI) method where Artificial Bee Colony and Particle Swarm Optimization algorithms were operated for the Multilayer Perceptron Neural Network. Two Electroencephalogram (EEG) datasets were used to test to test the accuracy and success of the study performed. Compared to conventional-MLPNN, higher success values were obtained on each dataset with the proposed methods. Experimental results demonstrate that combined SI and MLPNN algorithm has been increased the success of BP algorithm by avoiding local minima. For ABC data, respectively, ABC-MLPNN and PSO-MLPNN methods, 79.00% and 75.50% respectively for Boston data and 91.67% and 88.33% respectively for Selcuk data were obtained. On the other hand, with the MLPNN algorithm, the success rate was 68.50% for Boston data and 81.67% for Selcuk data. These results show that the success of MLPNN algorithm significantly increases with the weights obtained by using SI. In addition to this, this study showed that the SI-MLPNN algorithm can be used on non-linear and highly complex EEG data.

Keywords: Artificial neural networks, backpropagation, brain disorders, electroencephalogram, feature extraction/selection, swarm intelligence.

1. Introduction

Artificial Neural Networks (ANNs) can effectively make a decision about the class of the signal. Therefore, neural networks have been successfully applied for so many medical applications [1]. Spectral analysis is a well-known method for analyzing EEG signal (EEGs). Nowadays ANNs may offer a superior performance for analysis of EEGs, compared to the spectral analysis methods [2]. On the other hand, the Backpropagation (BP) Algorithm which is a technique of ANNs falls into the problem of local minima because of uses gradient information. To avoid this problem, a host of other algorithms have been explored for Multilayer Perceptron Neural Network (MLPNN) training. Swarm Intelligence-based techniques can be used in a number of applications such as controlling unmanned vehicles, self-assembly and interferometry, planetary mapping, controlling nanobots within the body, killing cancer tumors and data mining. Swarm Intelligence (SI) algorithms inspired from nature are one such alternative. Artificial Bee Colony (ABC) and Particle Swarm Optimization (PSO) are some of the SI algorithms that have been used for training ANNs. However, there is currently no intelligent theory based on the complexity of the diagnostic disease. Up to now, no study has been reported in literature

related to SI-based MLPNN classification for analysis of EEGs. EEG signals are non-linear signals that are quite difficult and complex to interpret in biomedical engineering. Another important contribution of this study is that the feature vector, which is extracted by different methods of extraction, is reduced by the eigenvector method to provide a faster and robust structure.

The rest of this paper has been organized as follows: Section 2 describes a brief overview the related materials and methods. The proposed method has been explained in Section 3, which is then followed by the performed experiments and obtained experimental results for the proposed methods has been explained in Section 4. Finally, we summarized the most relevant conclusions and discussion of this work in Section 5.

2. Materials and Methods

Signal processing techniques can generally be considered in three key stages required for the analysis of EEGs namely: pre-processing, feature extraction/selection and classification as shown in Figure 1.

¹ GSNSc, Computer Eng., Konya Technical Univ., Konya-42002, TURKEY
ORCID ID : 0000-003-0807-8550

² Department of EEE, Selcuk Uni., Konya-42002, TURKEY
ORCID ID : 0002-0799-2140

³ Department of Neurology, Selcuk Uni., Konya – 42002, TURKEY
ORCID ID : 0000-0001-5008-7915

* Corresponding Author: Email: semayildirim@selcuk.edu.tr

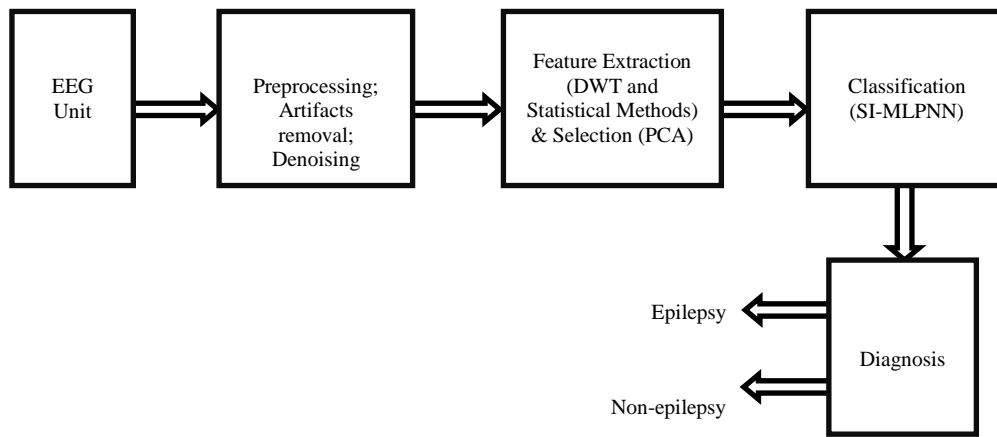


Fig.1. General Block diagram of the signal processing operations

The first stage of signal processing is the pre-processing that involves transforming raw data into an understandable format. This process, which is commonly used as a data mining application in advance, transforms data into a format that can be processed more easily and effectively for the user. Pre-processing usually contains sampling, denoising, normalization, filtering, artifact rejecting, etc.

Feature extraction/selection is very important role for classification methods. Feature extraction is the determination of a feature or a feature vector from a pattern vector. For pattern processing problems to be tractable requires the conversion of patterns to features, which are condensed representations of patterns, ideally containing only salient information [3]. In the feature extraction stage, numerous different methods can be used so that several diverse features can be extracted from the same raw data. Feature selection methods provides us a way of reducing computation time, improving prediction performance, and a better understanding of the data in machine learning or pattern recognition applications [4]. The focus of feature selection is to select a subset of variables from the input which can efficiently describe the input data while reducing effects from noise or irrelevant variables and still provide good prediction results [5].

The last step in signal processing, classification process is to check the identity of the input vectors according to the feature vectors stored in the database. Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. The simplest type of classification problem is binary classification. In binary classification, the target attribute has only two possible values. In the model build (training) process, a classification algorithm finds relationships between the values of the predictors and the values of the target. Different classification algorithms use different techniques for finding relationships. These relationships are summarized in a model, which can then be applied to a different data set in which the class assignments are unknown [6].

