EasyChair Preprint

№ 10449

Fundamentals of Spectral Clustering for Extracting Microstates of EEG

Vladimir da Rocha Junior and Patrick Marques Ciarelli

September 27, 2023

Fundamentals of Spectral Clustering for Extracting Microstates of EEG

Vladimir da Rocha Cordeiro Junior[1] and Patrick Marques Ciarelli[1]

[1] Universidade Federal do Espírito Santo / Department of Electrical Engineering, Vitória, Brazil

Abstract— **The present work proposes a novel approach for multichannel Electroencephalogram (EEG) microstates extraction based on the fundamentals of spectral clustering algorithm as a lower cost alternative technique from the classical model. This approach involves microstates generated from the Laplacian matrix spectrum with special attention to the graph metric distances effects on the model performances. Results have demonstrated the potential of the technique to soften the clustering stages and encompass the variation present on the EEG. Two groups of subjects EEG have been used in this work, control and schizophrenic adolescents, and the experiments have presented the minimum of 79.64% explained variance for 6 microstates.**

Keywords— **EEG, Microstates, Spectral Clustering, Graphs**

I. Introduction

According to the World Health Organization (OMS), by 2019 almost one billion people was suffering from mental disorders [1], these statistics also include 14% of the world's teenager's population, and over 800,000 people die by suicide every year due to their effects. The situation is so alarming that the World Federation for Mental Health (WFMH) has promoted a suicide prevention campaign for that year [2] and has highlighted the situation for teenagers since 2018 [2]. Earlier diagnosis is essential for prevention.

In this context, Electroencephalogram (EEG) microstates are prominent candidates for an earlier diagnosis based on biomarkers [3][4][5], able to make the earlier diagnosis simple and objective. EEG microstates is a different manner of interpreting the information restrained in the EEG time series [6][7]. The EEG signal is one of the most important information sources for brain functional analysis due to its low cost and non-invasive nature [5][8].

However, the traditional clustering method to obtain microstates generally does not represent well the variations of the EEG signals with few microstates, needing a larger amount of microstates, making it difficult to analyze. Furthermore, with more microstates, more time is taken to process the EEG signal.

This work proposes a novel approach for designing EEG microstates, based on the Spectral Clustering techniques. The method focuses on the preprocessing stages in order to speed up the other steps on the microstates extraction, with natural dimension reduction and higher information a lower cost alternative technique achievement. Thus, with few microstates is possible to achieve a better representation of EEG signals.

II. Materails and methods

A. *The database*

The database (EEG of healthy adolescents and adolescents with symptoms of schizophrenia) used in this work is composed of EEG records from adolescents that have been analyzed by specialists and divided into two groups: 39 healthy male subjects (age between 11 to 13 years old), and 45 male subjects with symptoms of schizophrenia (age between 10 to 14 years old). The middle age in both groups is 12 years and 3 months. Each file has the EEG record for one subject. Each file has 122,880 samples, where every interval of 7,680 samples represents one channel of the EEG, going from 1 to 16 channels. The sampling rate is 128 Hz representing one minute of record. This public domain database can be downloaded from the Laboratory for Neurophysiology and Neuro-Computer Interfaces [9] and its classification was carried out by the National Mental Health Research center (NMHRC) experts. Each file was preprocessed as X_n^c, $c = 1,2..16, n = 1,2...7{,}680$. (c refers to the number of channels and *n* represents the number of EEG samples). After preprocessing, the files have been concatenated in one full dataset of 84 subsets (39 + 45).

B. *EEG Microstates*

EEG can be interpreted as electric scalp potential topography that shifts in millisecond range as a consequence of the underlying brain activity. More specifically, the scalp electric topography is related to the location of neurons in the gray matter and the amplitudes are related to the number of

organized active neurons [6]. Far from randomness, these 8 to 12 Hz scalp topographies are regarded as "atoms of thought" [10] or EEG microstates. Several studies have demonstrated the differences in microstates statistics from control groups to groups with different neuronal disorders [3][4][5][11], suggesting microstates as likely neuronal biomarkers, making possible earlier detection for neuronal and psychiatric disorders, even before clinical signs. Despite the practical EEG application, the automated microstates extraction and their analyzes are challenging processes, where several studies have been carried out to improve the workflow and enlighten the classification approaches [3][5][12].

