© 2016 IJSRSET | Volume 2 | Issue 4 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

An intelligent System for Diagnosing Schizophrenia and Bipolar Disorder based on MLNN and RBF

M. I. Elgohary[1], Tamer. A. Alzohairy[2], Amir. M. Eissa[1], sally. Eldeghaidy[3], Hussein. M[*1]
*[1]Department of physics, Al-azhar university, Cairo, Egypt
huscholes@gmail.com
[2]Department of computer science, Al-azhar University, Cairo, Egypt
3Department of physics, Suiz Canal University, Suiz canal, Egypt

ABSTRACT

This paper is concerned with the problem of discriminating the patients suffering from schizophrenia and bipolar disorder versus control based on EEG rhythms. The EEG rhythms are used to extract a feature vector for each patient. In this paper, the large set of features included in the EEG rhythms is reduced into smaller set of features using FFT segmentation. Two classes of classifiers which are multi-layer perceptron and radial basis function are used to discriminate the patient data based on features vector. Experimental studies have shown that the proposed algorithms give excellent results when applied and tested on the three classes. The multilayer neural network with backpropagation achieved a high performance rate equal to 98.67 % compared to radial basis function networks which achieved a performance rate equal to 87.33%.

Keywords: EEG, schizophrenia, Bipolar disorder, artificial neural networks, Backpropagation Algorithm, Radial basis function network.

I. INTRODUCTION

An essential criterion for successful treatment of a mental illness is a correct diagnosis. The clinical diagnostic guidelines are well established using the American Psychiatric Association to differentiate various psychiatric conditions [1]. However, the diagnostic process is a more difficult task than it may first appear because specific symptoms can appear in more than one diagnostic category, and diagnostic criteria can overlap to the point where confident differentiation is often impossible. Also the psychiatric expert can have difficulty distinguishing certain psychiatric conditions, e.g. psychotic depression from schizophrenia or differentiating major depressive disorder from bipolar depression (BD) [2].

Diagnosis of psychiatric conditions may prove problematic due to the overlapping nature of their symptoms. Offering antidepressant medication to a major depressive disorder (MDD) patient can be quite helpful while the same medication might induce mania in a BD patient. It is very important therefore that a correct psychiatric diagnosis be obtained before treatment is initiated. [3]

There are previous studies based on the EEG for diagnosis. For example, the study [4] based on a sample of depression and normal control subjects reported that frontal brain asymmetry is a potential marker for depression. In [5], EEG data is analyzed to compare normal subjects versus subjects suffering various mental disorders. Li and Fan [6] use an artificial neural network (ANN) fed with EEG data to differentiate schizophrenia and depression patients. [7] M. Liu employed Principal Components Analysis (PCA) with nonlinear SVM, [8] Alba-Sanchez, F. utilized Self-Organizing Maps (SOM) in pattern recognition of mental disorder disease. Hiesh, M. use Support Vector Machine (SVM) to analyze Gamma band synchronization test data [9].

In this paper we propose an automated machine learning procedure that can diagnose specific forms of psychiatric illness based on the patient's resting electroencephalogram (EEG). The classes of psychiatric illness classified are schizophrenia (SCZ), and bipolar disorder (BD). A comparison with sample of healthy or 'normal' (N) subjects is given. The proposed methodology could in principal be extended to other forms of illness.

Classification of psychiatric disorders using artificial neural networks is a major focus in this paper. The use of two classifier Backpropagation and Radial basis function neural networks for distinguishing between the two classes of diagnostic illness: SCZ, BD against N is demonstrated.

This paper is structured as follows. In Section II, subjects and methods. Neural networks as classifiers are presented in section III. In Section V, the results and discussion are given. Finally, some conclusions are remarked in section VI.

II. METHODS AND MATERIAL

A. Subject

The EEG data were obtained from DR. ABOU ELAZAYEM PSYCHIATRIC HOSPITAL in Egypt [10]. The subjects included 70 normal persons who have no history of neurological or psychiatric disease, 80 schizophrenic and 80 bipolar patients. All patients were hospitalized and diagnosed with schizophrenia or bipolar according to the criteria of Diagnostic and Statistical Manual of Mental Disorders by independent psychiatrists.

