An intelligent System for Diagnosis of Schizophrenia and Bipolar Diseases using Support Vector Machine with Different Kernels

M.I. El Gohary ., T.A.Al Zohairy , A.M. Eissa., S. El Deghaidy , H.M. Hussein

ABSTRACT

Bipolar disorder and schizophrenia overlap in symptoms and may share some underlying neural behavior. The discrimination between the two diseases is one of the problems that face psychiatric experts. The present work will suggest a suitable solution to this problem based on artificial methods. The support vector machine (SVM) is used for discrimination dependent on recording of the EEG rhythms for patient. The large set of signals included in the EEG rhythms is reduced into smaller set after Fast Fourier Transform (FFT) segmentation. Different kernels are applied on the SVM which are linear, polynomial, quadratic and radial basis function. The application of SVM with different kernels for the EEG discrimination of the patients suffering schizophrenia and bipolar diseases is the aim of this work. Analysis of results have shown that the suggested algorithms will solve the discrimination problem between the two diseases. This is will be done using EEG waves and the support vector machine with linear and quadratic kernels, which have achieved a high performance rate reaching 98 % and 97.667% respectively compared to the other kernels applications.

***Keywords*— EEG, schizophrenia, Bipolar Disorder, support vector machine (SVM).**

I. Introduction

Schizophrenia (SZ) and bipolar disorder (BD) are two common psychiatric illnesses that share significant overlapping symptoms [1] such as cognitive features [2], genetic risk [3], and medication response [4]. The neurobiological underpinning of these disorders might provide basis for understanding their similarities and differences. Evidence suggests that SZ and psychotic bipolar disorder are strongly heritable [5]. So, an effective classification method is required to distinguish between SZ and BD patients in order to apply the right treatment to patient [6].

Recent studies [7] apply neuroimaging methods for diagnosis of schizophrenia and bipolar disorders to reveal discrete patterns of functional and structural abnormalities in neural systems which are critical for emotion regulation, mean while some other research works are employing traditional statistical methods that rely on the basic assumption of linear combinations but not appropriate for such tasks [8].

Classification is considered as a useful tool for medical diagnosis [9]. Fundamentally, classification approach could be established by medical experts to enable better understanding of diagnosis. Recent research studies contributed to the classification of diseases using techniques such as expert systems, artificial neural networks and SVM [10].

In this present work, an automated machine learning procedure that can diagnose specific forms of psychiatric illness using EEG of non-excited (without stimulation) patient’s suggested. Classification of psychiatric disorders using support vector machine with different kernels among three classes of diagnostic illness: SCZ, BD and healthy (N) are explained.

This research work presented as follows: In Section (II) subjects and methods are given. Support vector machine as a classifier is discussed in section (III). In Section (IV) the results and discussion are provided. Finally, conclusions is remarked.

II. SUBJECTS AND METHODS

A. Subjects

The EEG data were obtained from Abou-Elazayem Psychiatric Hospital in Egypt [11]. The subjects included 70 healthy persons who have no history of neurological or psychiatric disease, 80 schizophrenic and 80 bipolar disorder patients. All patients were hospitalized and diagnosed with schizophrenia or bipolar according to the criteria of diagnostic and statistical manual of mental disorders [12] 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 scalps 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 50 Hz using a 12 bit AD-converter and the data were recorded on a hard disk. For each subject, recordings covered the EEG activity in a resting condition (without stimulation) for time ranging approximately from 3 to 5 minutes.

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 dealing with high dimension data. Finding significant features that produce higher classification accuracy is an important issue.

The main steps for vector feature extraction of each subject are the EEGLAB toolbox [13], which is used to filter the original spontaneous EEG time series and remove the artifacts. A size of clear 2000 time-samples is selected which represent a 20 second intervals from each subjects. Fast forier transform (FFT) for 2000 points of all-time series obtained from 16 channels are calculated. The 2000 points of each channel are partitioned into 8 intervals and the mean value of each interval is calculated leading to 8 points. So, for each subject, there is 8 points for each one of the 16 channels which results in a single 128 feature vector for each subject and these will be used as input to the SVM for classification [14].

