

The research of EEG feature extraction and classification for subjects with different organizational commitment

Rui Zhang^{1,2}, Ziyang Wang², and Yu Liu^{2,*}

¹School of Artificial Intelligence, University of Chinese Academy of Sciences, China

²Institute of Automation, Chinese Academy of Sciences, China

Abstract. With the development of EEG analysis technology, researchers have gradually explored the correlation between personality trait (such as Big Five personality) and EEG. However, there are still many challenges in model construction. In this paper, we tried to classify the people with different organizational commitment personality trait through EEG. Firstly, we organized the participants to complete the organizational commitment questionnaire and recorded their resting state EEG. We divided 10 subjects into two classes (positive and negative) according to the questionnaire scores. Then, various EEG features including power spectral density, microstate, functional brain network and nonlinear features from segmented EEG sample were extracted as the input of different machine learning classifiers. Next, several evaluation metrics were used to evaluate the results of the cross-validation experiment. Finally, the results show that the EEG power in α band, the weighted clustering coefficient of functional brain network and the Permutation Entropy of EEG are relatively good features for this classification task. Furthermore, the highest classification accuracy rate can reach 79.9% with 0.87 AUC (the area under the ROC). The attempts in this paper may serve as the basis for our future research.

Keywords: EEG feature extraction, Machine learning, Power spectral density, Functional brain network, Organizational commitment.

1 Introduction

The concept of Organizational Commitment was put forward by Becker[1], an American sociologist. Its initial definition was regarded as a psychological phenomenon that individuals have to remain in the organization with the increase of "unilateral investment" in the organization. The traditional organizational commitment analysis is carried out by means of psychological questionnaire. However, there have been some studies of EEG analysis for other personality traits similar to organizational commitment, especially Big Five Personality[2]. Butt et al.[3] used wearable multi-modal sensors to

* Corresponding author: yu.liu@ia.ac.cn

record a variety of signals including EEG during public speeches, and divided subjects into two classes based on the Big Five personality scale. They extracted the power spectral density features of different channels to classify the subjects. Finally, good results were obtained in the classification of different personality traits, and the highest classification accuracy rate reached more than 90%. In addition, some researchers used emotion-induced EEG to analyse the subjects' Big Five personality traits. For instance, Annisa et al.[4] used Discrete Wavelet Transform (DWT) to extract EEG signal features from the data of the public multimodal database ASCERTAIN for implicit personality and emotion recognition, and used Support Vector Machines for classification. Finally, the classification accuracy of the extraversion and neuroticism reached 69% and 75.9%. The above-mentioned studies mainly used relatively complex experimental paradigms to record EEG data, while Jach et al.[5] only recorded resting state EEG. They used multivariate pattern analysis to predict the scores of the Big Five personality traits, and finally concluded that the agreeableness personality score can be predicted based on the EEG power of the alpha wave and lower beta wave, and the neuroticism personality score can be predicted by the EEG power of theta wave. Although there has been no research group concentrating on EEG analysis for Organizational Commitment so far, there are many potentially useful features in EEG which can be extracted for Organizational Commitment analysis. Therefore, in this paper, we aim to extract various features of resting state EEG to classify subjects with different organizational commitment personality trait.

In the research of EEG analysis, an important content is the feature extraction from EEG. One of the most commonly EEG features is frequency domain features, which are represented by power spectral density (PSD)[6]. Di et al.[7] mainly used power spectrum analysis when using resting state EEG for identification. The power spectrum density of each channel were obtained by means of the fast Fourier transform of the Hamming window. Finally, the classification accuracy of multiple classifiers was up to 99%. Besides, Hazarika et al.[8] used wavelet transform to extract EEG features in the identification of patients with mental disorders, and then selected artificial neural networks for classification. Finally, their method correctly classified more than 66% of the normal class and 71% of the schizophrenia class.