In an acoustically and electrically shielded room where the subjects were seated comfortably in a reclining chair, the EEG data were obtained from 16 surface electrodes placed on the scalp according to the standard international 10/20 system, namely the 16 channels, Fp1, Fp2, F3, F4, F7, F8, C3, C4, P3, P4, T3, T4, T5, T6, O1, O2 with reference to linked earlobes. The digitization of 16 channels EEG was performed with a sampling rate of 100Hz using a 12 bit AD-converter and the data were recorded on a hard disk. For each subject, recordings covered the EEG activity of a resting condition for time ranging from 3 to 5 minutes approximately.

B. Feature Extraction

Feature selection is a very important step in pattern recognition. The idea of feature selection is to choose a subset of features that improve the performance of the classifier especially when we are dealing with high dimension data. Finding significant features that produce higher classification accuracy is an important problem.

The main steps for vector feature extraction of each subject are using the EEGLAB toolbox [11] to filter the original spontaneous EEG time series and remove the artifacts. A size of clear 2000 time-samples is selected which represents a 20 seconds time interval from each subjects. Fast Fourier Transform (FFT) for 2000 points of all-time series obtained from 16 channels were calculated. The 2000 points of each channel are partitioned into 8 intervals and the mean value of each interval is computed which results in 8 points. So, for each subject, there is 8 points vector for each one of the 16 channels which results in a single 128 feature vector that will be used as the ANN input vector for each subject.

III. Artificial Neural Networks

Artificial Neural Networks (ANNs) are the interconnection of simple processing nodes which functionality is modeled from the neuron in the brain. The ANN consists of an input layer, an output layer and at least one hidden layer. The input values to the network are fed from the input layer through the hidden layer to the output layer. The values obtained as inputs are processed within hidden layer and forwarded to either the nodes of the next hidden layer or nodes of the output layer. There are different types of NNs, in this paper we are interested which are feedforward neural network (FFNN) with backpropagation training algorithm and RBF network with k means clustering algorithm. They are suitable for the purpose of classifying conflict in the two diseases schizophrenia and bipolar.

A. Multilayer feed-forward neural network (MLFN) and its dynamics

The MLFF network architecture is shown in fig.1, in which the input layer is composed of n nodes and an additional node for input bias, output layer of m nodes for the m outputs of the network and 2 hidden layers.

Following is the summary of the BPN [13]:

Step 1: Initialize weights and offsets.

Set all weights and node offsets to small random values.

Step 2: Present input and desired outputs.

Present a continuous valued input vector $x_0, x_1, \ldots, x_{N-1}$

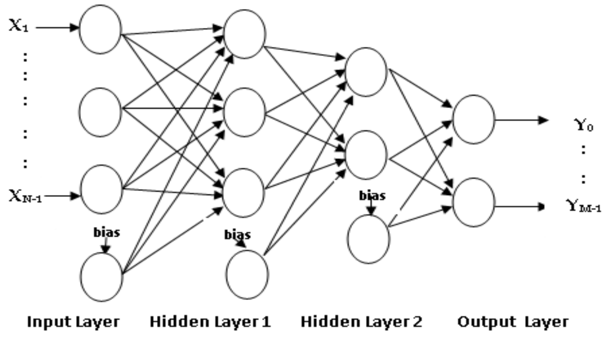

Fig. 1. MLFF Network architecture.

and specify the desired output $d_0, d_1, \ldots, d_{M-1}$. Because the network is used as a classifier then all desired outputs are typically set to zero except for that corresponding to the class the input is from. That desired output is +1. The input from the training set are presented cyclically until stabilization.

Step 3: Calculate actual output.