III. Support Vector Machine Algorithms

Support vector machine techniques [15] can be classified into three types namely, linearly separable, linearly inseparable and non-linearly separable.

A. Linearly Separable SVM

Linearly Separable classification separates the high dimensional data into two groups {+1, -1} without any overlapping or misclassification (Fig.1). SVM produces a number of decision margins where the best margin is identified by using perceptron algorithm.

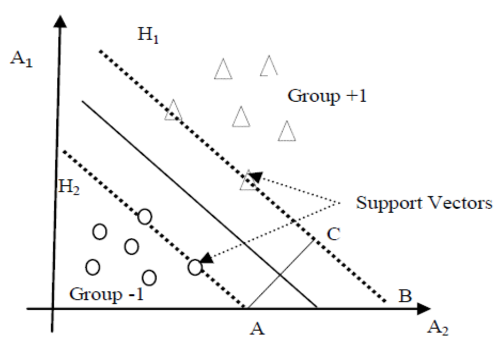

Fig.1: Linearly Separable SVM

The main objective of SVM is maximizing the margin width in order to reduce the misclassification error. The margin width can be calculated by drawing a line (AC) between H1 and H2 and forming a triangle ABC.

The distance between two hyperplanes is measured by calculating the length of AC.

$$\overline{AC} = \frac{2}{\|w\|}$$

The optimal hyperplane is given by the equation,

$$w_1.x_1 + w_2.x_2 - b = 0$$

The hyperplanes H_1 and H_2 are represented by

$$w_1.x_1 + w_2.x_2 - b = 1$$
$$w_1.x_1 + w_2.x_2 - b = -1$$

where w_1 , w_2 define the positions of hyperplanes H_1 and H_2 respectively. x_1 , x_2 are data points and b takes value of +1, 0, - 1 which shows how far the hyperplanes are away from the original line.

The maximum margin width is $\frac{2}{\|w\|}$ and minimum margin width is $\frac{1}{2}\|w\|$ under the constraint

$$y_i(w_1.x_1 + w_2.x_2 - b) \geq 1 \; for \; i = 1,2,..,m \qquad (5)$$

The parameters such as weight vector (w), bias (b), number of support vectors (m) are essential for classification. Margin width can be calculated from the values of 'w' and 'b' using optimization methods, which called primal problem. If the values of the parameters have larger values then it is difficult to calculate these parameters by primal method. This problem is considered as dual problem. Meanwhile these values can be calculated using Lagrange multipliers as follow:

Lagrangian multiplier for primal problem is given by

$$L\,(w,b,\alpha) = \left[\tfrac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i \,[y_i w - x_i +] - 1\right] \qquad (6)$$

where α_i is the Lagrangian Multiplier.

When applying the derivative of L with respect to w and b to zero, get

$$L\,(w,b,\alpha) = w - \sum_{i=1}^{n} \alpha_i x_i y_i = 0$$

This implies that,

$$w = \sum_{i=1}^{n} \alpha_i x_i y_i \qquad (7)$$

Apply from equation (7) into (6), get

$$L\,(w,b,\alpha) = \sum_{i=1}^{n} \alpha_i \tfrac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j x_i x_j y_i \, y_j$$

The associated Dual form is given by

$max \; L\,(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j x_i x_j y_i \, y_j$ subject to $\sum_{i=1}^{n} \alpha_i \, y_i = 0$. The optimum separating hyperplane (OSH) can be calculated by quadratic programming (QP).

B. Linearly Inseparable SVM

SVM for data contain noisy and faulty information which having possibilities of some error rate, it's impossible to construct a linear hyperplane without error for binary classification data as shown in Fig.2. Linearly inseparable classification can produce solutions for high dimensional data sets with overlapped or misclassified data. Slack variable ξ is used to represent the error term with slight modification in constraint (5) and allow misclassified points.

$$w \cdot x_i - b \geq +1 - \xi_i \text{ for } y_i = +1 - \xi_i \text{ where } \xi_i \geq 0 \; \forall i \qquad (8)$$

$$w \cdot x_i - b \leq -1 + \xi_i \text{ for } y_i = -1 + \xi_i \text{ where } \xi_i \geq 0 \; \forall i \qquad (9)$$