In addition, researchers can also use methods related to nonlinear dynamics theory to extract the feature of neuronal activity. Although the linear method has been widely used in the research of EEG signals and achieved good results, it cannot extract the complex and chaotic behaviour behind the EEG[9]. Therefore, some researchers applied nonlinear analysis methods to EEG-related research. The nonlinear features used in EEG analysis include complexity, entropy, etc. Zhao et al.[10] used correlation dimension, Kolmogorov entropy and Lempel-Ziv complexity feature analysis in the resting state EEG recognition study of heroin addicts, and detected the difference between the disease group and the normal control group. Ahmad et al.[11] combined linear and nonlinear methods to extract EEG features to identify cognitive states and resting states. Finally, they used a support vector machine for classification, with an accuracy of 87.5% (linear) and 92.1% (non-linear).

Furthermore, some researchers used functional brain network analysis[12] to construct brain network graphs, and then extracted the characteristic parameters of the graphs as EEG features. Wang et al.[13] used coherence analysis to construct the brain function connectivity matrix in the study of the identification of patients with psychogenic non-epileptic seizures, and then calculated the brain network clustering coefficient and global efficiency parameters as the characteristics of resting state EEG. Finally, they used linear discriminant analysis (LDA) to achieve a classification accuracy of 85%. In addition to coherence analysis, researchers can also use synchronization likelihood (SL)[14], phase

lock value (PLV)[15, 16], phase lag index (PLI)[17, 18], imaginary part of coherence (iCoh)[19] and other methods to construct functional brain networks in different tasks.

Based on the literature survey above, we conduct the subjects to complete the questionnaire of organizational commitment, and recording their resting state EEG. In this paper, we aim at extracting various EEG features and using different classification algorithm to achieve the classification for subjects with different organizational commitment personality traits.

2 Materials and methodology

2.1 EEG recording and preprocessing

In this research, subjects not only participated in the EEG recording experiment, but also filled out the organizational commitment questionnaire. Experimental tests show that the questionnaire has good reliability and validity. This paper analyses the Behavioural Organizational Commitment scale in the questionnaire. The scale contains 10 questions, with a full score of 40. We set the threshold at 21 points and those subjects who score above the threshold are class I subjects (with relatively high Behavioural Organizational Commitment scale score), and the rest are class II subjects (with relatively low Behavioural Organizational Commitment scale score). We selected 5 class I subjects (positive) and 5 class II subjects (negative) as the research objects of this paper. These subjects are all college students without mental illness.

We used Neuroscan's 64-electrode EEG equipment for EEG recording. In the process of online data recording, we used electrodes M1 and M2 as reference electrodes and recorded EEG with the sampling rate of 1000 Hz. Meanwhile, we recorded two-channel electro-oculogram(EOG) signal to assist in the removal of electro-oculogram artifacts. In this paper, we used the resting state EEG data of the subjects with their eyes closed, with a duration of 120 seconds. During the experiment, the subjects sat on a chair in a comfortable state to stay relaxed, and could not do any muscle movement at as much as possible to reduce the interference of artifacts.

Then, we preprocessed the data after obtaining the original EEG data. First, we eliminated the reference electrodes M1, M2 and the electrodes CB1, CB2, (retaining 60 channels of EEG data), and performed an average re-reference. Next, the EEG data was filtered by 1~30Hz band-pass filter in order to remove noise in different frequency bands. After that, the data was resampled in order to reduce the amount of data, and the sampling rate was reduced from 1000 Hz to 250 Hz. Since there are artifacts that are difficult to remove in EEG signals, such as electro-oculogram artifacts, we used independent component analysis (ICA)[6] to remove these artifacts.

Finally, the clean EEG data was obtained after we completed the above pre-processing work. In order to carry out subsequent analysis, the EEG data of all subjects was segmented, and the duration of each segment was 4 seconds (no overlap). In this way, the sample size of the data set has been further expanded. The subsequent analysis work is performed on a single segmented sample.

2.2 Methodology

In this paper, we extracted various features from resting state EEG in order to recognize different kinds of people based on machine learning. We used two methods to obtain the feature vectors including power spectral density analysis, functional brain network analysis. Here, the selected methods are introduced below briefly.

2.3 Power spectral density (PSD)

The most typical and commonly used feature of EEG signal is the rhythm or oscillation distributed in different frequency bands. Based on the Fourier transform, we can calculate the power spectral density of the EEG signal, which can describe the power distribution of the EEG signal along the frequency. We used periodogram method to calculate the spectrum of EEG signals, which is a simple, yet popular method. For the discrete EEG signal $X[n]$ ($n=1, 2, \dots, N$) with sampling rate F_s , the periodogram is calculated as follow[6]:

$$P(f) = \frac{1}{NF_s} \left| \sum_{n=1}^N X[n]w[n]e^{-j2\pi fn/F_s} \right|^2 \quad (1)$$

where $w[n]$ is the window function and N is the number of time points, and we choose the rectangular window function.