The common EEG microstate workflow encompasses several techniques, where the steps leading to the final segment labeling and statistics pass through years since the first approach [13]. Clustering the EEG information is a crucial step for the microstates achievement, where few alternatives have demonstrated to be viable, such as K-means [14], Modified K-means [15], and Topographic Atomize Agglomerative Hierarchical Clustering (TAAHC) [12].
As aforementioned, the microstates are a different manner of interpreting the information restrained in the EEG and demands a preprocessing stage of the information received from the ordinary EEG. Figure 1 illustrates an example of the international 10-20 electrodes arrangement, and Figure 2 presents the EEG as a time series for each channel.

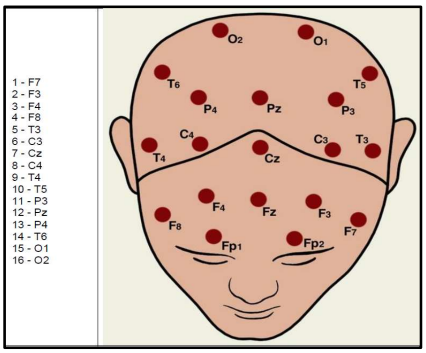

Figure 1. EEG 10-20 electrodes international system, adapted from [9].

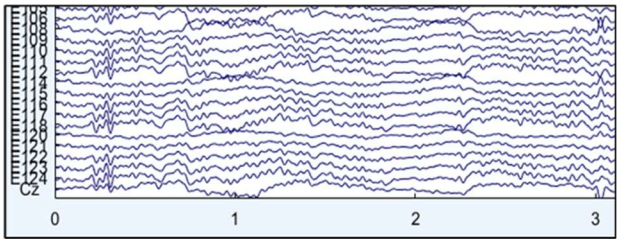

Figure 2. EEG multi-channel record, adapted from [12].

The microstate theory states that there are several segments in the EEG time series, each of them representing a specific electric topography on the scalp. However, few segments are more frequent in the alfa band (8 to 12 Hz) [6], as the EEG section is carried out with eyes closed, resting state, but not asleep. To find these prominent patterns, the Global Field Power (GFP), which is the same as the standard deviation among all the electrodes on the scalp, must be traced. Studies identified [10] that the microstates are found in the GFP due to the highest signal-to-noise ratio. Equation 1 defines the GFP, where u_i $(i = 1,2, \dots c)$ is the signal in the *i*th electrode, from $\boldsymbol{c}$ EEG electrodes, at time sample $\boldsymbol{t}$ [16].

$$GFP_t = \sqrt{\frac{1}{c}\sum_{i=1}^{c}(u_i(t) - \bar{u}(t))^2} \quad (1)$$

Once the patterns at GFP peaks are computed, an unsupervised learning algorithm is applied to cluster these topographies (or microstates). On average, 4 clusters can explain over 60% of the whole data set variation [17]. The Global Explained Variance (GEV) [16] is used to state how much information a set of microstates prototype encompasses from the EEG signal. Equation 2 presents the GEV_n for a given prototype, the GFP_n is the global field power, $\boldsymbol{u_n}$ is the n'th time sample, al_k is the prototype $k_i \in K$ (set of prototypes) assigned to u_n, N represents the number of time samples in the EEG.

$$GEV_n = Corr(u_n, al_n)^2 . (GFP_n^2 / \sum_{n'}^{N} GFP_{n'}^2) \quad (2)$$

The GEV is a measure of how similar each EEG sample (n) is from a microstate prototype. The most similar prototype to sample *n* is assigned to it. The sum of the GEV of every sample measures the amount of information a given prototype set encompass.
K-means clustering [14] is a basic clustering algorithm and belongs to a group named "*partition method*". Due to its simplicity and efficiency, it is one of the most widely used clustering algorithms. It encompasses the data set into a pre-set number of clusters, where every cluster is the nearest centroid of a data set partition. The algorithm recalculates the new centroids at every iteration until the convergence, which may occur at a pre-set number of iterations or the difference in the assigned points reaches a pre-set trigger.
Modified K-means was introduced specifically for treating the EEG signal [12] and to enhance the efficiency of K-means. There are two main differences between the techniques: the microstates topographies are polarity invariant, and the microstates activations are modeled.