Combine (5) and (6) get,

$$y_i\,(w \cdot x_i + b\,) \geq \xi_i - 1\;, \xi i \geq 0\;\; \forall i \qquad (10)$$

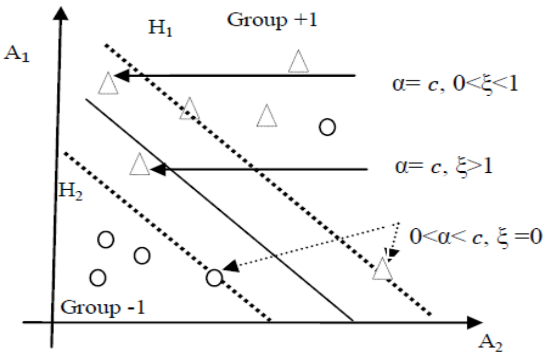

Fig.2: Linearly Inseparable SVM

Some data may be incorrectly classified and need to adjust the hyperplanes for proper classification. Increasing the margin on either side leads to increase in misclassification error rate. It can be minimized by the following function:

$$Min\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i$$

Where C is the constant to compromise between the size of the margin and ξ. In Figure.3, Data points will be correctly classified if αi and ξ_i are equal to zero or αi is equal to zero and ξ_i value is less than one. Data points will be misclassified, if α_i is equal to C and is greater than one. Data points are consider as Support Vectors, if α_i value must be greater than 0, and ξ_i is equal to zero.

Lagrangian formulation for dual form

$$Q(w, b, \xi\,, \beta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$-\textstyle\sum_{i=1}^{n}\alpha_i\;[y_i(w \cdot x_i + b\,) - 1 + \xi\,] - \beta_i\sum_{i=1}^{n}\xi_i \qquad (11)$$

The negative sign is used because the objective is to maximize margin width with respect to α_i , β_i and minimize with respect to w, b and ξ_i.

Differentiating with respect to w, ξ_i we get

$$w = \sum \alpha_i\, y_i\, x_i\;, \sum \alpha_i\, y_i = 0$$

And

$$C - \alpha_i - \beta_i = 0$$

Substitute in (11),

$$max\; L(\alpha) = \tfrac{1}{2}\textstyle\sum_{i,j=1}^{n}\alpha_i\,\alpha_j\,y_i y_j\,(x_i\,.\,x_j) + \sum_{i=1}^{n}\alpha_i \qquad (12)$$

with constraint

$$\sum \alpha_i\; y_i = 0\,,\; 0 \leq \alpha_i \leq c$$

we can get the value of w, b and ξ_i.

C. *Non Linearly Separable SVM*

Data that are not linearly separable can be converted into higher dimensional mapping for classification (Fig.3). The nonlinear mapping of original sample data which is transformed into higher dimensional mapping and is called Feature Mapping and its mapping function is denoted as $\varphi(x_i)$. In this case, Kernel functions are used to find the value of mapping function $\varphi(x_i)$.

$x_i^T\, x_j = k(x_i\,, x_j\,) = \varphi(x_i)^T\varphi(x_j)k\,(x_i\,, x_j\,)$ is called the Kernel function which is based on the inner product of two varients x_i , x_j. In original space dot product of x_i , x_j is used for calculation and it is converted into higher space dot which can be replaced by dot product as kernel function [16].

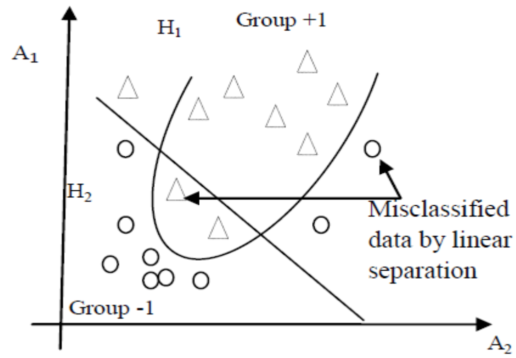

Fig .3: Non Linearly Separable

Some of the popular Kernel functions used in this work are:

Radial Kernel Function (RBF)

$$k\,(\,x_i\,,\; x_j\,) = e^{-\,\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Linear Kernel Function

$$k\,(\,x_i\,,\; x_j\,) = x_{i,}^T x_j$$

Polynomial Kernel Function

$$k\,(\,x_i\,,\; x_j\,) = [(x_{i,}^{\;T}\, x_j\,) + 1]^d$$

In the following example, SVM with linear kernel for finding optimal hyperplane of non-separable patterns is discussed.