First calculate the formulas

$$I_j = f_j^{h1}\left[\sum_i w_{ji}^{h1} x_i + b_j^{h1}\right] \quad (1)$$

$$Z_j = f_r^{h2}\left[\sum_j w_{rj}^{h2} I_j + b_r^{h2}\right] \quad (2)$$

Where x_i is the ith network input, w_{ji}^{h1} is the connection weight from the ith input to the jth neuron in the 1st hidden layer, w_{rj}^{h2} is the connection weight from the jth neuron in the 1st hidden layer to the rth neuron in the 2nd hidden layer, b_j^{h1} is the weight from the bias to the jth neuron in the 1st hidden layer, b_r^{h2} is the weight from the bias to the rth neuron in the second hidden layer, $f_j^{h1}(.)$ and $f_r^{h2}(.)$ are nonlinear sigmoid activation functions defined as

$$f(netinput) = \frac{1}{1+e^{-netinput}} \quad (3)$$

The network output is calculated by the following equation

$$\hat{y}(k+1) = f_k^O\left[\sum_k w_{kr}^O Z_r + b_k^O\right] \quad (4)$$

where w_{kr}^O is the weight connection of the kth neuron in the output layer to the r^{th} neuron in the 2nd hidden layer, b_k^O is the bias weight for the k^{th} output neuron, and f_k^O is the transformation function between 2nd hidden layer and output layer which is a linear function.

Step 4: compute the error signal

$$e_j(k) = d_j(k) - \hat{y}_j(k) \quad (5)$$

Step 5: Adapt weights.

Use a recursive algorithm starting at the output nodes and working back to the first hidden layer. Adjust weights by

$$w_{ij}(k+1) = w_{ij}(k) + \eta\delta_j x_i \quad (6)$$

In this equation $w_{ji}(k)$` is the weight from hidden node i or from an input to node j at time k, x_i is either the output of node i or is an input, η is the learning rate, and δ_j is an error term for node J. If node J is an output node then

$$\delta_j = (d_j - \hat{y}_j) \quad (7)$$

where d_j is the desired output of node j and $\hat{y}_j$ is the actual output.

If node j is an internal hidden node then

$$\delta_j = x_j(1-x_j)\sum_k \delta_j w_{jk} \quad (8)$$

Where k is over all nodes in the layers connected to node j.

Internal node thresholds are adapted in a similar manner by assuming they are connection weights on links from auxiliary constant-valued inputs equal to 1.

Step 6: Repeat by going to step 2.

B. Radial basis function neural network (RBFN) and its dynamics

1) Radial Basis Function network architecture

Fig.2 shows a typical RBF network, with q inputs $(x_1, \ldots, x_q)$, and p outputs $(y_1, \ldots, y_p)$. The hidden layer consists of h computing units connected to the output by h weight vectors $(\underline{\alpha}_1, \ldots, \underline{\alpha}_h)$.

Response of one hidden unit to the network input at the i^{th} instant, $\underline{x}_i$ can be expressed by

$$\Phi_k(\underline{x}_i) = \exp\left[-\frac{1}{(\sigma_k^i)^2}\left\|\underline{x}_i - \underline{\mu}_k^i\right\|^2\right], \ (k:1\ldots h) \quad (9)$$

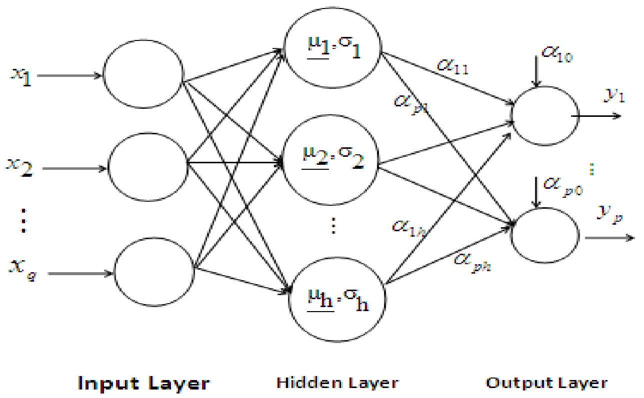

Fig. 2. RBF Network architecture.

