

A proposal for improving EEG microstate generation via interpretable deep clustering with convolutional autoencoders

Arjun Vinayak Chikkankod^{*,†}, Luca Longo[†]

Artificial Intelligence and Cognitive Load Lab, the Applied Intelligence Research Centre, Technological University Dublin, Dublin D07 H6K8, Ireland

Abstract

Electroencephalography-based microstates, characterised as quasi-stable states of mental activation, encapsulate the spatio-temporal dynamics of brain signals. They are representative template topographic maps for time intervals, usually in the order of 60-120ms, extracted from the whole time duration of an EEG-based experiment. This extraction is currently performed with shallow clustering algorithms such as k-means or hierarchical clustering and trained with many 1D vectors containing scalp-electrode activations, each representing an instance point in time. However, this approach ignores the spatial position of these electrodes, which we argue is essential information for improving cluster formation. This study contributes to the body of knowledge by introducing deep clustering, leveraging recent advancements in computer vision, and autonomously learning feature representations from EEG data during clustering, capturing the inherent structure and patterns more effectively. In addition, relevant advances in eXplainable Artificial Intelligence have enabled various attribution methods to interpret trained models, which can be exploited to understand the inherent mechanisms for microstate formation.

Keywords

EEG Microstates, Shallow clustering, Deep clustering, Convolutional autoencoders, Resting state.

1. Introduction

Microstates provide fine-grained temporal and spatial characteristics of multichannel, non-stationary EEG signals. They are quasi-stable brain activation states that can last for 60-120 ms [1]. Initially, microstate analysis was performed mainly for alpha-filtered EEG signal [2]. Nowadays, it is executed predominantly on a broadband signal and with a few narrowband signals to analyse different spectral profiles in rest, task, and activity-evoked states [3, 4]. EEG microstates can capture global patterns of dynamically varying brain representations over time [5]. They have a few distinct scalp potential topographies that capture the spatiotemporal aspects of EEG brain signals with varying degrees of global explained variance [5]. Generally, four microstate maps are often extracted using clustering methods, which can explain variance in brain dynamics ranging from 64-84% [5, 6]. Although the EEG microstate is a powerful tool

Late-breaking work, Demos and Doctoral Consortium, colocated with The 2nd World Conference on eXplainable Artificial Intelligence: July 17–19, 2024, Valletta, Malta

^{*}Corresponding author.

[†]These authors contributed equally.

✉ arjunvinayak.chikkankod@tudublin.ie (A. V. Chikkankod); luca.longo@tudublin.ie (L. Longo)

🆔 0000-0002-6472-1339 (A. V. Chikkankod); 0000-0002-2718-5426 (L. Longo)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

for brain dynamics analysis, its computation still needs to address numerous uncertainties and unanswered questions. Firstly, as multiple studies have pointed out, the four dominant cluster maps often represent up to 80% of the global variance in an EEG recording. However, studies conducted by Seitzman and colleagues could explain only 62% – 69% of the total variance with four cluster maps [7]. In addition, recent studies have revealed that anywhere from five to fifteen cluster maps are required to explain 80% of the global variance [7, 8, 9]. Secondly, the choice of the algorithm to generate the cluster maps is completely arbitrary [10]. Generally, modified k-means or agglomerative hierarchical clustering are used to render template maps. However, to date, clustering as a research field has produced hundreds of algorithms, each addressing specific challenges. In addition, both k-means and agglomerative hierarchical clustering are highly sensitive to noisy data and outliers. Microstate theory assumes that a finite number of non-overlapping, quasi-stable brain activation states lead to microstates. However, the problem within this theory is the lack of stability in the number of cluster maps learnt from shallow clustering (modified k-means, hierarchical clustering) and the resulting EEG microstates, which can explain up to 80% of the variance from the input EEG signal. Stability refers to the minimum variance in cluster map formation across subjects, tasks, and datasets that generalises EEG microstates. Given the above methodological gaps, this research aims to design a unified framework, leading to higher stability among the generated microstates, employing modern deep-clustering methods, expected to be robust to noisy data, outliers, and cluster morphologies. The Research Question (RQ) of our study is- To what extent do deep clustering methods improve the identification and stability of cluster maps and enhance the efficiency of microstate sequences across multiple resting and cognitive states over shallow clustering? The following section reviews relevant work in microstate theory and identifies knowledge gaps. It is then followed by section 3 presenting the design of a novel pipeline for microstate formation and analysis based on deep-clustering and notions borrowed from the field of eXplainable Artificial Intelligence (XAI)[11] for explainable microstate sequence formations. Eventually, section 4 presents the preliminary findings of a comparative experiment contrasting current shallow methods with the novel deep-clustering method under development.