Assume two classes of data to be classified using SVM with linear kernel. Each class consists of only one point. These points are:

$$\bar{x}_1 = A_1 = (1{,}1)\,, \bar{x}_2 = B_1 = (2{,}2) \qquad (1)$$

From SVM theory, we have the following two equations:

$$f(\overline{w}) = \tfrac{1}{2}\|\overline{w}\|^2 \qquad (2)$$

$$g_i(\overline{w}\,, b) = y_i[\langle \overline{w}\,, \bar{x}_i\rangle + b\,] - 1 \geq 0 \qquad (3)$$

We can expand $g_i(\overline{w}\,, b)$ to be:

$$g_1(\overline{w}\,, b) = (w_1\, x_{11} + w_2\, x_{12} + b) - 1 \geq 0 \qquad (4)$$

$$g_2(\overline{w}\,, b) = -(w_1\, x_{21} + w_2\, x_{22} + b) - 1 \geq 0 \qquad (5)$$

Next, we put the equations into the form of Lagrangian:

$$L(\overline{w}\,, b, \alpha\,) = f(\overline{w}) - \alpha_1 g_1(\overline{w}\,, b) - \alpha_2 g_2(\overline{w}\,, b)$$

$$L(\overline{w}\,, b, \alpha\,) = \frac{1}{2}\|\overline{w}\|^2 - \alpha_1\,(w_1\, x_{11} + w_2\, x_{12} + b - 1)$$

$$- \alpha_2\,(-(w_1\, x_{21} + w_2\, x_{22} + b)) - 1)$$

$$L(\overline{w}\,, b, \alpha\,) = \frac{1}{2}\|\overline{w}\|^2 - \alpha_1\,([w_1\, x_{11} + w_2\, x_{12} + b] - 1)$$

$$+ \alpha_2(w_1\, x_{21} + w_2\, x_{22} + b + 1) \qquad (6)$$

Solving the equations (8) to (12) obtained from the gradient of the following Lagrangian

$$\nabla L(\overline{w}\,, b, \alpha\,) = \nabla f(\overline{w}) - \nabla\alpha_1 g_1(\overline{w}\,, b) - \nabla\alpha_2 g_2(\overline{w}\,, b) = 0 \qquad (7)$$

Which are:

$$\frac{\delta}{\delta w_1} L(\overline{w}, b, \alpha) = \overline{w_1} - \alpha_1 x_{11} + \alpha_2 x_{21} = 0 \quad (8)$$

$$\frac{\delta}{\delta w_2} L(\overline{w}, b, \alpha) = \overline{w_2} - \alpha_1 x_{12} + \alpha_1 x_{22} = 0 \quad (9)$$

$$\frac{\delta}{\delta b} L(\overline{w}, b, \alpha) = -\alpha_1 + \alpha_2 = 0 \quad (10)$$

$$\frac{\delta}{\delta \alpha_1} L(\overline{w}, b, \alpha) = [w_1 x_{11} + w_2 x_{12} + b] - 1 = 0 \quad (11)$$

$$\frac{\delta}{\delta \alpha_2} L(\overline{w}, b, \alpha) = [w_1 x_{21} + w_2 x_{22} + b] + 1 = 0 \quad (12)$$