2.1. Electroencephalogram

Epilepsy is a disease where neurological disorders stem from temporary abnormal discharges in the brain's electrical activity [7], [8]. The crucial characteristic of this disease is repetitive

seizures. These seizures may sometimes not be noticed [9]. Electroencephalogram (EEG) is a clinical monitoring tool that records electrical activity of the brain signals which contain valuable information for understanding epilepsy. The detection of seizures occurring in the EEGs are an important component in the diagnosis and treatment of epilepsy. In general, long-term EEG monitoring records EEG for periods longer than routine EEG recordings of 20 to 30 minutes. However, interpretation of EEGs is a time consuming and expensive process because it involves large amounts of data. Large amounts of data are generated by EEG monitoring systems for electroencephalographic changes, and their complete visual analysis is not routinely possible. Prolonged EEG that lasts hours to days creates a large volume of EEG for interpretation. Computer-assisted analysis has become widely available to allow review of selected EEG during specific times of interest [10]. Therefore, developing automatic seizure detection methods is of great significance for reviewing EEGs [11].

In the last two decades, many researchers addressed to this problem. Fu [12] performed the classification of seizures based on the time-frequency imaging of EEGs using the Hilbert-Huang Transform (HHT) and Support Vector Machine (SVM) with Radial Basis Function (RBF). Joshi [13] carried out classification of EEGs using fractional linear prediction. A comparative study of wavelet families for EEGs classification was performed by Gandhi [14] using a Probabilistic Neural Network (PNN) with SVM. Lee [15] proposed a Neural Network with Weighted Fuzzy Membership (NEWFM) to classify EEGs, while Aydın [16] developed a classification using a Multilayer Neural Network (MLNN) architecture with respect to several time domain entropy measures on EEG series. A hybrid PSO integrating neural network with a fuzzy membership function (NEWFM) technique was proposed for epileptic seizure classification tasks by Abuhasel [17]. The work of Satapathy [18] analyzes the epileptic disorder in the human brain through EEGs analysis by integrating the best attributes of ABC and radial basis function networks (RBFNNs). Dehuri [19] employed for some Unique Client Identifier (UCI) data which include fisher iris, pima-diabetes and shuttle data classification using ABC trained MLPNN.

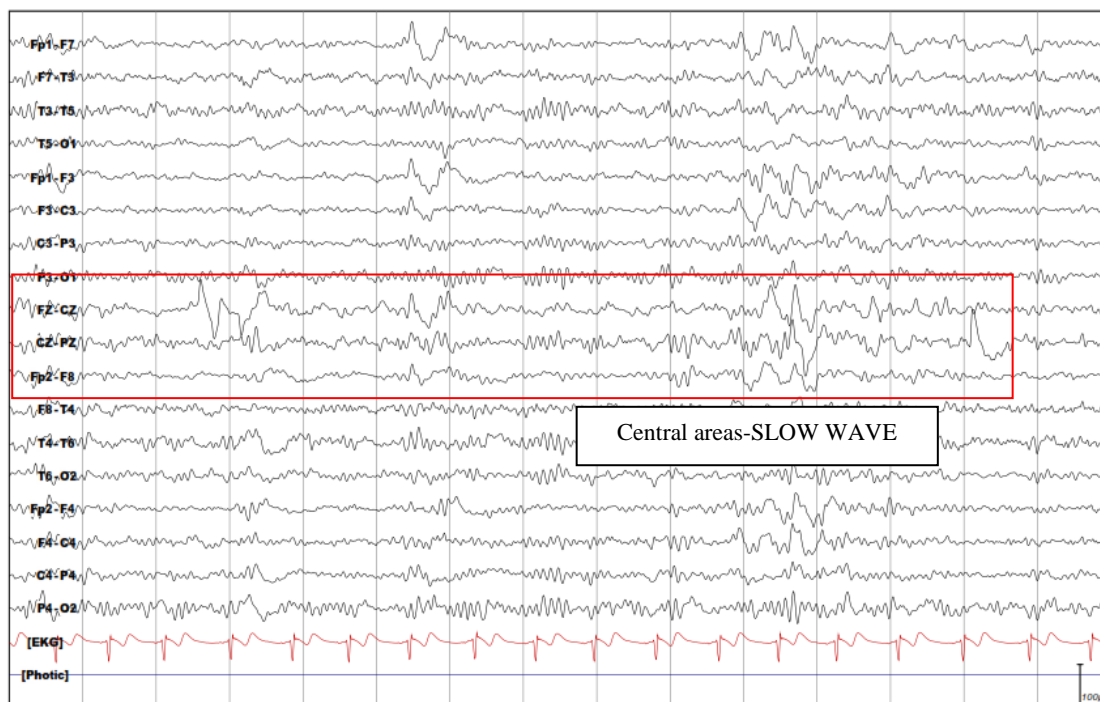


Fig.2. EEG record of an obtained epilepsy patient from Selcuk University Hospital (Page Size: 15 secs/page, Sensitivity: 150 μ Vp-p, Filters: LF 1.0Hz Notch, HF 15.0Hz Notch, Page 17 of 80).

Ground activity is well developed in posterior regions of hemispheres and contains 8-13 Hz alpha waves as shown in Figure 2. The neurologist interpreted that there was no significant asymmetry between the hemispheres in clinical information. Furthermore, slow-wave paroxysms were observed that condensed in the central areas at infrequent intervals along the trace.

2.2. EEG Dataset

In order to investigate the classification accuracy of the proposed methods, the data sets taken from Boston Children's Hospital and Selcuk University Medical Faculty Hospital were used. The first dataset collected from Boston Children's Hospital consists of EEG recordings from pediatric subjects with intractable seizures. Subjects were monitored for up to several days following withdrawal of anti-seizure medication in order to characterize their seizures and assess their candidacy for surgical intervention. Recordings, grouped into 23 cases, were collected from 22 subjects. All signals were sampled at 256 samples per second with 16-bit resolution. Two hundred recordings were included in this study, the first 100 recordings of which belonged to 22 seizure and non-epilepsy people in Boston Children's Hospital [20].

The second dataset has been obtained from Department of Neurology of Selcuk University Hospital, retrospectively. The EEGs taken from surface on brain were carried out on 60 patients using the 10-20 international system of electrode placement. The study used recordings belonging to 60 patients (30 epilepsies and 30 non-epilepsies) recorded between 2012 and 2015. All EEGs were obtained from routine EEG recording and all subjects were awake. The mean age of the 30 epilepsy patients was 38 (14 males, 16 female), and the mean age of the 30 non-epilepsy patients was 46 (14 males, 16 female); all EEG tracing has been electronically recorded. Additionally, EEG recordings on different dates for two patients who are epilepsy and undergoing

treatment were also included in our work. The EEG recordings consist of 18 channels at a sampling frequency of 200 Hz and an epoch of 15 seconds. Recording times were different in each case because the situation for each subject was different at the time they were recorded. This dataset, created by specialists, included EEGs with various waves such as sharp waves, spike, spike and slow waves, multiple spike and slow wave complexes. In this work, we used 5 epochs with both epilepsy and non-epilepsy patients, and the length of each epoch was composed of 3000 samples (5 epochs \times 200 Hz sampling frequency).