After obtaining the periodogram, we calculated the power in different rhythms (frequency bands) as the features of the EEG signal, that is, sum the power spectral density in each frequency band. The rhythms we study include: δ (1~4Hz), θ (4~8Hz), α (8~13Hz), β (13~20Hz)[20].

2.4 Functional brain network analysis

Functional brain network analysis is to perform connectivity analysis on multi-channel EEG to obtain a network connection matrix (i.e., brain network graph), and then to extract characteristic parameters in the graph structure as the resting state EEG features for subsequent analysis. Firstly, the EEG electrodes were regarded as the nodes of the functional brain network. Next, in order to separate the local EEG activity and eliminate the interference of low-frequency spatial signals generated in the deep brain structure, the spatial Laplace transform was used to convert the scalp EEG data into an estimated Current Source Density(CSD)[6]. Then, it's necessary to select appropriate connection analysis indicators to construct the adjacency matrix of network nodes. In this paper, we selected the phase lock value (PLV), which is used to evaluate the degree of phase synchronization. The calculation formula is as follow[21]:

$$PLV_{XY} = \left| \frac{1}{N} \sum_{n=1}^N e^{i\Delta\phi_{rel}(t_n)} \right| \quad (2)$$

where $\Delta\phi_{rel}(t)$ is the phase difference between $X(t)$ and $Y(t)$, and N is the data length (number of points). After that, a part of redundant connections in the network is eliminated by setting a threshold. After the above process, we have obtained a functional brain network. Finally, we extracted the structural parameters in the network diagram as the features of resting state EEG. In this paper, we chose the weighted clustering coefficient, strength and eigenvector centrality of the network graph as the EEG features. The weighted clustering coefficient[22] is the average "intensity" (geometric mean) of all triangles associated with each node. It takes into account the proportion of nodes' triangles formed for each node respect to the total number of triangles. The calculation expression is as follows:

$$C_w = \frac{3 * \#triangles}{\#all triplets} \quad (3)$$

The node strength is the sum of the connection weights between all nodes j connected to a certain node i . It is one of the basic network metrics that are ultimately required. The calculation formula is as follow[23]:

$$s_i = \sum_j W_{ij} \quad (4)$$

Eigenvector centrality[24] is a self-referential measure of centrality: nodes have high eigenvector centrality if they connect to other nodes that have high eigenvector centrality. The eigenvector centrality of node i is equivalent to the i -th element in the eigenvector corresponding to the largest eigenvalue of the adjacency matrix.

$$Wx = \lambda_1 x \quad (5)$$

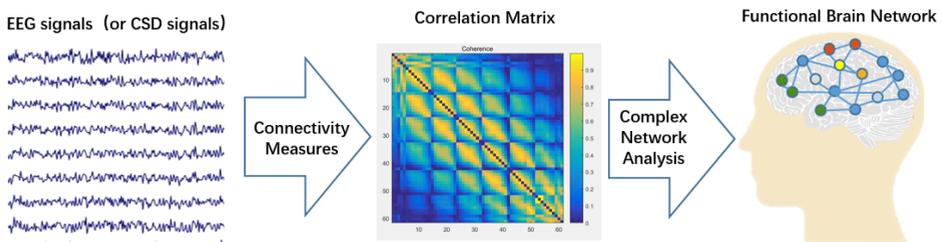


Fig. 1. Schematic flowchart of functional brain network analysis.

Features classification based on machine learning. We used EEG features and corresponding labels for supervised learning to obtain a model that can classify new observed features. In this paper, several classic machine learning models were selected for experiments, including non-linear and linear Support Vector Machines (SVM-RBF, SVM-Linear), Logistic Regression (LR), K-Nearest-Neighbors(KNN), Gaussian Naïve Bayes and Gradient Boosting[25].

