Machine Learning Approach to Diagnose Schizophrenia Based on Effective Connectivity of Resting EEG Data

K. Masychev[1], C. Ciprian[1] and M. Ravan[2]

1. Dept. of Computing Science, New York Institute of Technology, New York, NY, USA
2. Dept. of Electrical & Computer Eng., New York Institute of Technology, New York, NY, USA
{ kmasyche, ccipri02, mravan}@nyit.edu

Abstract— Schizophrenia is a severe mental disorder associated with neurobiological deficits. Despite the fact that the brain activity during tasks (i.e. P300 activities) are considered as biomarkers for diagnosing schizophrenia, brain activities at rest has the potential to reveal an intrinsic dysfunctionality in patients with schizophrenia and can be used to understand the cognitive deficits in these patients. In this study, we develop a machine learning (ML) algorithm based on eye closed resting-EEG data sets aiming to distinguish 63 schizophrenic patients (SCZs) from 70 healthy controls (HCs). The ML algorithm has three steps. In the first step an effective connectivity named symbolic transfer entropy (STE) is applied to the EEG waveforms. In the second step brain network properties are constructed from STE. In the third step the ML algorithm is applied to brain network properties to determine whether a set of features can be found that successfully discriminates SCZ from HC. The findings of this study revealed that the most discriminating features are shortest-path-length between different brain regions that could achieve an accuracy of 96.15%, with a sensitivity of 100% and specificity of 92.86% using 20% of data samples as test dataset that is not used for training. These findings imply that resting EEG could contribute to our ability to distinguish SCZs from HCs, and that the STE effective connectivity may prove to be a promising tool for the clinical diagnosis of schizophrenia.

I. Introduction

Schizophrenia is a severe neuropsychiatric disorder affecting approximately 1.1% of US population [1]. Schizophrenia is characterized by noticeable psychotic symptoms including hallucinations, delusions, reduction in performance, and thought disorder. Based on neuroimaging evidence on structural, functional, and effective brain connectivity, a core deficit of schizophrenia can be proposed as failure of effective functional integration within and between brain areas [2]-[4].

There are several studies that prove functional connectivity alteration in patients with schizophrenia in comparison to healthy controls in response to external cognitive or sensorimotor stimulation [5]-[7]. However, resting state electroencephalography (EEG) functional connectivity reflects the intrinsic inter-neuronal connections in specific circuits such as default mode network (DMN) that are attenuated or interrupted during cognitive or sensorimotor tasks [8]. Therefore, investigating resting state functional connectivity may reveal an intrinsic functional disintegration of brain regions for schizophrenic patients. Various studies demonstrated that the mental abnormalities in schizophrenia can be reflected by brain networks [9]-[10]. Thus measuring brain network ability can potentially improve understanding of disruption in brain connectivity underlying schizophrenia.