2.3. Feature Extraction/Selection

2.3.1. Discrete Wavelet Transform

Recently, many nonlinear and nonstationary methods [21], [13] have been suggested to extract signal processing parameters. The wavelet transform (WT) has been found to be particularly useful for analyzing signals that can best be described as a periodic, noisy, intermittent, and transient and so on. Its ability to examine the signal simultaneously in both time and frequency in a distinctly different way from the traditional Short-Time Fourier transform (STFT) has spawned an ever-increasing number of sophisticated wavelet-based methods for signal manipulation and interrogation. Wavelets are used to transform the signal under investigation into another representation which presents the signal information in a more useful form [22]. The main advantage of the WT is that it has a varying window size, being broad at low frequencies and narrow at high frequencies, thus leading to an optimal time-frequency resolution in all frequency ranges [23]–[28].

Discrete wavelet transform (DWT) is generally used because the calculation of wavelet coefficient ratios at every possible scale requires a great deal of effort and can result in a large amount of data [8]. Wavelets provide a time-scale information of a signal, enabling the extraction of features that vary in time [29]. DWT

analyzes the signal at different frequency bands and different resolutions by decomposing the signal into a coarse approximation and detailed information. DWT employs two sets of functions, called scaling functions and wavelet functions, which are associated with low-pass and high-pass filters, respectively. The decomposition of the signal into different frequency bands is simply obtained by successive high-pass and low-pass filtering of the time-domain signals [30].

2.3.2. Statistical Feature Extraction

Feature extraction plays an important role in pulling out special patterns (features) from the original data for reliable classification. The feature extraction stage must reduce the original data to lower dimensions that contain most of the useful information included in the original vector. It is therefore necessary to find out the key features that represent the whole dataset, depending on its characteristics [31]. Some statistical features are extracted from the data of each channel as the most representative values to describe the original signals. The following ten statistical features of each channel of EEG data are used as valuable parameters in the representation of the characteristics of the original EEGs.

- (i) Minimum
- (ii) Maximum
- (iii) Mean
- (iv) Median
- (v) Interquartile range
- (vi) Range
- (vii) Standard deviation
- (viii) Variance
- (ix) Kurtosis
- (x) Skewness

2.3.3. Principal Component Analysis

Principal Component Analysis (PCA) is the one of the most known methods for dimension reduction. The goal of PCA is to reduce the dimensionality of the data while retaining as much as possible of the variation present in the original dataset [32]. PCA transforms a high-dimensional dataset (of m dimensions) to a low-dimensional orthogonal feature (Eigenvector) space (of n dimensions, $m > n$) while retaining the maximum variance of the original high dimensional dataset. Each resulting orthogonal feature is referred to as a Principal Component (PC). Eigenvalues are scalar representations of the degree of variance within the corresponding PCs. PCs are ranked by their corresponding eigenvalues, and thus, the first PC captures the most significant variance in the dataset. The second PC is perpendicular to the first PC and it contains the next significant variance [33]. It is mostly useful for segmenting signals from multiple sources such as EEGs. The knowing of number of independent components in advance is very useful. Therefore, we preferred PCA that is one of the best dimension reduction methods for reduction the number of features on the second stage of signal processing.

2.4. Artificial Neural Network

ANNs are an information-processing system that is based on a simulation of the human cognitive process. In ANNs, knowledge about the problem is distributed through the connection weights of the links between neurons as shown in Figure 3.

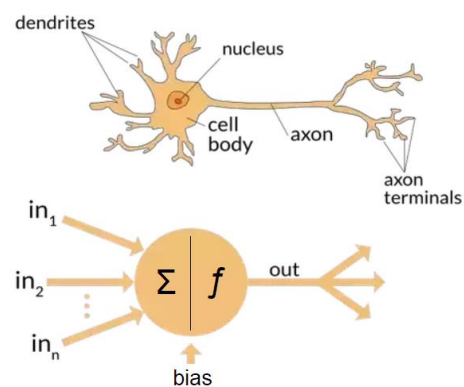


Fig.3. A biological and an artificial neuron [34]

The neural network has to be trained to adjust the connection weights and biases in order to produce the desired mapping. ANNs are widely used in the biomedical field for modeling, data analysis and diagnostic recognition. The ANN's capability to learn examples, the ability to reproduce arbitrary non-linear functions of input, and the highly parallel and regular structure makes them especially suitable for pattern recognition problems [35].

Multilayer perceptron neural network (MLPNN) is the most commonly used feedforward neural networks due to their fast operation, ease of implementation, and smaller training set requirements. The MLPNN consists of three sequential layers: input, hidden and output layers as shown in Figure 4.

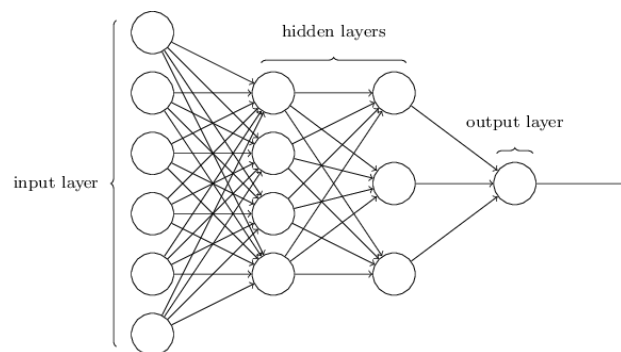


Fig.4. Multilayer Artificial Neural Network

The hidden layer processes and transmits the input information to the output layer. [36]. The training algorithm is an important part of the ANNs model. A good topology can be inefficient if trained by an inappropriate algorithm. A suitable training algorithm has a short training process while achieving better accuracy. There are many training algorithms used to train MLPNN and one of the most commonly used is Bayesian regularization BP, which is also used in this work. This algorithm updates the weight and bias values according to Levenberg-Marquardt optimization. It minimizes a combination of squared errors and weights, and then determines the correct combination to produce a network that generalizes well. The process is called Bayesian regularization [37].

2.5. Calculation of classification performance

The most straightforward way to evaluate the performance of classifiers is based on the confusion matrix analysis [38]. A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of

such a system is commonly evaluated using the data in the matrix [39]. Given a classifier and an instance, there are four possible outcomes. If the instance is positive and it is classified as positive, it is counted as a true positive (TP); if it is classified as negative, it is counted as a false negative (FN). If the instance is negative and it is classified as negative, it is counted as a true negative (TN); if it is classified as positive, it is counted as a false positive (FP) [40]. The evaluation of the proposed methods in classification problems is determined by computing the statistical parameters. Sensitivity (SEN), specificity (SPE) and classification accuracy (CA) values are calculated in Equals 1, 2 and 3:

$$SEN = \frac{TP}{TP + FN} \% \quad (1)$$

$$SPE = \frac{TN + FP}{TN + FP + TP + TN} \% \quad (2)$$

$$CA = \frac{TP + TN}{TP + TN + FP + FN} \% \quad (3)$$

3. Proposed Methods

SI is the name given to a relatively new interdisciplinary field of research, which has gained a wide popularity in recent times. Algorithms belonging to this field, draw inspiration from the collective intelligence emerging from the behavior of a group of social insects (like bees, termites and wasps). These insects even with very limited individual capability can jointly (cooperatively) perform many complex tasks necessary for their survival [41]. Ant Miner, ABC and PSO that are SI methods are frequently used in classification problems and are obtained successful results [19].