These equations are enough to find analytically the values of $\overline{w}, b$ *and* α analytically. Equating equations (11) and (12) we get:

$$[w_1 x_{11} + w_2 x_{12} + b] - 1 = [w_1 x_{21} + w_2 x_{22} + b] + 1 = 0$$

$$w_1 x_{11} + w_2 x_{12} - 1 = w_1 x_{21} + w_2 x_{22} + 1$$

$$[w_1 x_{11} + w_2 x_{12}] - [w_1 x_{21} + w_2 x_{22}] = 2$$

$$[w_1 + w_2] - [2w_1 + 2w_2] = 2$$

$$-w_1 - w_2 = 2 \xRightarrow{we\ get} w_1 = -(w_2 + 2) \quad (13)$$

Substituting from equation (13) into equation (8) and combining with equation (9) gives the following:

$$\overline{w_1} - \alpha_1 + 2\alpha_2 = 0$$

$$\overline{w_2} - \alpha_1 + 2\alpha_2 = 0$$

$$\alpha_1 = \alpha_2 \quad (14)$$

Which produce the next:

$$w_1 + \alpha_1 = 0 \quad (15)$$

$$w_2 + \alpha_1 = 0 \quad (16)$$

Equating equations (15) and (16) and putting the result back into Equation (13) this gives the following:

$$w_1 = w_2 = -1 \quad (17)$$

Using equation (17) in either of equations (15) or (16) will give:

$$\alpha_1 = \alpha_2 = 1 \quad (18)$$

Finally, using equation (18) in equations (11) and (12) gives:

$$b = 1 - (w_1 x_{11} + w_2 x_{12}) = 1 - (-1 - 1) = 3$$

$$b = -1 - (w_1 x_{21} + w_2 x_{22}) = -1 - (-2 - 2) = 3 \quad (19)$$

Note that this result also satisfying all of the conditions given in equation (3):

$$\alpha_1 (y_i [\langle \overline{w}, \bar{x}_1 \rangle + b] - 1) = 0$$

i.e.

$$\alpha_1 (w_1 x_{11} + w_2 x_{12} + b) - 1 = 0$$

$$([-1 - 1 + 3] - 1) = -2 + 3 - 1 = 0$$

$$\alpha_2 [w_1 x_{21} + w_2 x_{22} + b] - 1 = 0$$

$$([-2 - 2 + 3] + 1) = -4 + 3 + 1 = 0$$

which the inequality constraints:

$$\alpha_1 \geq 0 \ , \ \alpha_2 \geq 0 \quad (20)$$

IV. RESULTS

The Support vector machine with different kernels is programmed with matlab. MATLAB implementation of the feature vectors contained in a file called "svm_train" is used for training SVM classifier. Four different kernels are used to train the SVM model, linear, polynomial, quadratic and Gaussian radial basis functions. The number of samples used is 230 EEG power spectra for all cases. 80 different samples have been used as training data for the classifier while 150 different samples (50 from healthy (Normal), 50 from schizophrenic patients and 50 from bipolar disorder patients) have been used for testing.

The accuracy of the classifier is tested after training the SVM model. The function used for the classification process by MATLAB is "SVM-Test". Each of the different SVM kernel functions (linear, polynomial, quadratic and Gaussian RBF) are applied on the "SVM-Train" function independently.

Tables (1) to (4) show the performance of the SVM classifier when applying the four different kernels. According to the presented the tables it is found that linear SVM kernel the best performed with average classification accuracy of 98% with respect to the number of correctly classified examples. This is followed by quadratic, polynomial and RBF with a classification accuracy of 97.33%, 92% and 72.67 % respectively.

Table 1: Performance of Quadratic SVM kernel for Classification of EEG Power Spectra from the three classes N , SCZ, BD

	Tested Data	Number of correct sample	Number of incorrect sample	Percentage of Recognition
Normal	**50**	**50**	**0**	**100.00%**
Bipolar	**50**	**48**	**2**	**96%**
schizophrenia	**50**	**48**	**2**	**96%**
Total	**150**	**146**	**4**	**97.333%**

Table 2: Performance of Linear SVM kernel for Classification of EEG Power Spectra .

	Tested Data	Number of correct sample	Number of incorrect sample	Percentage of Recognition
Normal	**50**	**50**	**0**	**100.00%**
Bipolar	**50**	**49**	**1**	**98%**
schizophrenia	**50**	**48**	**2**	**96%**
Total	**150**	**147**	**3**	**98%**

Table 3: Performance of Polynomial SVM kernel Classification of EEG Power Spectra.