In the process of machine learning, we used the extracted different EEG features as the input of the classifiers. In order to evaluate the effect of pattern recognition, we used a cross-validation method: randomly selected the data of a subject in class I (positive sample) and a subject in class II (negative sample) as the validation set, and the EEG data of the remaining 8 subjects as the training set. Therefore, there are a total of $5 \times 5 = 25$ pairs of cross-validation trials. In order to evaluate the classification effect, we selected a variety of evaluation indicators including Accuracy, Precision, Recall, the receiver operating characteristic (ROC) curve and the area under the curve (AUC), etc. The results in this article are the average of these 25 sets of cross-validation experiments.

3 Experimental results

3.1 Classification based on PSD feature

In this section, the power spectral density of 60 channels of EEG are calculated so that the power features at different frequency sub-bands are extracted. Table 1 shows the results of the different PSD features in the classification based on various classifiers. In this table, we consider four evaluation metrics, i.e. accuracy, precision, recall and F1-score. According to the results, we can find that different classifiers have different performances when classifying different features. The conclusions of the analysis are as follows.

δ -PSD feature: The experimental results show that subjects cannot be classified based on δ -PSD feature. Table 1 shows that the various classifiers based on this feature perform poorly.

θ -PSD feature: Based on this feature, the nonlinear SVM classifier (with RBF kernels) preforms best among the given classification algorithm.

α -PSD feature: Obviously, compared with other PSD features, this feature performs best in different classifiers. Among the given classification algorithm, the nonlinear SVM performs best according to the four evaluation metrics, which obtains overall accuracy of 74.3% with 74.1% precision, 74.5% recall and 74.3% F1-score. Although Logistic Regression classifier has higher accuracy, its precision is too high and its recall is too low, which indicates that the prediction of the classifier is unbalanced.

β -PSD feature: Among the given classification algorithm, Logistic Regression classifier relatively performs well.

Table 1. Different PSD features classification accuracy, precision, recall, F1-score using various classifiers.

Feature	Classifier	Accuracy	Precision	Recall	F1-score
δ -PSD	Nonlinear SVM	0.466	0.462	0.420	0.440
	Linear SVM	0.482	0.481	0.460	0.470
	Logistic regression	0.486	0.483	0.412	0.445
	KNN	0.481	0.471	0.304	0.370
	Gaussian NB	0.378	0.189	0.074	0.107
	Gradient Boosting	0.469	0.468	0.459	0.463
	θ -PSD	Nonlinear SVM	0.627	0.614	0.683
Linear SVM		0.627	0.657	0.529	0.586
Logistic regression		0.611	0.629	0.539	0.580
KNN		0.599	0.594	0.625	0.609
Gaussian NB		0.495	0.497	0.772	0.604
Gradient Boosting		0.520	0.519	0.551	0.534
α -PSD		Nonlinear SVM	0.743	0.741	0.745
	Linear SVM	0.723	0.848	0.544	0.663
	Logistic regression	0.746	0.850	0.597	0.701
	KNN	0.649	0.709	0.505	0.590
	Gaussian NB	0.647	0.713	0.493	0.583
	Gradient Boosting	0.699	0.738	0.616	0.672
	β -PSD	Nonlinear SVM	0.496	0.496	0.52
Linear SVM		0.641	0.701	0.489	0.578
Logistic regression		0.659	0.696	0.564	0.623
KNN		0.523	0.524	0.507	0.515
Gaussian NB		0.415	0.423	0.468	0.445
Gradient Boosting		0.548	0.549	0.541	0.545

The receiver operating characteristic (ROC) curves for different classifiers and different PSD features are shown as Figure 2. Considering the area under the ROC curves (AUC), we can conclude that the α -PSD feature is more suitable for the classification of subjects

with different organizational commitment. Besides, the nonlinear SVM classifier can be preferred.