Topographic Atomize and Agglomerative Hierarchical Cluster (TAAHC) belongs to another clustering technique group, the "*hierarchical clustering*" [12]. It derives from de Atomize and Agglomerative Hierarchical Cluster in the manner it

measures the quality of the cluster. Basically, in the TAAHC it starts the algorithm with every sample having their own cluster, then, the worst cluster is removed (atomized) at every iteration, and the process continues until a minimum pre-set number of clusters, or a minimum of two clusters.

C. *Spectral Cluster*

Spectral cluster is an unsupervised learning algorithm based on the spectral graph theory [18], its basic concept is to scatter the data points into several groups such that points in the same group are similar, and points in different groups have low similarity. Beginning from a Given set of n patterns $(x_1, x_2, \dots x_n)$, converted into a similarity graph (G) of vertices (points) and edges (link between vertices), we can compute a symmetric matrix of weighted edges W. The weight between two vertices x_i and x_j is zero $(w_{ij} = 0)$ if they are not linked, otherwise, the weight is a positive value $w_{ij} \geq 0$, with $w_{ij} = w_{ji}$.

For the same graph (G), the degree matrix (D) is defined as a diagonal matrix, with the degrees $(d_1 \dots d_n)$ on it, where the degree of a vertex (x_i) represents the number of other vertices (x_j) that are linked to it, according to Equation 3.

$$d_i = \sum_{j=1}^{n} w_{ij} \tag{3}$$

The vertex degree equation shows the sum only happens over the vertices adjacent to x_i, and zero to any other vertex. If taken a subset with vertices $A \subset V$(the group of all the vertices in G), and its complement $\underline{A}$. The indicator vector is $1_A = (f_1, f_2 \dots f_n)', \in R^n$, where the entries $f_i = 1$, if $v_i \in A$, and 0, for the complements. Finally, the unnormalized Laplacian matrix (L) is defined by Equation 4, W is the weight matrix and the D is the degree matrix.

$$L = W - D \tag{4}$$

In practice, from a matrix of n patterns X_n^c, turned into a similarity graph G (by approximation) and, the matrices in the Equation 4 would have the form according to the Figure 3.

$$W = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \quad D = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \quad L = \begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 4 & -1 & -1 & -1 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & 0 & -1 & 0 & -1 & 2 \end{bmatrix}$$

Figure 3. Weight matrix (W), adjacency matrix (D) and the Laplacian matrix (L), adapted from [19]

Note that the Laplacian matrix (L) has the same entries as D on the diagonal and, out of the diagonal, L has -1 if there is an edge between nodes of row i and column j, and zero, otherwise. Also, note that the sums of the elements in a row or column always is equal to zero, as the constraint of the Laplacian matrix, and that this is a block matrix with L_i blocks corresponding to the real clusters in the data. So, the matter is to find the spectrum of L. Given that, all these features give the following properties to L:

a. Equation 5 is valid for every vector $f \in R^n$, and there are different ways to manipulate this equation to draw the Laplacian matrix eigenvectors and eigenvalues.

$$fLf' = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2 \tag{5}$$

b. L is symmetric and positive semi-definite.
c. The smallest eigenvalue of L is 0, the corresponding eigenvector is the constant 1_A of ones.
d. L has a non-negative, real-valued eigenvalues $0 = \lambda_1 \leq \lambda_2 \dots \leq \lambda_n$.

Once we have the matrix spectrum, some heuristics can be used to try to identify the best number of clusters. Traditional clustering techniques, k-means for instance, can be used to group the data over the matrix spectrum.