	Tested Data	Number of correct sample	Number of incorrect sample	Percentage of Recognition
Normal	**50**	**48**	**2**	**96%**
Bipolar	**50**	**47**	**3**	**94%**
schizophrenia	**50**	**43**	**7**	**86%**
Total	**150**	**138**	**12**	**92%**

Table 4: Performance of RBF SVM kernel Classification of EEG Power Spectra.

	Tested Data	Number of correct sample	Number of incorrect sample	Percentage of Recognition
Normal	**50**	**45**	**5**	**90%**
Bipolar	**50**	**37**	**13**	**74%**
schizophrenia	**50**	**28**	**22**	**56%**
Total	**150**	**109**	**41**	**72.67%**

V. CONCLUSION

The discrimination of Schizophrenia and bipolar disorder is a significant problem which requires the use of strong optimizing algorithms. Features extracted from EEG of 230 subjects (70normal, 80 schizophrenic patients and 80 bipolar patients) are used. By applying support vector machine using four different kernel functions (linear, polynomial, quadratic, radial basis) to 230 subjects, the experimental results have shown that the proposed algorithms can solve the discrimination problem using EEG rhythms and the support vector machine where linear and quadratic kernels have achieved a high performance rate equal to 98 % and 97.667% respectively compared to the two other kernels.

REFERENCES

[1] Keshavan MS, Morris DW, Sweeney JA, Pearlson G, Thaker G, Seidman LJ, et al. (2011): A dimensional approach to the psychosis spectrum between bipolar disorder and schizophrenia: The Schizo-Bipolar Scale. Schizophr Res 133:250–254.

[2] Chai XJ, Whitfield-Gabrieli S, Shinn AK, Gabrieli JD, Nieto Castanon A, McCarthy JM, et al. (2011): Abnormal medial prefrontal cortex restingstate connectivity in bipolar disorder and schizophrenia. Neuropsychopharmacology 36:2009–2017.

[3] Bramon E, Sham PC (2001): The common genetic liability between schizophrenia and bipolar disorder: A review. Curr Psychiatry Rep 3: 332–337.

[4] Carpenter WT, Bustillo JR, Thaker GK, van Os J, Krueger RF, Green MJ (2009): The psychoses: Cluster 3 of the proposed meta-structure for DSM-V and ICD-11. Psychol Med 39:2025–2042.

[5] Narayanan, B., O'Neil, K., Berwise, C., Stevens, M. C., Calhoun, V. D., Clementz, B. A., & Pearlson, G. D. (2014). Resting state electroencephalogram oscillatory abnormalities in schizophrenia and psychotic bipolar patients and their relatives from the bipolar and schizophrenia network on intermediate phenotypes study. Biological psychiatry, 76(6), 456-465.

[6] Phillips, M. L., & Frank, E. (2006). Redefining bipolar disorder: toward DSM-V. American Journal of Psychiatry, 163(7), 1135-1136.

[7] Erguzel, T. T., Sayar, G. H., & Tarhan, N. (2015). Artificial intelligence approach to classify unipolar and bipolar depressive disorders. Neural Computing and Applications, 1-10.

[8] Ritchie MD, White BC, Parker JS et al (2003) Optimization of neural network architecture using genetic programming improves detection and modeling of gene-gene interactions in studies of human diseases. BMC Bioinformatics 4:28

[9] Freitas AA (2002) A survey of evolutionary algorithms for data mining and knowledge discovery. In: Ghosh A, Tsutsui S (eds) Advances in evolutionary computation. Springer, Berlin, pp 819–845

[10] Wong ML, Lam W, Leung KS et al (2000) Discovering knowledge from medical databases using evolutionary algorithms. IEEE Eng Med Biol Mag 19(4):45–55

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

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

[13] Brunner, C., Delorme, A., & Makeig, S. (2013). Eeglab–an open source matlab toolbox for electrophysiological research. Biomed Tech, 58, 1.

[14] 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.

[15] Byun, H., & Lee, S. W. (2002). Applications of support vector machines for pattern recognition: A survey. In Pattern recognition with support vector machines (pp. 213-236). Springer Berlin Heidelberg.

[16] Corinna Cortes, Vladimir Vapnik, "Support Vector Networks", Machine Learning pp.273-297, 1995.