Recently, there are an increasing number of studies that develop practical tools for diagnosing schizophrenia using machine learning (ML) techniques that applied to resting state EEG biomarkers. Here we briefly review the outcomes of six of the most recent studies in this area with the highest classification accuracy. In the first study, Boostani et al. (2009) [11] extracted several features including band power, Autoregressive (AR) model parameters, and fractal dimension from the recorded resting EEG data (286 features). They applied different classifiers to extracted features to classify the two groups of 13 schizophrenic patients (SCZs) and 18 healthy controls (HCs). The achieved the highest classification accuracy of 87.51% using Boosted version of Direct Linear Discriminant Analysis (BDLDA). In the second study, Sabeti et al. (2011) [12] applied ML algorithm to 20 SCZs and 20 HCs. They first selected the most informative EEG electrodes using the mutual information techniques. Several features including autoregressive model parameters, fractal dimension, and band power were used for classification. Using 20 EEG electrodes, the total number of features were 300 for each 1 sec features for 2min recorded data for each participant. Then, they employed genetic programming to select the best features from the selected electrodes. Two classification algorithms were used in this study: Linear Discriminant Analysis (LDA) and Adaptive boosting (Adaboost), where the final number of selected features were 65 and 80 for LDA and Adaboost, respectively. They obtained the classification accuracy of 85.90% and 91.94% using LDA and Adaboost classifier, respectively. In the third study, Liu et al. (2018) [13] applied ML approach to resting state EEG data of 40 clinically high-risk individuals (CHRs), 40 SCZs, and 40 HCs to investigate whether the EEG characteristics of these three groups can differentiate CHR and SCZ from each other and from HCs. Using von Neumann Entropy as linear eigenvalue statistics (LES) feature for each window of 200 samples EEG data (1500 features in total for 300000 samples EEG data), they showed Support Vector Machine (SVM) classifier achieved the highest classification performance of 91.16% for classifying SCZs from HCs and 73.31% for classifying the three classes of CHR, SCZ, and HCs. In the fourth study, Phang et al. (2019) [14] proposed a deep multi-domain connectome convolutional neural network (MDC-CNN) framework for classifying of resting state EEG derived brain connectome in patients with schizophrenia. By combining three connectivity features of 1) time-domain Vector Autoregressive (VAR) model coefficients, 2) the frequency-domain partial directed coherence (PDC), and 3) the network topology-based complex network (CN) measures (2730 features), they achieved the classification performance of 93.06% for classifying 45 SCZ and 39 HC. In the fifth study, Li et al. (2019) [15] used the inherent spatial pattern of network (SPN) features extracted from resting EEG data to classify 19 SCZ and 23 HC. Using 4 SPN filters, they achieved the highest accuracy of 88.10% with SVM classifier. In the sixth study, Oh et al. (2019) [16] applied a 11 layered CNN model to differentiate resting EEG of 14 SCZ and 14 HCs with deep learning. A total of 1142 EEG segments were used for each subject, where each segment consisted 6250 time samples and 19 electrodes. Therefore, the total number of sampling points were 1142×6250×19=135612500. The most significant features were automatically extract by CNN. Their proposed model achieved classification accuracies of 81.26% for subject based testing.

In most of these studies, the number of features is much higher than the number of subjects, which may cause overfitting. The other main disadvantage of these studies is using a small data set that affects the reliability of the classifiers performance. Particularly when the selected features display significant variability; larger training dataset is needed in order to have a reliable classification performance. Finally, some of these studies used deep learning which is costly to compute compared to traditional ML techniques and needs more training data to be consistent.

The objective of this study is to develop a new ML algorithm based on brain network properties constructed from effective connectivity to study distinguishing characteristics between 63 schizophrenics and 70 healthy brains using a small set of selected features.

Functional connectivity reflects the statistical dependencies of signals from different brain regions as typically revealed by cross-correlation, coherency, or phase lag index measures. In contrast effective connectivity (EC) more precisely measures the influence that a "node" exerts over another under a network model of causal dynamics and is inferred from a model of neuronal integration, which defines the mechanisms of neuronal coupling much more precisely than functional connectivity [17]. Measures of effective connectivity must meet four requirements to be a useful addition to already established methods [18]. These include: 1) independence of *a priori* definitions and models 2) the ability to detect strong non-linear interactions across brain function, 3) the ability to detect EC even if there is a wide distribution of interaction delays between the two signals, reflecting signal transmission through multiple pathways or over complex axonal networks 4) robustness against linear cross-talk between signals. Transfer entropy (TE) as a model free statistic that can measure the directed flow of information between two incidents, accomplishes all the above requirements and is therefore a suitable approach for the quantification of effective connectivity [18]. For these reasons, TE has gained growing application in neurological science.

Various techniques have been proposed to estimate TE (e.g. [19]-[21]). Most techniques, however, make great demands on the data, require fine-tuning of parameters, and are highly sensitive to noise, which limit their utility. Symbolic TE (STE) [21] which estimates TE through symbolization, is a robust, computationally efficient measure to quantify information flow in multidimensional dynamic systems. This makes STE a promising measure of the preferred direction of information flow between brain regions.

The novelty of this present study lies in applying the combination of brain network properties constructed from STE and ML methods to the characterization of schizophrenia. The combination of these two techniques results in a significant improvement in the ability to characterize schizophrenia with small number of features, relative to previous studies.