Off course there are several constraints that have to be regarded, since at one step we turned the problem of L to work with zeros and ones weights, but the real problems bring non-binary wights, which demands a trade-off relaxing adjustment for L. Also, the construction of the graph (G) has a seed of randomness [18][19] to be regarded in order not to impair the results. In the end, even with the relaxing adjustment, the spectral cluster outperforms the traditional spectral methods [18] and fits to most clustering problems.

In this work, there is a point where the spectral clustering workflow is bent: the unnormalized Laplacian spectrum matrix is analyzed through the eigengap heuristic to identify a suitable number of eigenvectors to be kept on the spectrum matrix, U_n^m, $m \leq n$ (U is the spectrum of the symmetric Laplacian matrix), this procedure normally results in dimensionality reduction. Finally, K-means is applied over U_n^m to cluster the data into a pre-set number of clusters, that will operate as the prototypes mentioned in Section II. Note that this prototype works on the spectrum matrices spaces, this means that the whole operation takes place in the latent space and, since the U_n^m has all the rows of the initial data set, every pattern will be assigned to one of the clusters (EEG microstates) labels.

D. *The Spectral Microstates Workflow*

The complete workflow to design the microstates prototypes, in MATLAB©, is presented in Figure 4. The full dataset (see section II.A), split into training and test partitions, was developed as follows: the control group ranges from files 1 to 39

and the symptomatic range from 40 to 84, so, the first 70% from each group was separated to form the control and the symptomatic training subsets. In the sequence (third row of Figure 4), the potential prototypes are traced following the best signal-to-noise ratio: the EEG samples at the GFP peaks are extracted from a random sample of the training subsets (0.01 from each), this is the GFP peaks subset (a standard subset in this work), where the spectral clustering algorithm is carried out to draw U_n^m. The dimension, m, is set with the eigengap heuristic [18], regarding that two different groups should have vertex weight null (since they are not connected). The number of clusters is defined with K-means over U_n^m (fifth row of Figure 4). The output of this procedure is a set of latent clusters since the unsupervised learning occurs in the spectral space of the Laplacian matrix. For consistency, these latent clusters must explain the variance of the same data set they have arisen from (GFP peaks subset). This procedure is necessary for tracing the following goals:

1. Reduce the algorithm intrinsic loss due to the relaxing adjustment inherent from the algorithm [18].
2. Increase the consistency of results before different random seeds on the similarity graph design and clustering algorithm.
3. Define, if possible, the best parameters, like Laplacian normalization, number of eigenvectors for the spectrum matrix, metric "distance" for the similarity graph and K-means number of clusters.
4. Check the GEV achieved by the likely prototypes over the self-backfit (spectral prototypes compared to the sample in the latent space).

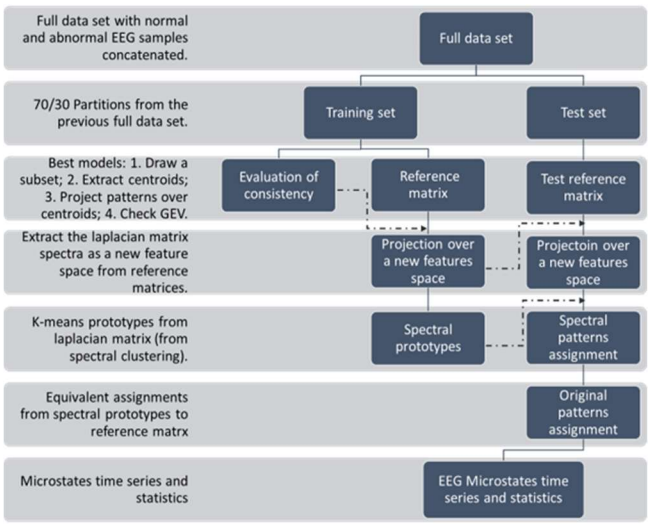

Figure 4. Spectral microstates workflow

After the evaluation of consistency, the most suitable parameters are used for designing the likely prototypes, which are re-evaluated over a random partition of the training subsets, 1% of each subset (fourth row in Figure 4). This new check is necessary for the previous evaluation was carried out over the GFP peaks subset. Following the workflow in Figure 4, every row receives a label based on its most similar prototype, based on the same metric distance used for extracting the prototypes. Since every row in the latent space has an equivalent row in X_n^c, the label is the same (fifth and sixth row of Figure 4). Once the labels are assigned, it is ready for the statistics or classification. In the sequence, the method was applied to a group of unknown groups (test sets) to evaluate whether the method can trace these common states in every instance.