2. State-of-the-art pipeline for microstate theory

The current, state-of-the-art pipeline for microstate formation and analysis is depicted in Figure 1. It is based on four steps: [A] to pre-processed EEG recordings for artefacts removal, facilitating the computations in the following steps; [B] to compute the Global Field Power (GFP) peaks; [C] to cluster the vectors with n electrode recordings at the time of each peak identified in B; the centroid of these clusters are the cluster maps or template maps. [D] to assign cluster labels for non-GFP points using backfitting procedures. Much exists in phase A for pre-processing and denoising EEG signals. We remind the readers of Makoto’s popular denoising pipeline. The Global Field Power (GFP) computation of phase B is essentially the standard deviation of all the electrodes for a given time. Formally: $GFP = \sqrt{\sum_{i=1}^N (v_i - \bar{v})^2 / N}$. The local maxima of the GFP are determined over all the EEG recordings by performing the first derivative and then annotating the zero-crossings. Phase C is the clustering of the scalp-electrode activations at the time location of the peaks of the GFP.

2. Distance-based hierarchical methods provide a hierarchy of associations among input instances. Agglomerative (bottom-up) or divisive (top-down) methods establish the hierarchical structure. The agglomerative method begins with a singleton cluster. It merges it over subsequent iterations to construct clusters with a bottom-up hierarchy. In contrast, the divisive method begins with all instances assigned to a single cluster and splits it over subsequent iterations to construct two clusters with a top-down hierarchy. AGNES (AGglomerative NESTing) and DIANA (Divisive Analysis) are the prominent agglomerative and divisive methods.
3. Density-based methods treat regions with high density as potential clusters. Moreover, such dense regions are likely to take arbitrary shapes, which makes them suitable for data with non-convex shapes. DBSCAN is a well-known density-based algorithm.
4. Grid-based methods extract grid-like configurations from each region using the grid data structure of fine granularity. Clusters are determined by computing the density of each region and then sorting the regions according to their densities. Grid-based methods are highly scalable, take minimal processing time, and are suitable for parallel process implementation. STING and CLIQUE are typical grid-based methods.
5. Probabilistic and generative models assume that the data from a particular cluster comes from a single distribution and a generative model, such as a mixture of Gaussians. A model's parameters can be estimated using the Expectation-Maximisation (EM) method, which aims to increase the maximum likelihood function on a given set of data points. Subsequently, they predict the generative probability of the underlying data points.

In microstate theory research, tools such as Cartool and those provided by EEGLAB are often used, including k-means, agglomerative clustering, MST (minimum spanning tree) based clustering, Non-Negative Matrix Factorization (NMF), ICA, and spectral clustering. Many other possibilities exist, but they are largely unexplored in the field, especially those based on deep learning. Deep clustering methods have surpassed shallow clustering methods' performance on benchmarking image datasets [10, 16]. We argue that integrating these into microstate theory research can advance the field and support enhanced microstate formation and analysis (red contribution of Figure 1C).

3. Design and methods

Following the previously identified research directions for microstate theory advancement, this section introduces an experiment where the described pipeline for microstate formation is extended with a novel deep clustering method in phase C. The raw EEG data recording needs to go to bandpass filtering using the Firwin filter, with a frequency range between 1.0 and 30.0 Hz. Subsequently, the original sampling rate, usually higher, is resampled, usually around 120 Hz. This adjustment helps balance data resolution with computational efficiency, ensuring the data is manageable while retaining sufficient detail for accurate analysis. Subsequently, the data must be referenced using an average reference method, reducing signal-related bias. Independent Component Analysis (ICA) is applied using the Infomax algorithm, configured to extract several components with a maximum number of iterations, allowing for convergence. This step is pivotal for removing eye, muscle and other artefacts from the EEG data. After identifying

and removing these components, the original signal was reconstructed and forwarded to the shallow clustering. The pre-processed EEG signal is used to construct a set of two-dimensional (2D) topographic maps, one map per point in time. 2D topographic maps are formed by projecting three-dimensional (3D) electrode information onto 2D Cartesian coordinates using an Azimuthal equidistant projection. Electrode values are interpolated for areas corresponding to non-electrode locations using the cubic method. This dataset is then used to train a deep clustering architecture, as explained in the following section.