where $\underline{\mu}_k^i$ is the center vector for k^{th} hidden unit at i^{th} instant, σ_k^i is the width of the Gaussian function at that time, and $\|.\|$ denotes the Euclidean norm. The overall network response is given by

$$\hat{\underline{y}}_i = \underline{f}(\underline{x}_i) = \underline{\alpha}_0^i + \sum_{k=1}^{h} \underline{\alpha}_k^i \Phi_k(\underline{x}_i) \qquad (10)$$

where $\underline{y}_i \in R^p$, $\underline{x}_i \in R^q$. The coefficient vector $\underline{\alpha}_k^i$ is the connecting weight vector of the k^{th} hidden unit to output layer, which is in the vector form of $\underline{\alpha}_k^i = [\alpha_{1k}^i, \ldots, \alpha_{lk}^i, \ldots, \alpha_{pk}^i]^T$. Thus the coefficient matrix of the network can be expressed as

$$A_{p\times h}^i = \begin{bmatrix} \alpha_{11}^i & \ldots & \alpha_{1k}^i & \ldots & \alpha_{1h}^i \\ \ldots & & \ldots & & \ldots \\ \alpha_{l1}^i & \ldots & \alpha_{lk}^i & \ldots & \alpha_{lh}^i \\ \ldots & & \ldots & & \ldots \\ \alpha_{p1}^i & \ldots & \alpha_{pk}^i & \ldots & \alpha_{ph}^i \end{bmatrix}$$

and the bias vector is $\alpha_0^i = [\alpha_{10}^i, \ldots, \alpha_{l0}^i, \ldots, \alpha_{p0}^i]^T$.

2) Training Radial Basis Function Network

Training RBF neural network consists of determining the location of centers and widths for the hidden layer and the weights of the output layer. It is trained using a two-phase approach: in the first phase, unsupervised learning occurs, which main objective is to optimize the location of center and width. In the second phase, the output layer is trained in a supervised mode using the least mean-square (LMS) algorithm to adjust the weight so as to obtain the minimum mean square error at the output. The following are the three steps of the hybrid learning method for an RBF neural network:

a) Find the cluster centers of the radial basis function; use the k-means clustering algorithm.

b) Find the width of the radial basis function using p-nearest neighbor.

c) Find the weight; use LMS.

a) Calculation of Centers

To calculate the centers of the radial basis function we use the k-means clustering algorithm. The purpose of applying the k-means clustering algorithm is to find a set of clustered centers and partition the training data into subclasses. The center of each cluster is initialized to a randomly chosen input datum. Then each training datum is assigned to the cluster that is nearest to itself. After training data have been assigned to a new cluster unit, the new center of a cluster represents the average of the training data associated with that cluster unit.

After all the new centers have been calculated, the process is repeated until it converges. The recursive k-means algorithm is given as follows:

1) Choose a set of centers $\{\underline{\mu}_1, \underline{\mu}_2, \ldots, \underline{\mu}_h\}$ arbitrarily and give the initial learning rate $\gamma(0)=1$.

2) Compute the minimum Euclidean distance

$$\begin{aligned} &L_i(k) = \left\|\underline{x}(k) - \underline{\mu}_i(k-1)\right\| \qquad \mathrm{i:1\ldots h} \\ &\mathrm{r = arg|min\ L_i}(k)| \end{aligned} \qquad (11)$$

3) Adjust the position of these centers as follows:

$$\begin{aligned} \underline{\mu}_i(k) &= \underline{\mu}_i(k-1) + \gamma(k)(\underline{x}(k) - \underline{\mu}_i(k-1)) \quad &(\mathrm{i=r}) \\ &= \underline{\mu}_i(k-1) \quad &(\mathrm{i \neq r}) \end{aligned} \qquad (12)$$

4) $k = k+1$, $\gamma(k) = 0.998\gamma(k-1)$ and go to 2.

b) Width Calculation

After the RBF centers have been found, the width is calculated. The width represents a measure of the spread of data associated with each node. Calculation of the width is usually done using the P-nearest neighbor algorithm. A number P is chosen and for each center, the P nearest centers are found. The root-mean squared distance between the current cluster and its P nearest neighbors is calculated, and this is the value chosen for σ. So, if the current cluster center is $\underline{\mu}_j$, the value of width is given by:

$$\sigma_j = \sqrt{\frac{1}{P}\sum_{i=1}^{P}(\underline{\mu}_j - \underline{\mu}_i)^2} \qquad (13)$$