III. Results

There are dozens of possibilities for the prototypes designing and any different parameter in the workflow will result in a new model. However, the most important group of parameters and hyperparameters are those related to the similarity graph. Depending on the metric distances used, the graph will better interpret relations among the patterns in the datasets, making the other steps clearer to extract the best eigenvector and clustering. One advantage of this approach is the natural handling of dimensionality following the eigengap heuristic, which grants a notion of the best number of eigenvectors for a given metric distance. Figure 5 presents the behavior of the training data set for the Spearman eigengap approach, projecting 200 eigenvalues.

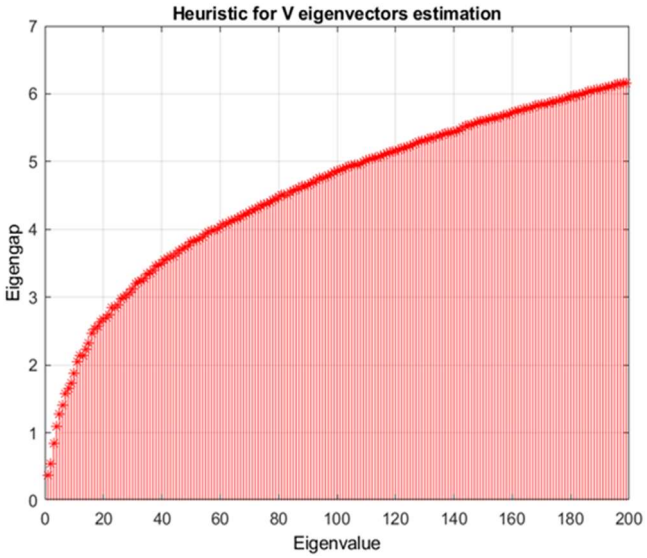

Figure 5. Eigengap heuristic for Spearman graph.

Although the projection is not clear, the most significant interval is between 1 to 5 eigenvalues, which leads to this natural shrinkage of dimensions. Then, the matrix U_n^m will keep the same number of patterns (n), but the most significant latent features or eigenvectors (m), ready for K-means clustering. Given the best latent dimension, the number of microstates were designed by the GEV variation with the number of clusters, regarding the minimum accepted variance is 60%. Figure 6 illustrates the GEV has highest variation between 4 to 6 clusters, Spearman on the left graphic and Jaccard on the right.

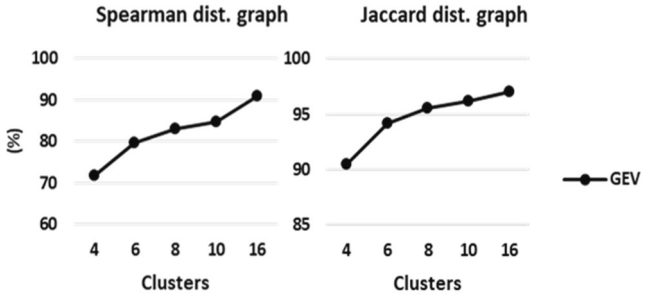

Figure 6. GEV growth for the number of clusters

In the end, two metrics could capture the information free from unwanted effects experienced during the experiments. Normally when the metric distance does not produce a proper graph, the samples tend to overfit in one cluster or the same microstates distribution is replicated for every EEG sample. Out of these effects, Jaccard and Spearman distances [20] have generated the most suitable information for the matrices D and U. The most reasonable GEV was achieved by combination of Jaccard (for the graph) and cosine distance for K-means and pairwise assignment, as highlighted in Table 1 (94.16% for training control group, and 91.68% for the training symptomatic). Regarding the techniques afore mentioned for eigenvectors (V) and for the number of cluster (K), the prototypes coincided in 6 clusters, but 4 dimensions for Spearman prototypes (6/4) and 3 dimensions for Jaccard ones (6/3).