II. Method

A. Subjects

63 SCZ subjects (age [years]: avg. = 37.27, std = 8.98, min = 17, max = 56, 37 male subjects (58.7%) and 26 female subjects (41.3%)) as well as 70 HC (age [years]: avg. = 37.74, std = 16.57, min = 18, max = 81, gender: 38 male subjects (54.3%) and 32 female subjects (45.7%)) were participated in this study. All subjects were unpaid volunteers, who were recruited from Hamilton Psychiatric Hospital, Hamilton, Ontario to investigate if EEG data can differentiate SCZ from HC. All participants in this study filled the informed consent and were aware of the nature of the study. All SCZ subjects met Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) criteria for schizophrenia.

B. EEG Data

An experienced technician recorded 3 Eye Closed (EC) and 3 Eye Opened (EO) resting EEG data each with the duration of 3.5 min in a sound proof electromagnetically shielded room using 10-20 EEG setup with 20 electrodes. All the recording sessions were scheduled in the morning and the subjects were requested to avoid consuming coffee, drug, alcohol, and smoking immediately before the session. The signals were notch filtered at 60 Hz and band-pass filtered between [0.5 and 80 Hz] during the recording and digitalized with the sampling frequency of 204.8 Hz.

C. Data Pre-Processing

We first visually selected the best EC EEG data with minimum contamination with noise and artefact for each participant. Then we used wavelet enhanced ICA (wICA) method to extract and cancel artifactual parts of the selected EEG data [23]. WICA uses wavelet threshold to enhance artefact removal with independent components analysis and therefore better recovering the neural activities hidden in artefactual components. To minimize the artefacts, we first band-pass-filtered the EEG signal with cut off frequencies of 0.5 Hz and 50 Hz using EEGLAB Toolbox [24]. EEG artifacts were then removed using wICA algorithm (available at https://www. matworks/ matlabcentral. /fileexchange/55413-wica-data-varargin).

D. EEG-STE

Consider two random processes X= (x_1, x_2, …, x_N) and Y= (y_1, y_2,…, y_N), where x_i and y_i are the ith samples (originating from two different brain regions). Symbolic transfer entropy (STE) estimates TE by using a method of symbolization. Symbols are defined by reordering the amplitude values of time series x_i and y_i. For a given, but otherwise arbitrary i, m amplitude values $X_i = \{x_i; x_{i+d}; \ldots; x_{i+(m-1)d}\}$ are arranged in an ascending order $\{x_{i+(k_{i1}-1)d} < x_{i+(k_{i2}-1)d} < \ldots < x_{i+(k_{im}-1)d}\}$, where d is the time delay sample, and m is the embedding dimension. To simplify the calculation of probability distributions, X_i is then transformed into symbol sequence containing discretized symbols. A symbol sequence is defined as $\hat{X}_i = \{k_{i1}; k_{i2}; \ldots; k_{im}\}$, which is a sequence of the indexes of original elements constitute order patterns. Given symbol sequences $\hat{X}_i$ and $\hat{Y}_i$, STE is defined as [21]

$$T_{Y,X}^{S} = \sum p(\hat{X}_{i+t}, \hat{X}_i, \hat{Y}_i) \log_2 \frac{p(\hat{X}_{i+t}|\hat{X}_i, \hat{Y}_i)}{p(\hat{X}_{i+t}|\hat{X}_i)}, \quad (1)$$

where p denotes the transition probability density, the sum runs over all symbols, and t denotes a time step. We used EEGapp pipeline [25] to calculate STE value between each two EEG electrodes. In this app, first for each time segment of 10 sec, the STE value is computed for a set of d values (d=1:2:30) and the embedding dimension of m = 3. The STE with maximum value is then selected that corresponds to the correct time delay between the electrodes. Finally, the average of selected STEs among all segments is considered as the final STE between two electrodes. This results in 380 STE pairs of effective connectivity values at each frequency band of δ (1Hz -4Hz), θ(4Hz-8Hz), α (8Hz-13Hz), β (13Hz-30Hz),. and γ (30Hz-50Hz).