3.1. Artificial Bee Colony

The colony of artificial bees consists of three groups of bees: employed, onlookers, and scout bees. The employed bees are those, which randomly search for food-source positions. Onlookers are those bees waiting in the hive's dance area. In the ABC algorithm, onlookers and employed bees perform the exploration process in the search space, while on the other hand, scouts control the exploration process [42].

In ABC every bee explores a possible solution – in the current case the optimum weight matrix for the given network configuration. The training error $err(x)$ is used as the fitness values; this indicates the extent of conformance of the network output with actual output. Minimizing the error (fitness value) will lead to the best set of weights for the given network configuration and the network is said to be trained. For any classifier, its performance is dependent on the chosen loss function. Selection of proper loss function $err(x)$ for a given problem is often difficult. Different classification techniques in machine learning employ different loss functions to get better classification accuracy. Commonly neural network classifiers employ means square error (mse) minimization or cross entropy. In our study, we use the loss function such as root mean sum of squared residuals (error) in the training data as the fitness values of the ABC. This serves as a qualitative performance measure of the network learning and is given in Eq. 4 [19].

$$mse = \frac{1}{2} \sum_i (y_i - d_i)^2 \quad (4)$$

where y_i is the activity level of the i^{th} node in the top layer and d_i is the desired output of the i^{th} node.

Our objective is to minimize this fitness value. At each time step the randomness amplitude and speed of convergence of each bee are changed towards its food source. The random factor prevents the swarm getting stuck in the wrong place and speed of convergence is used to identify the rate at which bees converge to a solution. Training basically involves presenting the training samples as input vectors through a neural network, calculating the error of the output layer, and then adjusting the weights of the network to minimize the error [19].

3.2. Particle Swarm Optimization

PSO that is an evolutionary optimization algorithm suggested by Kennedy and Eberhart in the mid-1990s and it attempts to simulate the movements and choreographies of the birds. PSO can be incorporated into the classification methods because it is robust and adaptable. On the other hand, modified PSO with w that is the inertia weight tries to keep away from the local minima [43]. Inertia weight is an important parameter in PSO, which significantly affects the convergence and exploration-exploitation trade-off in PSO process [44]. Big and small values of inertia variable result in the avoidance of local minima [45].

$$V_i^{t+1} = wV_i^t + c_1 \text{Rand}() (P_i - X_i^t) + c_2 \text{Rand}() (P_g - X_i^t) \quad (5)$$

$$W_{ij} = X_i^{t+1} = X_i^t + V_i^{t+1} \quad (6)$$

$$X_j = \sum_i y_i w_{ij} \quad (7)$$

Firstly, the total weighted input x_j is calculated using Eqs. 5, 6, and 7 for the PSO-MLPNN method. X indicate the particle position and V indicate the particle velocity. P_i and P_g are the particle best ($pbest$) and global best ($gbest$), respectively. The term i shows the particle index, and t is the time step. $\text{Rand}()$ denotes a normally distributed one-dimensional random number with a mean of zero and a standard deviation of one. Parameters c_1 and c_2 are the cognitive and social learning rates [46], where w is the inertia suggested by [43]. In Eq. 7, y_i is the activity level of the i^{th} node in the previous layer and w_{ij} is the weight of the connection between the i^{th} and j^{th} node.

Secondly, the node computes the activity y_j using some function of the total weighted input. In this work, we used the sigmoid function since the error rate is less than the others.

$$y_j = \left[1 + e^{-x_j} \right]^{-1} \quad (8)$$

The error (E) when the activities of all output nodes have been determined is computed, which is defined by the expression:

$$E = \frac{1}{2} \sum_i (y_i - d_i)^2 \quad (9)$$

where y_i is the activity level of the i^{th} node in the top layer and d_i is the desired output of the i^{th} node.

In our study, the MLPNN and SI methods were combined to prevent the MLPNN from getting out of the trap and to ensure high classification success. For proposed methods ABC-MLPNN and PSO-MLPNN, after randomly assigning initial weights, activation function values were calculated, and training of the network was continued using the BP with new weights until the

stop criteria were achieved. Input values and hidden layer neuron numbers of the network (consisting of one hidden layer) determine the dimension (N) of the ABC and PSO. The block diagram of the MLPNN method combined using SI is shown in Figure 5. The weights obtained by using SI between

input-hidden layer and hidden output layer are represented by W_{SI} . A sigmoid function was used as the activation function of the input-hidden layer and a linear function was used as the activation function of the hidden output layer.

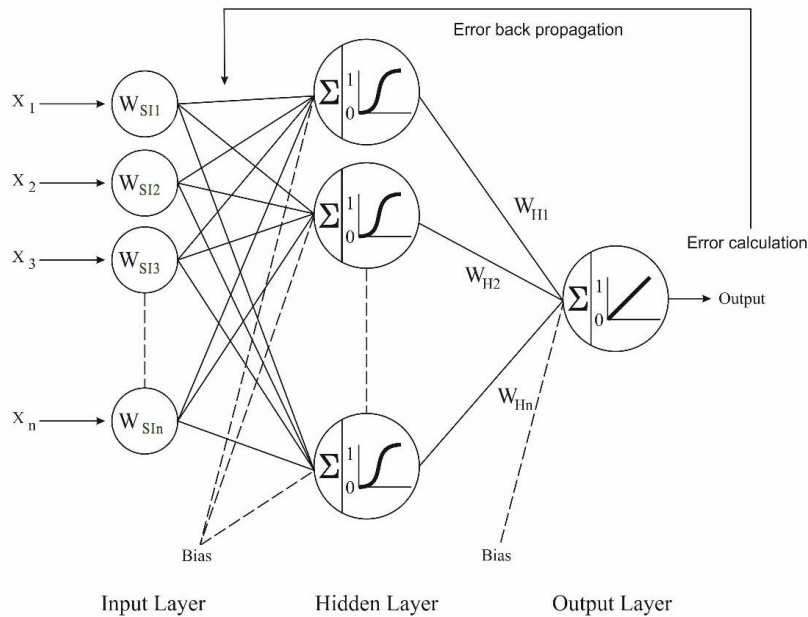


Fig.5. This is a flow diagram of the active nodes used in the hidden and output layers of the SI-MLPNN. Each of the inputs is multiplied by a weight (W_{SI} : the ABC or PSO values), and the summed.