3.1. Deep Clustering

Deep clustering identifies cluster groups using features learned by deep neural networks [10]. We designed a novel deep clustering architecture based on a self-supervised autoencoder [17] that iteratively learns representations and clusters from 2D topographic maps using a single objective function. It attempts to learn an approximation of the identity function using the weights W and bias b ($h_{W,b}(x) \approx x$). The designed AE uses clustering as an additional learning mechanism. In detail, it uses an auxiliary distribution and optimises cluster assignment by minimising the KL-divergence loss. In a nutshell, an autoencoder for deep clustering can be represented as $L_{AEDC} = L_{AE} + L_{ST}$ where L_{AE} corresponds to the *KL* loss function, whereas L_{ST} is the neighbourhood constraint clustering loss. The model utilises probabilistic outputs of the softmax function for the clustering assignments. This model, subject to reconstruction and neighbourhood constraints, systematically generates class-level explanations. The resulting explanations are then mapped to template (cluster) maps, rendering the process transparent and the results justifiable within the context of XAI principles [18, 19, 20]. After training such architecture, the emerging clusters are analysed, and their centroids are extracted and mapped to the original corresponding topographic maps, which become the template maps. Therefore, the number of template maps equates to the number of clusters. Deciding the number of clusters can follow the outcome of other research studies [4] or can be optimally established via a grid search. The final phase is backfitting, which forms a sequence of microstates, as described in section 2.

A comparative experiment is designed, as depicted in Figure 2. Here, the traditional shallow clustering method used within microstate theory research, namely a modified k-means, is contrasted with the novel application of the deep-clustering autoencoder, as designed in section 3.1. In modified k-means, the input vectors are polarity invariant, which means their absolute values are computed. The Test-Retest dataset was selected because it contains EEG recordings for various tasks: eyes open and closed, subtraction, music, and memory tasks [21]. The clustering phase (c) results are evaluated by assessing clustering effectiveness. Instead, the microstate formation phase (D) is evaluated by analysing the stability of the generated microstate sequence. Clustering effectiveness was quantified using internal clustering measures, namely the Silhouette score, the Davies-Bouldin Index (DBI), and the Calinski-Harabasz Index (CHI). Instead, the stability of the microstates was assessed through the duration parameter of each microstate. This parameter indicates how consistently each microstate is maintained over time, providing insights into the temporal stability of the microstate sequences.