E. Brain Network Properties

In this study, the following 8 brain network properties of the effective connectivity at each frequency band were investigated: 1) clustering coefficients for each node (EEG electrode) (20 features), 2) global-efficiency (1 feature), 3) local-efficiency (20 features), 4) characteristic-path (1 feature), 5) shortest-path-length between each two nodes (380 features), 6) density (1 feature), 7) eccentricity (20 features), and 8) node-strength (20 features). Therefore, the total numbers of features for all 5 frequency bands are N_c=463×5= 2315 network properties for each subject are treated as candidate features for the ML portion of this study.

F. Machine Learning

In this study the data set consists of EC resting EEG for each of the M_t = 133 subjects, and their corresponding labels; 1 for the 63 SCZ, and 2 for the 70 HC subjects. ML algorithm employ a training set consisting of labelled samples from SCZ and HC subjects to the class of subject. "Features", defined as variables whose values differ between the SCZ and HC classes, are identified from a list of candidate variables, using various types of feature selection algorithms. These selected features then define a feature space. The job of a classifier is to optimally partition the available training samples into two separate regions (i.e. a SCZ region and a HC region) in the feature space. The class of a previously unseen sample can then be determined by extracting the selected features from the sample and plotting the corresponding point in the feature space. The proximity of each subject's point, in the feature space, to the regions in this feature space occupied by others who are known to be either SCZ or HC, then determines that subject's class.

In this study, first N_c = 2315 brain network properties values are extracted as candidate features from the EEG data of all subjects.

The second step is feature selection to reduce the number of features and therefore avoid over-fitting. In this step we first randomly divide the dataset into the train and test set to have 80% of the samples in each class for training and the remaining 20% for testing. We then use only the training data to do the feature selection, *i.e.* to identify a set of N_r most discriminating features ($N_r \ll Nc$) to distinguish between SCZ and HC. This ensures, that there is no data leakage and we are not using information that is in the test set to help with feature selection. Using training data, we used the filter based Relief algorithm [26] for feature selection, which is noise-tolerant and robust to feature interactions. The key idea of the algorithm is to select features according to how their values are similar for the neighboring samples in one class and difference for the neighboring samples in different classes [26].

In order to avoid choosing features that are dominant in few subjects, we used stratified k fold cross validation (SKF-CV) procedure [27] with k=5 to select the most discriminating features between two classes among all training subjects. The SKF-CV procedure is an iterative process, where in each iteration all the features associated with one particular fold are omitted from the dataset. The iterations repeat until all folds have been omitted once. Here, at each iteration, the Relief approach determines a list of $2N_r$ most discriminating features from the remaining k-1 folds. After completing all iterations, the N_r most repeated features were selected as the final set of selected features.

The third step is to indicate the class (label) of subjects based on the selected features. In our study, we compare the performance of four widely used classifiers: Support Vector Machine (SVM), Gaussian Naïve Bayes (GNB), K Nearest Neighbors (KNN), Random Forest (RF), and Linear Discriminant Analysis (LDA) algorithms. A Gaussian radial basis kernel function, which was determined using the sequential minimal optimization technique [28] is considered for the SVM algorithm, 3 neighbors is considered for KNN algorithm, and 12 trees are considered for the RF algorithm. These methods were implemented using the Statistics and Machine Learning Toolbox in MATLAB R2019b.

The final step is to evaluate the performance of the ML algorithms. Here we used unseen test dataset to effectively estimate and compare the performance of the algorithms.

III. RESULTS

Using Relief method, the lowest number of discriminating features between SCZ and HC that gave adequate classification performance was $N_r = 15$. The list of these 15 selected features are shown in Table 1. This number of features is much lower than 107 training samples that will avoid over-fitting (the feature to sample ratio is $15/107 \times 100 = 14.0\%$). From Table 1, all the features are selected from shortest-path-length between each two nodes, where 8 of selected features are from the occipital areas at different frequency bands (features 1-8). The other features are from centro-temporal (features 9), parieto-temporal (feature 10), frontal (features 11 and 13), and fronto-temporal (feature 12, 14, and 15), areas. Figure1 show the boxplot of these features for both SCZ and HC, where the black vertical line drawn from the minimum to the maximum data value, and the blue box drawn from the lower to upper quartile with a red horizontal line marking the median of the features. The standard deviation is approximately equal to three quarters of the difference between the upper and lower quartile. From the figure, the median of 10 features (features 1-8, 14-15) are significantly higher (p-value ≤ 0.05) for SCZs than HCs using Wilcoxon rank sum test. Rank sum tests the null hypothesis that data in two groups are samples from continuous distributions with equal medians, against the alternative that they are not.