A typical value of P is 2, in which case σ is set to be the average distance from the two nearest neighboring cluster centers.

c) Weight Estimation

Learning in the output layer is performed after calculation of the centers and widths of the RBF in the hidden layer. The objective is to minimize the error between the observed output and desired one. It is commonly trained using the LMS algorithm [12] and is summarized as follows:

Training sample: Input signal vector = $\underline{\Phi}(k)$

Desired response = $\underline{d}(k)$

User-selected parameter: $0 < \eta < 1$

Initialization: Initialize the weights $\underline{\hat{w}}(0)$.

Computation: For k = 1, 2, …. Compute

$\underline{e}(k) = \underline{d}(k) - \underline{\hat{w}}^T(k)\underline{\Phi}(k)$.

$\underline{\hat{w}}(k+1) = \underline{\hat{w}}(k) + \eta\,\underline{\Phi}(\mathrm{k})\,\underline{\mathrm{e}}(\mathrm{k})$.

3) Classifying schizophrenia and bipolar using NNs

Multilayer perceptron with backpropagation and radial basis function with k means clustering algorithm are programmed using C++ programming language [14]. The input layer for both neural networks consists of 128 source nodes as mentioned before. The initial weights for multilayer perceptron with BP and RBF neural networks are taken randomly from the interval [0, 1]. The learning rate given the value 0.05.Choosing the number of the hidden layer nodes for MLP and RBF is shown in the section 4. The number of output layer nodes depends on diseases needed to be discriminated. In our paper the corresponding desired outputs in both MLP and RBF are (1,0,0),(0,1,0), (0,0,1) for Normal, schizophrenia and bipolar disorder respectively. The number of samples is 230 EEG power spectra for all cases with 80 samples representing BP, 80 samples representing Sc and 70 samples representing normal. 80 samples (20 normal, 30 BP and 30 Sc) have been used as training samples for the two classifiers used while 150 samples have been used for testing. After training two classifiers, selected 150 samples are used to test the two classifiers (50 from Normal, 50 from schizophrenic patients and 50 from bipolar disorder patients).

IV. RESULTS AND DISCUSSION

Before applying the two models of ANN for discrimination, the network needs to be trained to optimize the network performance. The effect of the number of hidden neurons for MLP is presented in Table 1. With a learning rate of 0.05, the best performing hidden node setup is 30 for first hidden node and 15 for second hidden node compared with other combinations of hidden nodes. Table 2 shows the effect of the number of different RBF clusters with a fixed learning rate, the best performance is obtained when the number of clusters is 10.

Table 1: Effect of the Number of Hidden Nodes on Performance of MLP ANN (testing results)

No of Nodes in 1st hidden layer	No of Nodes in 2nd hidden layer	Accuracy Normal %	Accuracy schizophrenia %	Accuracy bipolar %
20	10	99.06	95.45	96.37
30	15	99.43	98.34	98.13
40	20	97.40	91.16	92.78
50	25	93.53	93.44	97.31
60	30	93.13	97.42	95.17
70	35	98.10	96.51	96.18

Table 2: EFFECT of the Number of clusters in Hidden layer on Performance of RBF based ANN

No of clusters in hidden layer	Accuracy Normal %	Accuracy bipolar %	Accuracy schizophrenia %
4	77.06	86.34	82.38
10	89.79	88.07	84.16
20	71.85	86.05	82.23
30	71.87	83.71	83.60
40	84.05	75.29	81.64
50	66.87	79.67	75.45

After training both networks, 150 selected test sample are used (50 from Normal, 50 from schizophrenic patients and 50 from bipolar disorder patients). The BP ANN was tested with structure of four layers back-propagation networks (128 input nodes, 30 first hidden nodes, 15 second hidden nodes and 3 output nodes). The radial basis function neural network used consists of (128 input node with 10 clusters and 3 output nodes).The performance of MLP with backpropagation neural network is given in Table3; it shows that The MLP neural network performed best with average classification accuracy of 98.67% with respect to all test samples. For RBF network with k means clustering algorithm, Table4 shows the performance of network of RBF neural network where classifying accuracy reached 87.33%.