4. Experimental Results

The success of combined proposed methods was compared with the performance of a conventional-MLPNN method by using two different EEG datasets, of which one were publicly available (Boston Children’s Hospital) and the remaining dataset (Selcuk University Hospital) was obtained retrospectively. For best model selection in the classification problems described, a ten-fold, cross-validation technique was used [47]–[49]—where ten-fold, cross-validation is the most common in data mining and machine learning. Afterwards, all of the datasets values underwent a normalization operation before being trained by pre-processing as follows:

$$x_s^{norm} = \frac{x_s - x_{min}}{x_{max} - x_{min}} \quad (10)$$

where x_s is the value of the s^{th} segment to be normalized, and x_{max} and x_{min} are the maximum and minimum values of the data.

A ten-fold cross-validation technique was used for the best model selection in the defined classification problems [47]–[49], with a ten-fold cross-validation technique being the most common in data mining and machine learning. Meanwhile, the target values are kept as 1 and 0, which in turn, represent epileptic activities and non-epilepsy EEGs respectively.

The present model in this study consists of three following steps as mentioned in section of Materials and Methods:

1. **Feature Extraction:** The EEGs, consisting of many data points, can be compressed into a few features by performing spectral analysis of the signals with the WT. These features characterize the behavior of the EEGs.

Using a smaller number of features to represent the EEGs is particularly important for recognition and diagnostic purposes [25], [26], [50]. DWT is one of the nonstationary methods used for extracting features. The determination of convenient wavelets and the number of levels of decomposition is very significant for signal analysis. After arrangement of datasets, feature extraction must be performed so that the same dataset can be identified with fewer features. The obtained feature vector can sometimes be used directly or a feature selection process can be carried out to decrease the number of features. Firstly, the fourth-order Daubechies, which has proven to be the most appropriate wavelet function for epileptic EEG analysis wavelength, has chosen for wavelet function in this study. [51]. The number of decomposition levels must chose according to on the effective frequency components of the signal [52]. The fourth order Daubechies wavelet that is wavelet function was used to decompose for sub-band in the time domain using fifth level. EEG recordings were divided into sub-band frequencies such as delta (δ), theta (θ), alpha (α) and beta (β) by using DWT. Then a set of power features in time-domain was extracted from the wavelet sub-band frequencies δ (0–4 Hz), θ (4–8 Hz), α (8–16 Hz) and β (16–32 Hz). Statistical features are indicative factors in determining qualities of signs. Thus, a second feature vector was obtained by subtracting 10 different statistical features from all channels on each dataset. Then, DWT and statistical properties were combined and unique feature vector was constructed. All feature vectors were computed using the MATLAB (Version 7.11, R2010b)

software package to achieve results faster and more accurately.

2. *Feature Selection:* Selection of the proposed model inputs is very important for success of the classifier [21]. In that context, if the number of inputs is selected unnecessarily high, performance of the system might decrease because of the difficulty of the calculation on the network. However, if the number of inputs is selected unnecessarily low, the system may not give the result accurately and positively. Therefore, selection of the number of inputs is very significant for these systems. There is no doubt that, PCA

that is a sophisticated method reduces the size of features [53]. PCA method was operated for feature selection. New feature vectors (eigenspace) which are given in the “Feature Selection” section of Table 1 was determined by using PCA. By considering the relationship between features through feature selection, those features with the most significant relationship create a new eigenvector. The eigenvectors present in this table were used to determine the success of classification methods as a feature vector.

Table 1. The number of properties and feature vectors of Boston and Selcuk dataset (322 and 252 features were obtained from each dataset by feature extraction techniques. However, the 13 and 24 feature values obtained after PCA application have been used to test the performance of classifiers instead of these values).

Dataset	The number of records	Channel number	The number of features (Feature Extraction)	The number of eigenvector (Feature Selection)	New Feature vector
Boston	200	23	322	13	13 x 200
Selcuk	60	18	252	24	24 x 60

3. *Classification:* Signal processing is to determine the success of the system using classifiers. The BP algorithm has difficulties in handling local optima and cannot yield optimal adjustable weights for MLPNN. For the MLPNN method, which has a single hidden layer, the number of different numbers of neurons has been tested and twelve neurons were chosen as optimum neuron numbers after attempts [19]. The best weights and optimum number of neurons have been investigated operating a program written in Delphi language. The main difference between the network types lies in the type of activation function used by

the hidden neurons. In MLPNN, a common type of activation function used by the hidden neurons has a sigmoid function. Neurons in the output layer usually have linear transfer functions [54]. Classification success was also investigated by using a sigmoid transfer function as shown in Figure 6. Thus, high classification success was aimed for by choosing the best network structure. Linear and sigmoid functions used as the output function were tried for some folds of the present datasets, and their performances were examined.

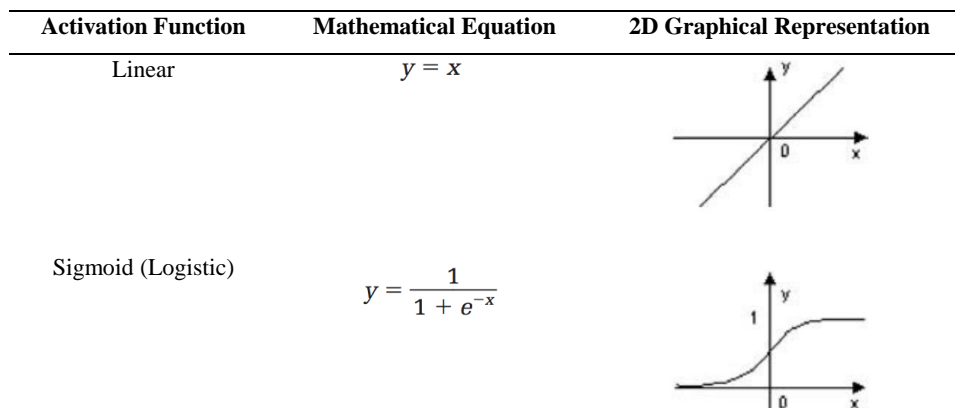


Fig.6. The activation functions used in proposed methods [55].

According to the data obtained, there was no remarkable difference between them; however, the results obtained with the linear transfer function were slightly better than the sigmoid function. So, the linear transfer function has been performed for all other data. The simplest stop criterion for active learning is when the training set reaches the desired size or a predefined threshold. In this context, we examined each stop criterion methods for ending classification. In this work, the termination

criterion is mostly obtained with maximum iteration for all classification methods.