Table 2 shows the sensitivity, specificity, and total accuracy for SVM, GNB, KNN, RF, and GNB classification approaches using 20% unseen test dataset. From Table 2, the SVM and KNN classifiers can discriminate SCZ from HC with the highest classification performance of 96.15%, followed by the RF classifier with the performance of 92.31% and GNB and LDA classifiers with the performance of 88.46%. The high classification performance across different classification algorithms prove the selected features are highly discriminating between the two classes

It is worth noting that most of the selected features are from the areas that are also identified in previous studies. Tohid et al. (2015) [29] conducted a systematic review that reports the results of the relevance of people with schizophrenia to the occipital lobe. They found out there is enough evidence that support the concept of decrease in the volume of the occipital lobe in SCZ patients. In another study, Abdul-Rahman et al. (2012) [30] investigated the relationship between arcuate fasciculus abnormalities and the psychotic symptoms in schizophrenia. They found that the fractional anisotropy reductions of the arcuate fasciculus in SCZ occurred in

Table 1. The 15 discriminating features between SCZ and HC groups.

Feature #	Frequency Band	Shortest Path Length
1	δ	O1 and Oz
2	δ	O1and O2
3	θ	O1and Oz
4	θ	O1 and O2
5	θ	O2 and O1
6	α	O1 and O2
7	β	O1 and O2
8	β	O2 and O1
9	θ	C4 and T4
10	θ	P4 and T4
11	δ	F7 and F8
12	α	F8 and T4
13	δ	F8 and F7
14	δ	Fz and T3
15	δ	T3 and F3

SCZ: schizophrenia, HC: Healthy Control

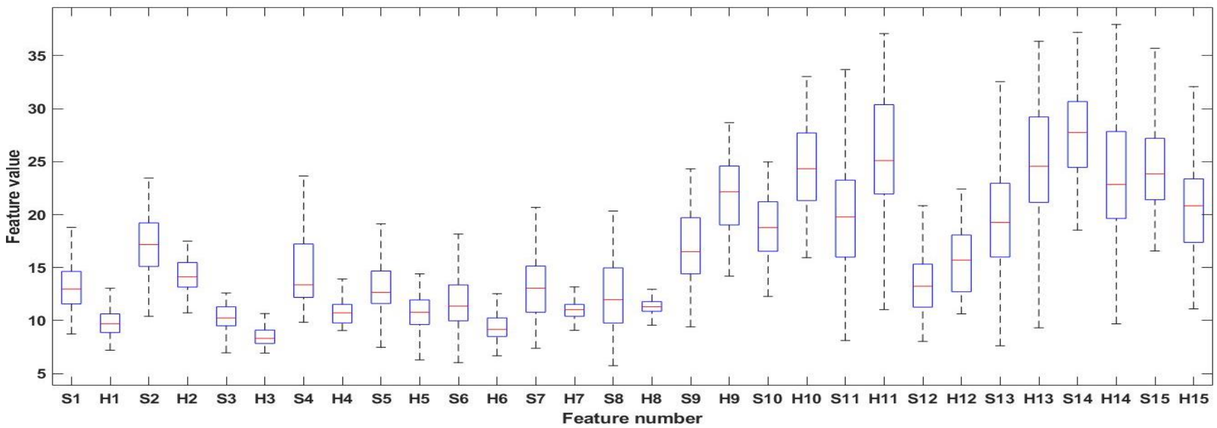

Figure1. Box plot of 15 features for schizophrenic and healthy subjects. S*i*: *i*th feature for schizophrenic patients, H*i*: *i*th feature for healthy controls. The red horizontal lines show the median of the features.