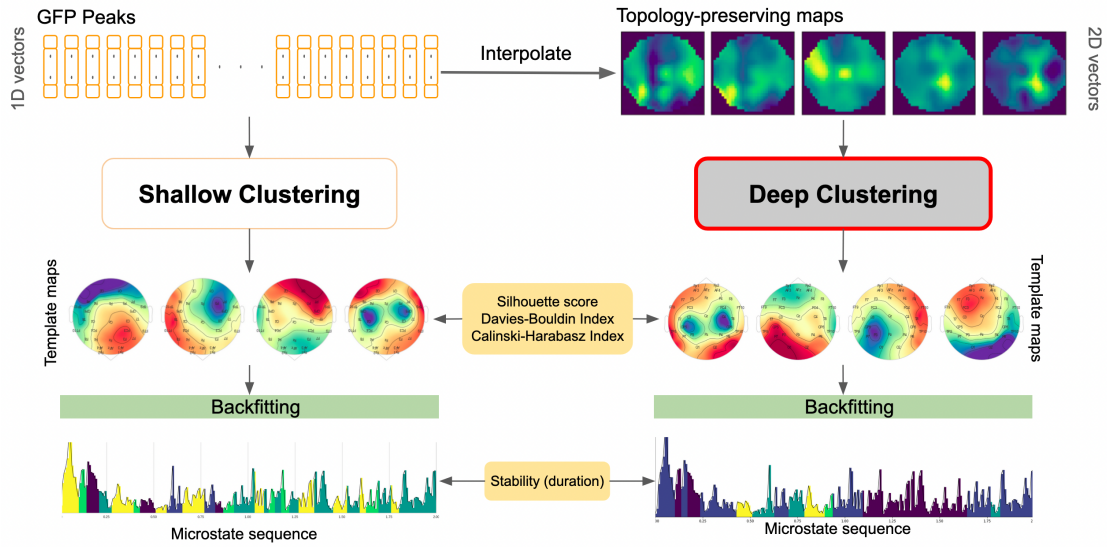


Figure 2: A diagram illustrating the comparative experiment.

4. Preliminary Results

Table 1 presents the results of clustering effectiveness. The Silhouette score ($[-1, 1]$) gauges the compactness and separation of the clusters; a higher score indicates well-defined clusters. The Davies-Bouldin Index (DBI) assesses cluster delineation, with a lower score suggesting more distinct clusters (0 being the ideal). Conversely, a higher Calinski-Harabasz Index (CHI) denotes better cluster separation and formation. Preliminary analyses reveal that deep clustering using the designed AE outperforms the Modified K-Means approach across all three metrics (Table 1). The duration metric, the mean uninterrupted time a specific microstate label is maintained, is expressed in milliseconds. This metric assesses the temporal stability of microstate sequences. In the comparative analysis, the Modified K-Means algorithm yielded a microstate duration of $80ms$, whereas the designed AE Deep clustering architecture achieved a slightly longer duration of $85.5ms$. This increase suggests that microstates generated by the deep learning method remain stable for a longer duration before transitioning to the next state [5].

Table 1

Preliminary results of clustering effectiveness for Mod K-Means and Deep Clustering.

	Mod K-Means	Deep Clustering
Silhouette score	-0.03	0.29
Davies-Bouldin Index (DBI)	33.57	1.67
Calinski-Harabasz Index (CHI)	6.44	3727.14

5. Conclusion

Our study introduces a novel deep clustering as a potential alternative to traditional shallow clustering methods for generating EEG microstates. Preliminary findings indicate that the deep clustering method, based on a convolutional autoencoder, outperforms the modified k-means in the cluster map identification and stability of the microstate, potentially leading to more robust and reliable microstate analysis. With just four microstate templates, we could effectively summarise the brain activity captured by 64 electrodes over 5 minutes. This dimensionality reduction simplifies the analysis and makes the complex spatiotemporal dynamics of brain signals more interpretable. By mapping these microstates to specific mental configurations, we can gain insights into the neural correlates of various cognitive processes. However, the evaluation was conducted on a single dataset, and further validation across diverse datasets is required to generalise the findings. The study primarily focused on resting-state EEG data, and future research should investigate the applicability of deep clustering to task-related EEG data. Future research directions include exploring different deep clustering architectures to improve microstate analysis further. Saliency and class activation maps can reveal neural patterns influencing microstate generation, offering insights into brain function and cognition.

Acknowledgments

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

- [1] D. Lehmann, H. Ozaki, I. Pal, Eeg alpha map series: brain micro-states by space-oriented adaptive segmentation, *Electroencephalography and Clinical Neurophysiology* 67 (1987) 271–288. doi:[https://doi.org/10.1016/0013-4694\(87\)90025-3](https://doi.org/10.1016/0013-4694(87)90025-3).
- [2] D. Lehmann, Multichannel topography of human alpha eeg fields, *Electroencephalography and Clinical Neurophysiology* 31 (1971) 439–449. doi:[https://doi.org/10.1016/0013-4694\(71\)90165-9](https://doi.org/10.1016/0013-4694(71)90165-9).
- [3] W. Hu, Z. Zhang, L. Zhang, G. Huang, L. Li, Z. Liang, Microstate detection in naturalistic electroencephalography data: A systematic comparison of topographical clustering strategies on an emotional database, *Frontiers in Neuroscience* 16 (2022). doi:10.3389/fnins.2022.812624.
- [4] V. Férat, M. Seeber, C. Michel, T. Ros, Beyond broadband: Towards a spectral decomposition of electroencephalography microstates, *Human Brain Mapping* 43 (2022). doi:10.1002/hbm.25834.
- [5] C. M. Michel, T. Koenig, Eeg microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review, *NeuroImage* 180 (2018) 577–593. doi:<https://doi.org/10.1016/j.neuroimage.2017.11.062>, *brain Connectivity Dynamics*.