Table 1 Results for general samples backfit

Metric distance for similarity graph	Metric distance for k-means clustering	Laplacian normalization	K	V	GEV for general training control (%)	GEV for general training symptomatic (%)
Spearman 6/4	Cosine	None	6	4	79.64	80.30
Jaccard 6/3	Cosine	None	6	3	**94.16**	**91.68**

K: number of spectral clusters V: number of eigenvectors in U

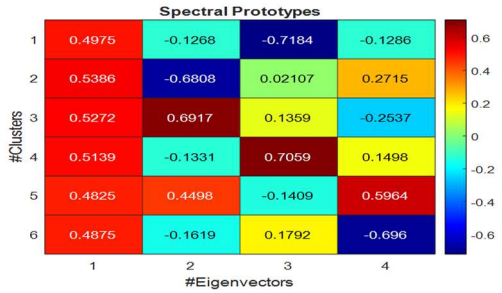

Figure 7. Spearman 6/4 prototypes (rows)

These protype performed the average GEV of 73.74% among all the EEGs for the control test set, and 77.57% among the EEG for the symptomatic test set. The performance of the entire groups of spectral prototypes is presented in Table 2. Again the Jaccard 6/3 prototypes performed better for explaining the data in the latent space, after every sample on the each EEG teste set received their label, and the standard deviation ranged from 2.43% in the control, to 2.58% in the symptomatic test set. Regarding the standard deviation for the Spearman 6/4, the GEV performed for one subject was inferior to those achieved to the rest of the test set, the reason for the standard deviation of 14.45%.

Table 2 Results for individual samples GEV

Model	GEV for control test set	Std for control test	GEV for symptomatic test set	Std for symptomatic test
Spearman6/4	73.74	14.45	77.57	2.26
Jaccard 6/3	**92.31**	2.43	**91.51**	2.58

Even though all the procedure takes place in the latent domain, every pattern in the EEG receives a label relative to its latent prototype.

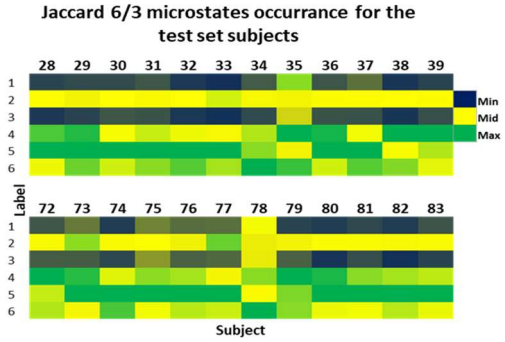

Figure 8. Jaccard spectral microstates ocurrence, control subjects range from 28 to 39 and symptomatic subjects range from 72 to 83.

Groups are distributed along the columns, blue for lesser and green for the most frequent. This distribution was drawn by the spectral prototypes designed by Jaccard 3/6. Similar effects have been observed using the other models. Thus, it is

possible to extract the statistics and trace the dependencies in the microstates time series for classification of each EEG file in control or symptomatic subject. Figure 8 shows the frequency of each label (1 to 6 in the rows) among the EEG file for each subject in the test set, 28 to 39 for the control group, and 72 to 84 for the symptomatic (one pattern has been omitted for convenience in the figure).