Table 2. Classification performance of different classifiers using test data.

Classifier	Sensitivity	Specificity	Total Accuracy
SVM	100	92.86	96.15
KNN	100	92.86	96.15
RF	100	85.71	92.31
GNB	83.33	92.86	88.46
LDA	83.33	92.86	88.46

left premotor cortex and supplementary motor cortex. The reductions in fractional anisotropy may reflect decreased myelination, loss of axonal membrane, or loss of coherence in the frontal arcuate fasciculus, which suggest disruptions in fronto-parieto-temporal connections [31], [32]. Furthermore Okugawa et al. (2002) [33] described bilateral reduction in grey matter volume of the temporal lobes. Finally, all the features are selected from shortest-path-length property of brain network. Li et al. (2019) [15] used four network properties clustering coefficient, characteristic path length, global efficiency, and local efficiency to recognize SCZ from resting EEG data. They achieved an accuracy of 54.76% with SVM classifier and 71.43% with LDA classifier. These low accuracies show that these features cannot encompass the entire spatial information related to network topologies. On the other hand, Sun et al. (2013) [34] found out among the network properties, shortest path length and clustering coefficient parameters have the ability to discriminate anatomical changes in neuron for SCZ patients using diffusion tensor imaging. They showed SCZ patients displayed comparatively lower clustering coefficient and longer path lengths that are consistent with our finding, where the shortest path length was longer for most of the selected features (10/15).

IV. Conclusions

In this study, for the first time, we used brain network properties constructed from STE to develop an EEG based ML algorithm for diagnosing schizophrenia. Based on the results, the algorithm could successfully distinguish 63 SCZs from 70 HCs by using 15 discriminating features with the high accuracy of 96.15%.

We note that the performances indicated in Table 2 are indeed higher than typical values obtained from previous studies with much lower number of features and complexity compared to the studies using deep learning approaches. We submit this improvement in performance is due to the effectiveness of STE method that was employed in the present study. Furthermore, the sample of 133 subjects is higher than most of the previous studies, which can increase the probability of accurate detection of schizophrenia.

Finally, the selected features are mostly from the shortest path lengths between areas that are in accordance with other research studies related to SCZ. It may, therefore, be reasonable to hypothesize that the STE effective connectivity extracted from resting EEGs could contribute towards a better understanding of the underlying pathophysiology of schizophrenic illnesses.

References

[1] *"Schizophrenia" NIMH Archived from the original on 4 Oct. 2016.* Retrieved 29 Dec.*2015.*

[2] K. J. Friston, "The disconnection hypothesis," *Schizophr. Res.*, vol. 30, no. 2, pp. 115-125, 1998.

[3] M. Ribolsi et al., "Abnormal brain lateralization and connectivity in schizophrenia," *Rev. Neurosci.* vol. 20, no. 1, pp. 61-70, Feb. 2009.

[4] A. Schmitt et al., "Schizophrenia as a disorder of disconnectivity," *Eur. Arch.Psychiatry Clin. Neurosci.*, vol. 261, Art. no. 261, pp. S150–S154, Nov. 2011.

[5] E. Pachou et al., Working memory in schizophrenia: an EEG study using power spectrum and coherence analysis to estimate

cortical activation and network behavior," *Brain Topogr.* vol. 21 pp. 128–137, 2008.

[6] T. Fujimoto et al., "Dysfunctional cortical connectivity during the auditory oddball task in patients with schizophrenia," *Open Neuroimag. J.* vol. 7, pp. 15-26, 2013.

[7] M. Ravan et al., "A machine learning approach using auditory odd-ball responses to investigate the effect of Clozapine therapy," *Clin. Neurophysiol.*, vol. 126, no. 4, pp. 721-730, Apr. 2015.

[8] M..D. Greicius et al., "Functional connectivity in the resting brain: a network analysis of the default mode hypothesis," *Proc. Natl. Acad. Sci. USA*, pp. 253-258, 2003.