Table 3: Performance of MLP with BP algorithm in Classification of EEG Power Spectra from the three classes N, SCZ, and BD

subjects	No. of tested data sets	Correct classifications	Incorrect classifications	classification performance
Normal	50	50	0	100%
schizophrenia	50	49	1	98 %
Bipolar	50	49	1	98%
Total	150	148	2	98.67 %

Table 4: Performance of RBF algorithm in Classification of EEG Power Spectra from the three classes N, SCZ, and BD

subjects	No. of tested data sets	Correct classifications	Incorrect classifications	classification performance
Normal (N)	50	45	5	90.00%
Schizophrenia	50	43	7	86.00%
Bipolar	50	43	7	86.00 %
Total	150	131	19	87.33%

V. CONCLUSION

The discrimination of Schizophrenia and bipolar disorder is a significant problem which requires the use of intelligent algorithms. Two classifiers used for discrimination which are MLP with BP neural network and RBF neural network. Features extracted from EEG of 230 subjects (70 normal, 80 schizophrenic patients and 80 bipolar patients) are used. Applying the two classifiers on 230 subjects have showed that performance rate for the MLP with BP is 98.67 % while for radial basis function networks is 87.33%. The performance rates obtained from the study have shown that feedforward neural network with BP is preferred to solve the conflict in the diagnostic process between schizophrenia and bipolar diseases.

VI. REFERENCES

[1] American Psychiatric Association, & American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (DSM). Washington, DC: American psychiatric association, 143-7.

[2] V. Kusumakar, "Antidepressants and antipsychotics in the long-term treatment of bipolar disorder," *Journal of Clinical Psychiatry*, vol. 63, Suppl 10, pp. 23–28, 2002.

[3] Sachs, G. S., Printz, D. J., Kahn, D. A., Carpenter, D., & Docherty, J. P. (2000). The expert consensus guideline series: medication treatment of bipolar disorder. Postgrad Med, 1, 1-104.

[4] A. J. Niemiec and B. J. Lithgow, "Alpha-band characteristics in EEG spectrum indicate reliability of frontal brain asymmetry measures in diagnosis of depression," in Proceedings Int. Conf. of the IEEE Eng. in Medicine and Biology Society, pp. 7517–7520, Sept. 2005.

[5] P. Coutin-Churchman, et al., "Quantitative spectral analysis of EEG in psychiatry revisited: drawing signs out of numbers in a clinical setting," Clinical Neurophysiology, vol. 114, no. 12, pp. 2294–2306, Dec. 2003.

[6] Li, Y. J., & Fan, F. Y. (2006, January). Classification of Schizophrenia and depression by EEG with ANNs. In 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference (pp. 2679-2682). IEEE.

[7] M. Liu, “A Study of Schizophrenia Inheritance through Pattern Classification”, 2nd International Conference on Intelligent Control and Information Processing, Changsa, 2011.

[8] Alba-Sanchez F., “Assisted Diagnosis of Attention-Deficit Hyperactivity Disorder through EEG Bandpower Clustering with Self-Organizing Maps”, 32nd Annual International Conference of the IEEE EMBS, Argentina, September 2010.

[9] Hiesh, Ming-Hsien, et al. "Classification of schizophrenia using genetic algorithm-support vector machine (ga-svm)." 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2013.

[10] http://www.elazayem.com/elazayem%20hospital.htm

[11] Arnaud Delorme, Scott Makeig. “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis”. Neurosci Methods. Vol. 134. pp.9-21. 2004.

[12] El-Gohary, M. I., Al-Zohairy, T. A., Eissa, A. M., Eldeghaidy, S. M., & El-Hafez, H. M. A. (2015). EEG Discrimination of Rats under Different Architectural Environments using ANNs. International Journal of Computer Science and Information Security, 13(12), 24.

[13] Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. In The elements of statistical learning (pp. 485-585). Springer New York. ISO 690

[14] Rao, V., and Rao, H., (1996). C+ + Neural Networks and Fuzzy Logic, 1st edn. BPB Publications, New Delhi.