IV. Conclusion

In this paper, we proposed the spectral clustering approach over the EEG data as a novel alternative for the extraction of EEG microstates, partially operating during the preprocessing stage and partially in the new features generation. In other words, the algorithm is not used to cluster the data, it is applied to generate the Laplacian data spectrum before the clustering process. The preprocessing stage is an important step for the performance of several approaches, in the sense can soften some complex challenges, which has been noticed during the experiments. Among several parameters, only few have generated the suitable distributions (spectrums) for triggering the models to catch the slight differences and generate better spectral microstates. This stage is so important that the clustering algorithm over the spectrum becomes viable. This means that other models could possibly carry out the same task easily, once the spectral clustering algorithm has preprocessed the data. The metric chosen for assessing the hypothesis was the GEV. In this case, the appraisal took place in the latent space (one intrinsic aspect of the spectral clustering) and achieved high rates of information variability (one model achieved a minimum of 91.51%, as seen in Table 2). However, it is difficult to compare the results with classic approaches since they work in different domains and this task is for future works. When the whole model was applied over each EEG file in the test section, and the instances in the time domain have been assigned, some patterns have arisen as a time series (Figure 8), which state the efficiency of the hypothesis to extract the information able to support classification approaches. In the end, the novel approach has achieved the aims of the work, leaving several possibilities for further studies that can secure the spectral microstates as a viable alternative for analysis of EEG signals.

Acknowledgement

We thank the financial support for the research from the project of the Fundação de Amparo à Pesquisa do Espírito Santo (FAPES), number 598/2018.

Conflit of interest

The authors declare that they have no conflict of interest.

References

1. Mental health at https://www.who.int/news-room/fact-sheets/detail/mental-disorders
2. Publications at https://wfmh.global/what-we-do/publications
3. Alves, Lorraine M., et al. "Graph Theory Analysis of Microstates in Attention-Deficit Hyperactivity Disorder." *Congresso Brasileiro de Automática-CBA*. Vol. 2. No. 1. 2020.
4. Tait, Luke, et al. "EEG microstate complexity for aiding early diagnosis of Alzheimer's disease." *Scientific reports* 10.1 (2020): 1-10.
5. López, Silvia, I. Obeid, and J. Picone. *Automated interpretation of abnormal adult electroencephalograms*. Diss. 2017.
6. Koenig, Thomas, et al. "Millisecond by millisecond, year by year: normative EEG microstates and developmental stages." *Neuroimage* 16.1 (2002): 41-48.
7. Pascual-Marqui, Roberto D., Christoph M. Michel, and Dietrich Lehmann. "Segmentation of brain electrical activity into microstates: model estimation and validation." *IEEE Transactions on Biomedical Engineering* 42.7 (1995): 658-665.
8. Park, Su Mi, et al. "Identification of Major Psychiatric Disorders From Resting-State Electroencephalography Using a Machine Learning Approach." *Frontiers in Psychiatry* (2021): 1398.
9. EEG of healthy adolescents and adolescents with symptoms of schizophrenia at http://brain.bio.msu.ru/eeg_schizophrenia.htm
10. Lehmann, Dietrich, et al. "Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. Visual imagery and abstract thoughts." *International Journal of Psychophysiology* 29.1 (1998): 1-11.
11. Ku, David W., and Patrick M. Ciarelli. "Classificação Automática de Sinais Anormais de EEG por meio de Microestados e Aprendizado de Máquina." *Simpósio Brasileiro de Automação Inteligente-SBAI*. Vol. 1. No. 1. 2021.
12. Poulsen, Andreas Trier, et al. "Microstate EEGlab toolbox: An introductory guide." *BioRxiv* (2018): 289850.
13. Lehmann, D. "Human scalp EEG fields: evoked, alpha, sleep, and spike-wave patterns." *Synchronization of EEG activity in epilepsies*. Springer, Vienna, 1972. 307-326.
14. Lloyd, Stuart. "Least squares quantization in PCM." *IEEE transactions on information theory* 28.2 (1982): 129-137.
15. Arthur, David, and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Stanford, 2006.
16. Murray, Micah M., Denis Brunet, and Christoph M. Michel. "Topographic ERP analyses: a step-by-step tutorial review." *Brain topography* 20.4 (2008): 249-264.
17. Michel, Christoph M., and Thomas Koenig. "EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review." *Neuroimage* 180 (2018): 577-593.
18. Von Luxburg, Ulrike. "A tutorial on spectral clustering." *Statistics and computing* 17.4 (2007): 395-416.
19. Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. *Mining of massive data sets*. Cambridge university press, 2020.
20. Chomboon, Kittipong, et al. "An empirical study of distance metrics for k-nearest neighbor algorithm." *Proceedings of the 3rd international conference on industrial application engineering*. 2015.