[9] K. J. Friston and C. D. Frith, "Schizophrenia: a disconnection syndrome?," *Clin Neurosci,* vol. 3, no. 2, pp. 89-97, 1995

[10] .D. I. Leitman et al., "Sensory deficits and distributed hierarchical dysfunction in schizophrenia," *Am J Psychiat,* vol. 167, no. 7, pp. 818-827, Jul 2010.

[11] R. Boostani et al., "An efficient classifier to diagnose of schizophrenia based on the EEG signals," *Expert Syst. Appl.,* vol. 36. pp. 6492-6499, 2009.

[12] M. Sabeti et al., "A new approach for EEG signal classification of schizophrenic and control participants," *Expert Syst. Appl.*, vol. 38, pp. 2063–2071, 2011.

[13] H. Liu et al., "A data driven approach for resting-state EEG signal classification of schizophrenia with control participants using random matrix theory," *arXiv.org,* 2018.

[14] C. R. Phang et al., "Classification of EEG-based brain connectivity networks in schizophrenia using a multi-domain connectome convolutional neural network," *arXiv.org,* 2019.

[15] F. Li et al., "Differentiation of schizophrenia by combining the spatial EEG brain network patterns of rest and task P300," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp.594-602, 2019.

[16] S. Oh et al., "Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals," *Appl. Sci.* vol. 9, no. 14, 2870, 2019.

[17] K. J. Friston, "Functional and effective connectivity in neuroimaging: A synthesis," *Hum. Brain Map.*, vol. 2, no. 1-2, pp. 56-78, 1994.

[18] R. Vicente et al., "Transfer entropy—a model-free measure of effective connectivity for the neurosciences," *Jour. Comput. Neurosci.*, vol. 30, no. 1, pp: 45-67, Feb. 2011.

[19] A. Kaiser and T. Schreiber, "Information transfer in continuous processes," *Physica D: Nonlinear Phenomena*, vol. 166, no. 1-2, pp. 43-62, Jun. 2002.

[20] . F. Verdes, "Assessing causality from multivariate time series," *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.*, vol. 72, no. 2, Aug. 2005.

[21] M. Lungarella et al., "Information transfer at multiple scales," *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.*, vol. 76, no. 5, Nov. 2007.

[22] M. Staniek and K. Lehnertz, "Symbolic transfer entropy," *Phys. Rev. Lett.*, vol. 100, pp. 158101-1: 158101-4, Apr. 2008.

[23] N. P. Castellanos et al., "Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component Analysis," *J. Neurosci. Methods*, vol. 158, no. 2, pp. 300-312, 2006.

[24] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9-21, Mar. 2004.

[25] EEGapp, BIAPT lab, McGill University, https://github.com/BIAPT/EEGapp/wiki, (accessed 23 December 2017).

[26] H. Liu and H. Motoda, *Computational methods of feature selection*. Chapman & Hall, 2008.

[27] T. Hastie et al., *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. Springer; Feb. 2009.

[28] J. Platt, "Sequential minimal optimization: a fast algorithm for training support vector machines," *Microsoft, Redmond*, WA, USA, Tech. Rep. MSR-TR-98-14, Apr. 1998.

[29] H. Tohid et al., "Alterations of the occipital lobe in schizophrenia," *Neurosciences*., vol. 20, no. 3, pp. 213-24, 2015.

[30] M. F. Abdul-Rahman et al. "Arcuate fasciculus abnormalities and their relationship with psychotic symptoms in schizophrenia," *PLoS One,* vol. 7, no. 1, e29315, 2012.

[31] F. Aboitiz and V. R. Garcia, "The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective," *Brain Res. Rev*. vol. 25, pp. 381-396, 1997.

[32] T. Paus et al., "Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges," *Eur. J. Neurosci.* vol. 8, pp. 2236-2246, 1996.

[33] G. Okugawa et al. "Reduced grey and white matter volumes in the temporal lobe of male patients with chronic schizophrenia," *Eur. Arch. Psychiatry Clin. Neurosci*. vol. 252, pp. 120-123, 2002.

[34] Y. Sun et al., "Structural connectivity analysis reveals topological aberrations in patients with schizophrenia," *Proc. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013.