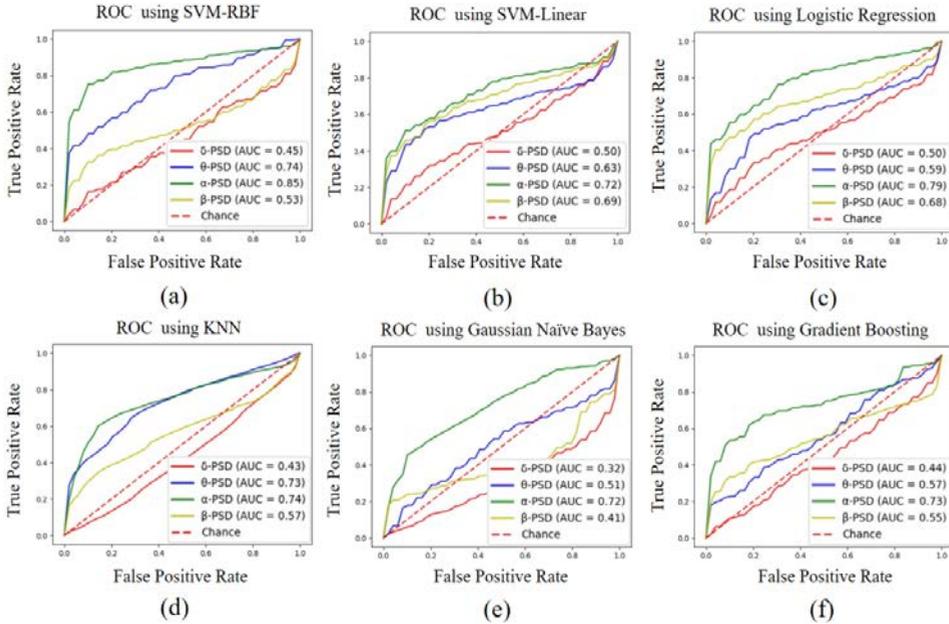


Fig. 2. ROC analysis of the various classifiers using various EEG PSD features.

3.2 Classification based on functional brain network feature

In this section, we construct the functional brain networks based on phase lock value (PLV). Then, three kinds of features (strength, weighted clustering coefficient and eigenvector centrality) are extracted from the network graphs. Since the network has 60 nodes (corresponding to 60 channels), the dimension of the feature is 60. Finally, these features are inputted independently to various classifiers. The results of classification are shown in Table 2. By contrast, we find that Eigenvector Centrality extracted from the functional brain network is not an effective classification feature. However, Strength and Weighted Clustering Coefficient can be used as the input features to obtain relatively good classification results.

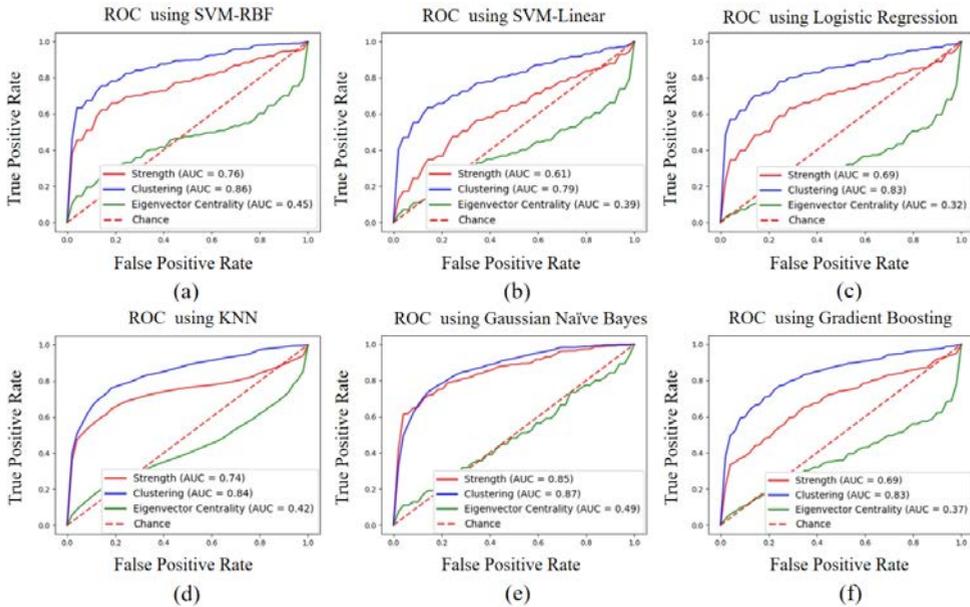
Strength: The best performing classifier is Gaussian Naïve Bayes classifier with overall accuracy of 75.7%, 82.5% precision, 65.2% recall and 72.9% F1-score. The second is nonlinear SVM classifier which achieves overall accuracy of 69.4%.