- [6] T. Koenig, L. Prichep, D. Lehmann, P. V. Sosa, E. Braeker, H. Kleinlogel, R. Isenhardt, E. R. John, Millisecond by millisecond, year by year: normative EEG microstates and developmental stages, *Neuroimage* 16 (2002) 41–48.
- [7] B. A. Seitzman, M. Abell, S. C. Bartley, M. A. Erickson, A. R. Bolbecker, W. P. Hetrick, Cognitive manipulation of brain electric microstates, *Neuroimage* 146 (2017) 533–543.
- [8] D. D’croz-Baron, L. Bréchet, M. Baker, K. Tanja, Auditory and visual tasks influence the temporal dynamics of eeg microstates during post-encoding rest, *Brain Topography* 34 (2021) 3. doi:10.1007/s10548-020-00802-4.
- [9] A. Custo, D. Van De Ville, W. M. Wells, M. I. Tomescu, D. Brunet, C. M. Michel, Electroencephalographic resting-state networks: Source localization of microstates, *Brain Connectivity* 7 (2017) 671–682. doi:10.1089/brain.2016.0476, PMID: 28938855.
- [10] S. Zhou, H. Xu, Z. Zheng, J. Chen, Z. Li, J. Bu, J. Wu, X. Wang, W. Zhu, M. Ester, A comprehensive survey on deep clustering: Taxonomy, challenges, and future directions, *ArXiv abs/2206.07579* (2022).
- [11] L. Longo, et al., Explainable artificial intelligence (xai) 2.0: A manifesto of open challenges and interdisciplinary research directions, *Information Fusion* 106 (2024) 102301. doi:https://doi.org/10.1016/j.inffus.2024.102301.
- [12] J. Han, M. Kamber, J. Pei, *Data mining concepts and techniques*, third edition (2012).
- [13] U. Lal, A. V. Chikkankod, L. Longo, A comparative study on feature extraction techniques for the discrimination of frontotemporal dementia and alzheimer’s disease with electroencephalography in resting-state adults, *Brain Sciences* 14 (2024). doi:10.3390/brainsci14040335.
- [14] U. Lal, A. V. Chikkankod, L. Longo, Fractal dimensions and machine learning for detection of parkinson’s disease in resting-state electroencephalography, *Neural Comput. Appl.* (2024).
- [15] D. Xu, Y. Tian, A comprehensive survey of clustering algorithms, *Annals of Data Science* 2 (2015) 165–193. doi:10.1007/s40745-015-0040-1.
- [16] Y. Li, P. Hu, Z. Liu, D. Peng, J. Zhou, X. Peng, *Contrastive clustering*, 2020.
- [17] A. V. Chikkankod, L. Longo, On the dimensionality and utility of convolutional autoencoders latent space trained with topology-preserving spectral eeg head-maps, *Machine Learning and Knowledge Extraction* 4 (2022) 1042–1064. doi:10.3390/make4040053.
- [18] X. Guo, X. Liu, E. Zhu, J. Yin, *Deep clustering with convolutional autoencoders*, 2017, pp. 373–382. doi:10.1007/978-3-319-70096-0_39.
- [19] C. A. Ellis, R. L. Miller, V. D. Calhoun, Improving explainability for single-channel eeg deep learning classifiers via interpretable filters and activation analysis*, in: *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2023, pp. 2474–2481. doi:10.1109/BIBM58861.2023.10385647.
- [20] I. Hussain, R. Jany, R. Boyer, A. Azad, S. A. Alyami, S. J. Park, M. M. Hasan, M. A. Hossain, An explainable eeg-based human activity recognition model using machine-learning approach and lime, *Sensors* 23 (2023). URL: <https://www.mdpi.com/1424-8220/23/17/7452>. doi:10.3390/s23177452.
- [21] Y. Wang, W. Duan, D. Dong, L. Ding, X. Lei, A test-retest resting and cognitive state eeg dataset (2022). doi:10.18112/openneuro.ds004148.v1.0.1.