Afterwards, we separately calculated SEN, SPE and CA values (Equals 7, 8 and 9), which are statistical measures of the performance of a binary classification test for all methods. Classification successes for each method are shown in Tables 2, 3 and 4, respectively.

Table 2. The Results of ABC-MLPNN for every fold

Fold Number	Boston Dataset				Selcuk Dataset			
	Correct Data	SEN (%)	SPE (%)	CA (%)	Correct Data	SEN (%)	SPE (%)	CA (%)
Fold 1	18	83.33%	92.86%	90.00%	6	100.00%	100.00%	100.00%
Fold 2	14	70.00%	70.00%	70.00%	5	100.00%	66.67%	83.33%
Fold 3	16	71.43%	100.00%	80.00%	6	100.00%	100.00%	100.00%
Fold 4	16	77.78%	81.82%	80.00%	6	100.00%	100.00%	100.00%
Fold 5	16	92.31%	57.14%	80.00%	5	100.00%	50.00%	83.33%
Fold 6	17	66.67%	92.86%	85.00%	5	66.67%	100.00%	83.33%
Fold 7	17	84.62%	85.71%	85.00%	5	75.00%	100.00%	83.33%
Fold 8	14	85.71%	61.54%	70.00%	6	100.00%	100.00%	100.00%
Fold 9	14	66.67%	75.00%	70.00%	5	100.00%	75.00%	83.33%
Fold 10	16	90.00%	70.00%	80.00%	6	100.00%	100.00%	100.00%

Table 3. The Results of PSO-MLPNN for every fold

	Boston Dataset				Selcuk Dataset			
	Correct Data	SEN (%)	SPE (%)	CA (%)	Correct Data	SEN (%)	SPE (%)	CA (%)
Fold 1	15	57.14%	84.62%	75.00%	6	100.00%	100.00%	100.00%
Fold 2	13	80.00%	60.00%	65.00%	6	100.00%	100.00%	100.00%
Fold 3	16	71.43%	100.00%	80.00%	4	50.00%	100.00%	66.67%
Fold 4	16	72.73%	88.89%	80.00%	5	75.00%	100.00%	83.33%
Fold 5	14	90.91%	44.44%	70.00%	4	75.00%	50.00%	66.67%
Fold 6	16	57.14%	92.31%	80.00%	6	100.00%	100.00%	100.00%
Fold 7	15	73.33%	80.00%	75.00%	6	100.00%	100.00%	100.00%
Fold 8	16	76.92%	85.71%	80.00%	5	100.00%	75.00%	83.33%
Fold 9	16	75.00%	87.50%	80.00%	5	75.00%	100.00%	83.33%
Fold 10	14	75.00%	62.50%	70.00%	6	100.00%	100.00%	100.00%

Table 4. The Results of MLPNN for every fold

	Boston Dataset				Selcuk Dataset			
	Correct Data	SEN (%)	SPE (%)	CA (%)	Correct Data	SEN (%)	SPE (%)	CA (%)
Fold 1	12	37.50%	75.00%	60.00%	5	66.67%	100.00%	83.33%
Fold 2	13	80.00%	60.00%	65.00%	6	100.00%	100.00%	100.00%
Fold 3	11	55.56%	54.55%	55.00%	5	66.67%	100.00%	83.33%
Fold 4	14	63.64%	77.78%	70.00%	5	75.00%	100.00%	83.33%
Fold 5	16	50.00%	85.71%	80.00%	4	100.00%	33.33%	66.67%
Fold 6	15	92.31%	57.14%	75.00%	5	66.67%	100.00%	83.33%
Fold 7	15	81.82%	66.67%	75.00%	6	100.00%	100.00%	100.00%
Fold 8	13	70.00%	60.00%	65.00%	5	100.00%	75.00%	83.33%
Fold 9	14	62.50%	100.00%	70.00%	3	100.00%	100.00%	50.00%
Fold 10	14	100.00%	57.14%	70.00%	5	75.00%	100.00%	83.33%

Table 5. Confusion matrix using ABC-MLPNN with ten-fold cross-validation (90-10% training-test for Fold 1 of Boston dataset).

Fold 1	Predicted Class		Total
	0	1	
Actual Class	0	5 (TP) 1 (FP)	6
	1	1 (FN) 13 (TN)	14
Total	6	14	20

The confusion matrix of Fold 1 for Boston dataset that is performed via ABC-MPLNN is given in Table 5 as an example. While 5 correct data have been obtained for 6 data of EEG that includes epileptic activities, 13 correct data have been achieved for 14 data of non-epilepsy EEG. Thus, when TP (5) and TN (13) values are collected, it is seen that 18 out of 20 test data are

correctly estimated. When classification performance is calculated for fold 1 using statistical parameters,

$$SEN = \frac{5}{6} = 83.33\%$$

$$SPE = \frac{13}{14} = 92.86\%$$

$$CA = \frac{18}{20} = 90.00\%$$

values are obtained.

Average sensitivity, specificity and CA values of the proposed methods, and conventional-MLPNN method, on all datasets are given in Table 6.

Table 6. Average classification success (\pm standard deviation) of all methods for every dataset

Method	Dataset	Sensitivity (% \pm Std.dev.)	Specificity (% \pm Std.dev.)	CA (% \pm Std.dev.)
ABC-MLPNN	Boston	78.85 \pm 0.0964	78.69 \pm 0.1429	79.00 \pm 0.0699
	Selcuk	94.17 \pm 0.1245	89.17 \pm 0.1845	91.67 \pm 0.0878
PSO-MLPNN	Boston	72.96 \pm 0.0612	78.60 \pm 0.1729	75.50 \pm 0.0550
	Selcuk	87.50 \pm 0.1768	92.50 \pm 0.1687	88.33 \pm 0.1372
MLPNN	Boston	69.33 \pm 0.1942	69.40 \pm 0.1494	68.50 \pm 0.0747
	Selcuk	85.00 \pm 0.1610	90.83 \pm 0.2168	81.67 \pm 0.1459

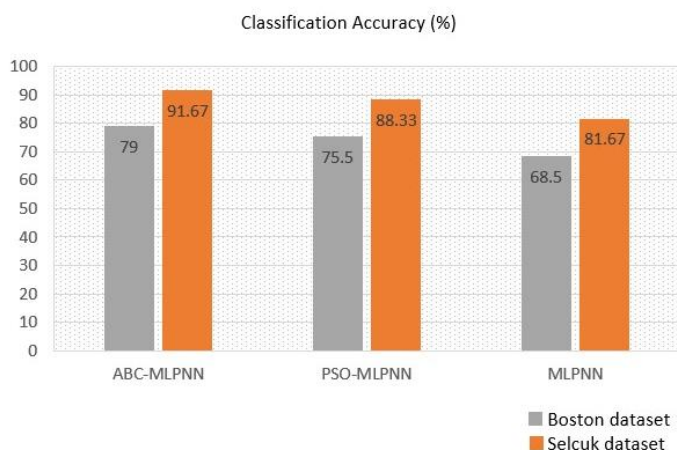


Fig.7. Comparison of traditional-MLPNN and proposed classification methods.