Weighted Clustering Coefficient: This is a relatively good feature which performs well in all given classifiers. All classifiers achieve the overall accuracy more than 70%, which indicates this feature is robust to the classification task. The Gaussian Naïve Bayes classifier performs best among all the given classifiers, with overall accuracy of 79.9%, 88.6% precision, 68.7% recall and 77.4% F1-score.

The receiver operating characteristic (ROC) curves for different functional brain network features and different classifiers are shown in Figure 3. It can be seen that the area under the curve (AUC) is 0.87 when we use Weighted Clustering Coefficient as the input feature of Gaussian Naïve Bayes classifier. Therefore, Weighted Clustering Coefficient is a relatively good feature in the classification task of subjects with different scores on the Organizational Commitment Scale.

Table 2. Functional network features classification accuracy, precision, recall, F1-score using various classifiers.

Feature	Classifier	Accuracy	Precision	Recall	F1-score
Strength	Nonlinear SVM	0.694	0.702	0.674	0.688
	Linear SVM	0.594	0.613	0.510	0.557
	Logistic regression	0.628	0.659	0.528	0.587
	KNN	0.659	0.691	0.574	0.627
	Gaussian Naïve	0.757	0.825	0.652	0.729
	Gradient Boosting	0.614	0.620	0.588	0.603
Clustering	Nonlinear SVM	0.791	0.840	0.720	0.775
	Linear SVM	0.709	0.739	0.647	0.690
	Logistic regression	0.748	0.779	0.691	0.732
	KNN	0.781	0.884	0.647	0.747
	Gaussian Naïve	0.799	0.886	0.687	0.774
	Gradient Boosting	0.756	0.777	0.717	0.746
Eigenvector	Nonlinear SVM	0.447	0.441	0.394	0.416
	Linear SVM	0.431	0.414	0.334	0.370
Centrality	Logistic regression	0.384	0.343	0.253	0.291
	KNN	0.426	0.357	0.183	0.242
	Gaussian Naïve	0.487	0.488	0.517	0.502
	Gradient Boosting	0.365	0.300	0.201	0.240

**Fig. 3.** ROC analysis of the various classifiers using various functional brain network features.

4 Conclusion

In this paper, we investigated the various EEG features including PSD and functional brain network features for the classification of subjects with different scores on the organizational commitment questionnaire. We used different classification algorithm in order to find relatively good classifiers suitable for the corresponding features. In order to evaluate the experimental results, we use accuracy, precision, recall, F1-score, ROC and AUC as

evaluation metrics in cross validation. According to the experimental results above, we conclude that the EEG power in α band, the weighted clustering coefficient of functional brain network of EEG are relatively good features for this classification task. Besides, it can achieve better classification results when we choose appropriate classifiers. The SVM classifier with RBF kernels based on multi-channel α -PSD features achieve overall accuracy of 74.3%, with 0.85 AUC; The Gaussian Naïve Bayes classifier based on multi-channel Weighted Clustering Coefficient features achieve overall accuracy of 79.9%, with 0.87 AUC. However, we only analysed the data of 10 subjects so far, so there may be problems in the results due to some accidental factors. Therefore, in future studies, we will consider recording more subjects' EEG for experimental analysis.

The authors would like to thank the staff of Institute of Psychology, Chinese Academy of Sciences for assisting in recording EEG data.