As seen in Table 6 and Figure 7, the ABC-MLPNN method shows the highest success rate (91.67%) with the Selcuk dataset while the PSO-MLPNN and MLPNN methods show a success rate of 88.33% and 81.67%, respectively, on the same dataset. Consequently, the CA of ABC-MLPNN and PSO-MLPNN methods is higher than the CA of MLPNN method for both Selcuk dataset and Boston dataset. Therefore, it can be clearly seen that the successes of the proposed methods are higher than the success of a conventional-MLPNN method. As a result, it was determined that the success of ANN combined with SI techniques, such as ABC and PSO, was higher than that of the conventional-ANN method. The performances of the proposed methods on different datasets were investigated, and it was seen that the success of SI-MLPNN methods were higher than that of the conventional-MLPNN method.

5. Conclusions and Discussion

Automated detection of EEGs in normal and ictal situations is important in the field of epileptic activities. ANNs have been widely used by researchers to classify the EEG signals [56]. There is no doubt that the ANN algorithm is one of the most commonly used algorithms to test the accuracy and success of a system's operation. However, BP usually traps in local minima. This greatly affects the success and efficiency of the system. SI has been extensively used for training neural network because of the stochastic nature of the algorithm, which makes it very robust and flexible. Therefore, we investigated the combined ABC-MLPNN and PSO-MLPNN methods for avoid this problem and tested the successes of proposed methods and conventional-MLPNN method on two different EEG datasets.

As mentioned in the experimental results section, both the satisfactory accuracy of the proposed methods by performing

ABC-MLPNN and PSO-MLPNN but also a suitable eigenvector feature set was found. We suggest that the proposed methods contribute to the detection of signals indicating brain disorders and that long-term EEG recording that are difficult to interpret can be used to diagnose epileptic activities using techniques such as embedded systems.

The proposed methods can provide valuable contributions to the neurologist in treatment and diagnosis. In addition, these methods, to be further developed into a user-level program, allow fast and robust classification of EEGs.

5.1. Conflicts of Interest Statement

The authors declare that they have no competing interests.

5.2. Acknowledgements

We would like to thank to the doctors who shared their knowledge and experience and to the Department of Neurology of Selcuk University Hospital (Non-Invasive Clinical Research Ethics Committee No.2014/423) that provided the dataset.

References

- [1] W. G. Baxt, "Use of an artificial neural network for data analysis in clinical decision-making: the diagnosis of acute coronary occlusion," *Neural Comput.*, vol. 2, no. 4, pp. 480–489, 1990.
- [2] I. Gule, E. D. Ubeyli, and N. F. Guler, "A mixture of experts network structure for EEG signals classification," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2006, pp. 2707–2710.
- [3] E. D. Übeyli, "Lyapunov exponents/probabilistic neural networks for analysis of EEG signals," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 985–992, 2010.
- [4] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, 2014.
- [5] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *J. Mach. Learn. Res.*, vol. 3, no. 3, pp. 1157–1182, 2003.
- [6] "Classification." [Online]. Available: https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/classify.htm#i1005746. [Accessed: 12-Feb-2019].
- [7] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of neural science*, vol. 4. McGraw-Hill New York, 2000.
- [8] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *J. Neurosci. Methods*, vol. 123, no. 1, pp. 69–87, 2003.
- [9] L. D. Iasemidis *et al.*, "Adaptive epileptic seizure prediction system," *Biomed. Eng. IEEE Trans.*, vol. 50, no. 5, pp. 616–627, 2003.
- [10] W. O. Tatum IV, "Long-term EEG monitoring: a clinical approach to electrophysiology," *J. Clin. Neurophysiol.*, vol. 18, no. 5, pp. 442–455, 2001.
- [11] L. Guo, D. Rivero, and A. Pazos, "Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial

- neural networks,” *J. Neurosci. Methods*, vol. 193, no. 1, pp. 156–163, 2010.
- [12] K. Fu, J. Qu, Y. Chai, and Y. Dong, “Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM,” *Biomed. Signal Process. Control*, vol. 13, pp. 15–22, Sep. 2014.
- [13] V. Joshi, R. B. Pachori, and A. Vijesh, “Classification of ictal and seizure-free EEG signals using fractional linear prediction,” *Biomed. Signal Process. Control*, vol. 9, pp. 1–5, Jan. 2014.
- [14] T. Gandhi, B. K. Panigrahi, and S. Anand, “A comparative study of wavelet families for EEG signal classification,” *Neurocomputing*, vol. 74, no. 17, pp. 3051–3057, Oct. 2011.
- [15] S.-H. Lee, J. S. Lim, J.-K. Kim, J. Yang, and Y. Lee, “Classification of normal and epileptic seizure EEG signals using wavelet transform, phase-space reconstruction, and Euclidean distance,” *Comput. Methods Programs Biomed.*, vol. 116, no. 1, pp. 10–25, 2014.
- [16] S. Aydın, H. M. Saraoğlu, and S. Kara, “Log energy entropy-based EEG classification with multilayer neural networks in seizure,” *Ann. Biomed. Eng.*, vol. 37, no. 12, pp. 2626–2630, 2009.
- [17] K. A. Abuhasel, A. M. Iliyasa, and C. Fatichah, “A Hybrid Particle Swarm Optimization and Neural Network with Fuzzy Membership Function Technique for Epileptic Seizure Classification,” *J. Adv. Comput. Intell. Intell. Informatics*, vol. 19, no. 3, pp. 447–455, 2015.
- [18] S. K. Satapathy, S. Dehuri, and A. K. Jagadev, “ABC optimized RBF network for classification of EEG signal for epileptic seizure identification,” *Egypt. Informatics J.*, 2016.
- [19] S. Dehuri, S. Ghosh, and S.-B. Cho, *Integration of swarm intelligence and artificial neural network*. World Scientific, 2011.
- [20] A. H. Shoeb, “Application of machine learning to epileptic seizure onset detection and treatment.” Massachusetts Institute of Technology, 2009.
- [21] I. Güler and E. D. Übeyli, “Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients,” *J. Neurosci. Methods*, vol. 148, no. 2, pp. 113–121, 2005.
- [22] P. S. Addison, *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. CRC press, 2017.
- [23] I. Daubechies, “The wavelet transform, time-frequency localization and signal analysis,” *IEEE Trans. Inf. theory*, vol. 36, no. 5, pp. 961–1005, 1990.
- [24] S. Soltani, “On the use of the wavelet decomposition for time series prediction,” *Neurocomputing*, vol. 48, no. 1, pp. 267–277, 2002.
- [25] M. Akay, “Wavelet applications in medicine,” *IEEE Spectr.*, vol. 34, no. 5, pp. 50–56, 1997.
- [26] M. Unser and A. Aldroubi, “A review of wavelets in biomedical applications,” *Proc. IEEE*, vol. 84, no. 4, pp. 626–638, 1996.
- [27] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi, “Wavelet toolbox,” *MathWorks Inc., Natick, MA*, vol. 15, p. 21, 1996.
- [28] K. N. Le, K. P. Dabke, and G. K. Egan, “Hyperbolic wavelet power spectra of nonstationary signals,” *Opt. Eng.*, vol. 42, no. 10, pp. 3017–3037, 2003.
- [29] S. Prabhakar, A. R. Mohanty, and A. S. Sekhar, “Application of discrete wavelet transform for detection of ball bearing race faults,” *Tribol. Int.*, vol. 35, no. 12, pp. 793–800, 2002.
- [30] A. Moreno-Muñoz, *Power quality: mitigation technologies in a distributed environment*. Springer Science & Business Media, 2007.
- [31] Y. Li, P. P. Wen, Siuly, Y. Li, and P. P. Wen, “Clustering technique-based least square support vector machine for EEG signal classification,” *Comput. Methods Programs Biomed.*, vol. 104, no. 3, pp. 358–372, Dec. 2011.
- [32] S. Lakhina, S. Joseph, and B. Verma, “Feature reduction using principal component analysis for effective anomaly-based intrusion detection on NSL-KDD,” 2010.
- [33] U. Rajendra Acharya *et al.*, “Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework,” *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9072–9078, Aug. 2012.
- [34] “What is the differences between artificial neural network (computer science) and biological neural network? - Quora.” [Online]. Available: <https://www.quora.com/What-is-the-differences-between-artificial-neural-network-computer-science-and-biological-neural-network>. [Accessed: 14-Feb-2019].
- [35] I. A. Basheer and M. Hajmeer, “Artificial neural networks: fundamentals, computing, design, and application,” *J. Microbiol. Methods*, vol. 43, no. 1, pp. 3–31, 2000.
- [36] U. Orhan, M. Hekim, and M. Ozer, “EEG signals classification using the K-means clustering and a multilayer perceptron neural network model,” *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13475–13481, 2011.
- [37] L. Guo, D. Rivero, J. Dorado, J. R. Rabunal, and A. Pazos, “Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks,” *J. Neurosci. Methods*, vol. 191, no. 1, pp. 101–109, 2010.
- [38] G. E. Batista, R. C. Prati, and M. C. Monard, “A study of the behavior of several methods for balancing machine learning training data,” *ACM Sigkdd Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, 2004.
- [39] T. R. Patil and S. S. Sherekar, “Performance analysis of Naive Bayes and J48 classification algorithm for data classification,” *Int. J. Comput. Sci. Appl.*, vol. 6, no. 2, pp. 256–261, 2013.
- [40] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
- [41] S. Das and A. Konar, “A swarm intelligence approach to the synthesis of two-dimensional IIR filters,” *Eng. Appl. Artif. Intell.*, vol. 20, no. 8, pp. 1086–1096, 2007.
- [42] F. S. Abu-Mouti and M. E. El-Hawary, “Optimal distributed generation allocation and sizing in distribution systems via artificial bee colony algorithm,” *IEEE Trans. power Deliv.*, vol. 26, no. 4, pp. 2090–2101, 2011.
- [43] Y. Shi and R. Eberhart, “A modified particle swarm optimizer,” in *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on*, 1998, pp. 69–73.
- [44] J. C. Bansal, P. K. Singh, M. Saraswat, A. Verma, S. S. Jadon, and A. Abraham, “Inertia weight strategies in particle swarm optimization,” in *Nature and Biologically Inspired Computing (NaBIC), 2011 Third World Congress on*, 2011, pp. 633–640.
- [45] P. C. Chen and Y. K. Hwang, “SANDROS: a dynamic graph search algorithm for motion planning,” *Robot. Autom. IEEE Trans.*, vol. 14, no. 3, pp. 390–403, 1998.
- [46] R. C. Eberhart and J. Kennedy, “A new optimizer using particle swarm theory,” in *Proceedings of the sixth international symposium on micro machine and human science*, 1995, vol. 1, pp. 39–43.
- [47] U. R. Acharya, S. V. Sree, A. P. C. Alvin, and J. S. Suri, “Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework,” *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9072–9078, 2012.
- [48] Y. Kaya, M. Uyar, R. Tekin, and S. Yıldırım, “1D-local binary pattern based feature extraction for classification of epileptic EEG signals,” *Appl. Math. Comput.*, vol. 243, pp. 209–219, 2014.
- [49] D. Wang, D. Miao, and C. Xie, “Best basis-based wavelet packet

- entropy feature extraction and hierarchical EEG classification for epileptic detection,” *Expert Syst. Appl.*, vol. 38, no. 11, pp. 14314–14320, 2011.
- [50] N. Hazarika, J. Z. Chen, A. C. Tsoi, and A. Sergejew, “Classification of EEG signals using the wavelet transform,” in *Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International Conference on*, 1997, vol. 1, pp. 89–92.
- [51] A. Subasi, “EEG signal classification using wavelet feature extraction and a mixture of expert model,” *Expert Syst. Appl.*, vol. 32, no. 4, pp. 1084–1093, May 2007.
- [52] E. D. Übeyli, “Statistics over features: EEG signals analysis,” *Comput. Biol. Med.*, vol. 39, no. 8, pp. 733–741, 2009.
- [53] P. Berg and M. Scherg, “A multiple source approach to the correction of eye artifacts,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 90, no. 3, pp. 229–241, 1994.
- [54] M. Y. Rafiq, G. Bugmann, and D. J. Easterbrook, “Neural network design for engineering applications,” *Comput. Struct.*, vol. 79, no. 17, pp. 1541–1552, 2001.
- [55] “Mind: How to Build a Neural Network (Part One).” [Online]. Available: <https://stevenmiller888.github.io/mind-how-to-build-a-neural-network/>. [Accessed: 14-Feb-2019].
- [56] Q. Gui, Z. Jin, and W. Xu, “Exploring EEG-based biometrics for user identification and authentication,” *2014 IEEE Signal Process. Med. Biol. Symp. IEEE SPMB 2014 - Proc.*, 2015.