References

1. H. S. Becker, "Notes on the concept of commitment," *American journal of Sociology*, vol. 66, no. 1, pp. 32-40, 1960.
2. W. Li *et al.*, "EEG responses to emotional videos can quantitatively predict big-five personality traits," *Neurocomputing*, vol. 415, pp. 368-381, 2020.
3. A. R. Butt, A. Arsalan, and M. Majid, "Multimodal Personality Trait Recognition using Wearable Sensors in Response to Public Speaking," *IEEE Sensors Journal*, vol. PP, no. 99, pp. 1-1, 2020.
4. F. Q. Annisa, E. Supriyanto, and S. Taheri, "Personality Dimensions Classification with EEG Analysis using Support Vector Machine," in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2020, pp. 79-82: IEEE.
5. H. K. Jach, D. Feuerriegel, and L. D. Smillie, "Decoding Personality Trait Measures from Resting EEG: An Exploratory Report," *Cortex*, vol. 130, 2020.
6. L. Hu and Z. Zhang, *EEG Signal Processing and Feature Extraction*. Springer, 2019.
7. Y. Di, X. An, F. He, S. Liu, Y. Ke, and D. Ming, "Robustness Analysis of Identification Using Resting-State EEG Signals," *IEEE Access*, vol. 7, pp. 42113-42122, 2019.
8. N. Hazarika, J. Z. Chen, A. C. Tsoi, and A. Sergejew, "Classification of EEG signals using the wavelet transform," *Signal processing*, vol. 59, no. 1, pp. 61-72, 1997.
9. S. Mahato and S. Paul, "Detection of major depressive disorder using linear and non-linear features from EEG signals," *Microsystem Technologies*, vol. 25, no. 3, pp. 1065-1076, 2019.
10. Q. Zhao *et al.*, "Nonlinear dynamic analysis of resting EEG alpha activity for heroin addicts," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016, pp. 533-537.
11. R. F. Ahmad, A. S. Malik, H. U. Amin, N. Kamel, and F. Reza, "Classification of cognitive and resting states of the brain using EEG features," in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2016, pp. 1-5.
12. H. He *et al.*, "Resting-state functional network connectivity in prefrontal regions differs between unmedicated patients with bipolar and major depressive disorders," *Journal of affective disorders*, vol. 190, pp. 483-493, 2016.

13. Wang *et al.*, " Study on brain functional network analysis and classification recognition in patients with cardiogenic non-epileptic seizure based on resting state EEG," *Journal Biomedical Engineering*, vol. 32, no. 01, pp. 8-12, 2015.
14. H. Lei, J. Fan, S. Zhou, J. Dong, and X. Zhu, " Resting state brain function network EEG study in patients with obsessive-compulsive disorder," *Chinese Mental Health Journal*, vol. 6, 2018.
15. F. Al-Shargie, U. Tariq, M. Alex, H. Mir, and H. Al-Nashash, "Emotion recognition based on fusion of local cortical activations and dynamic functional networks connectivity: An EEG study," *IEEE Access*, vol. 7, pp. 143550-143562, 2019.
16. V. Gonuguntla, R. Mallipeddi, and K. C. Veluvolu, "Identification of emotion associated brain functional network with phase locking value," in *2016 38Th Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc)*, 2016, pp. 4515-4518: IEEE.
17. M. Fraschini *et al.*, "EEG functional network topology is associated with disability in patients with amyotrophic lateral sclerosis," *Scientific reports*, vol. 6, no. 1, pp. 1-7, 2016.
18. M. Yu *et al.*, "Different functional connectivity and network topology in behavioral variant of frontotemporal dementia and Alzheimer's disease: an EEG study," *Neurobiology of aging*, vol. 42, pp. 150-162, 2016.
19. M. Bola and B. A. Sabel, "Dynamic reorganization of brain functional networks during cognition," *Neuroimage*, vol. 114, pp. 398-413, 2015.
20. J. C. Henry, "Electroencephalography: basic principles, clinical applications, and related fields," *Neurology*, vol. 67, no. 11, pp. 2092-2092-a, 2006.
21. J. P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human brain mapping*, vol. 8, no. 4, pp. 194-208, 1999.
22. M. Hardmeier, F. Hatz, H. Bousleiman, C. Schindler, C. J. Stam, and P. Fuhr, "Reproducibility of functional connectivity and graph measures based on the phase lag index (PLI) and weighted phase lag index (wPLI) derived from high resolution EEG," *PloS one*, vol. 9, no. 10, p. e108648, 2014.
23. L. Wang, X. Guo, J. Sun, Z. Jin, and S. Tong, "Cortical networks of hemianopia stroke patients: a graph theoretical analysis of EEG signals at resting state," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 49-52: IEEE.
24. M. Hardmeier *et al.*, "Cognitive dysfunction in early multiple sclerosis: altered centrality derived from resting-state functional connectivity using magnetoencephalography," *PLoS One*, vol. 7, no. 7, p. e42087, 2012.
25. C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.