# Interpretable EEG Microstate Discovery via Variational Deep Embedding: A Systematic Architecture Search with Multi-Quadrant Evaluation

Saheed Faremi[1,2,*,†], Andrea Visentin[1,3,†] and Luca Longo[1,2,†]

[1]School of Computer Science and Information Technology, University College Cork, Western Road, Cork, T12 XF62, Ireland
[2]Artificial Intelligence and Cognitive Load Research Lab, University College Cork, Ireland
[3]Insight RI Research Centre for Data Analytics, University College Cork, Ireland

**Abstract**

EEG microstate analysis segments continuous brain electrical activity into brief, quasi-stable topographic configurations that reflect discrete functional brain states. Conventional approaches such as Modified K-Means operate directly in electrode space with hard assignment, offering no learned latent representation, no generative decoder, and no mechanism to decode latent configurations into verifiable scalp topographies, limiting both model transparency and interpretability. To address this, we present a Convolutional Variational Deep Embedding (Conv-VaDE) model that jointly learns topographic reconstruction and probabilistic soft clustering in a shared latent space. Conv-VaDE enables generative decoding of cluster prototypes into verifiable scalp topographies, replacing opaque hard partitioning with probabilistic soft assignment. A polarity invariance scheme and a four-dimensional grid search over cluster count ($K \in [3, 20]$), latent dimensionality, network depth, and channel width are conducted to systematically reveal how each architectural design choice shapes the quality, stability, and interpretability of learned EEG microstate representations. The model is evaluated on the LEMON resting-state eyes-closed EEG dataset with ten participants using topographic template formation, clustering stability, and global explained variance (GEV). The architecture search reveals that depth $L = 4$ appears consistently across all 18 best-performing configurations, yielding a best-case GEV of 0.730 and a silhouette of 0.229 at $K = 4$ across the model sweeps, where moderately deep networks with compact channel widths and small latent dimensionality dominate across the full $K$ range. These results establish that principled architecture search, rather than model scale, is the key to interpretable and stable EEG microstate discovery via variational deep embedding.

**Keywords**

EEG microstates, Variational autoencoder, Gaussian mixture model, Polarity invariance, Hyperparameter sweep, Clustering validation, Mechanistic Interpretability

## 1. Introduction

EEG microstates are brief, quasi-stable spatial distributions of scalp electric potential, typically lasting 60–120 ms, that partition continuous brain activity into a sequence of discrete functional states [1, 2]. Lehmann and colleagues established that microstate temporal statistics, including mean duration, occurrence rate, transition probabilities, and fractional coverage, are sensitive biomarkers, with subsequent work linking them to cognitive processing and neuropsychiatric conditions such as schizophrenia and depression [3, 4, 2].

Modified K-Means (ModKMeans), the predominant clustering algorithm for microstate analysis, assigns each time point to the nearest microstate class based on spatial correlation, treating opposite-sign maps as equivalent. This hard assignment discards information about transitional or ambiguous regions where the scalp topography may genuinely reflect a mixture of two or more underlying microstates. Furthermore, the convention of fixing $K = 4$ microstates, while grounded in early group-level findings, may not capture the full complexity of individual brain dynamics; the optimal cluster count is known to

*Late-breaking work, Demos and Doctoral Consortium, colocated with the 4th World Conference on eXplainable Artificial Intelligence: July 01–03, 2026, Fortaleza, Brazil*

*Corresponding author.
†These authors contributed equally.
126100912@umail.ucc.ie (S. Faremi); andrea.visentin@ucc.ie (A. Visentin); luca.longo@ucc.ie (L. Longo)
0009-0007-8428-7019 (S. Faremi); 0000-0003-3702-4826 (A. Visentin); 0000-0002-2718-5426 (L. Longo)

vary across datasets, recording conditions, and clinical populations [1]. Critically, ModKMeans operates directly in electrode space with no learned representation and no decoder, offering no mechanism to reconstruct or verify the cluster centres it discovers as scalp topographies. Recent deep learning architectures applied to EEG microstates [5, 6, 7] address some of these shortcomings by learning nonlinear mappings that enable reconstruction of cluster centres as scalp topographies. However, these approaches either rely on post-hoc clustering of a frozen latent space or frame microstates as supervised classification, and the challenge of determining the optimal number of microstates remains unresolved.

This gap motivates a generative approach that couples a learned decoder with a probabilistic clustering mechanism capturing assignment uncertainty. Variational Deep Embedding (VaDE) [8] combines a variational autoencoder with a Gaussian mixture model prior to jointly learn structured representations and probabilistic cluster assignments. The present study extends VaDE to EEG microstate analysis through a Convolutional Variational Deep Embedding (Conv-VaDE) model and investigates how cluster count, latent dimensionality, network depth, and channel width affect the quality, stability, and interpretability of learned microstate templates in a systematic four-dimensional grid search.

## 2. Related Work

*Conventional microstate analysis.* Microstate analysis was formalised by Lehmann et al. [2] and systematised by Michel and Koenig [1], who established the pipeline of bandpass filtering, Global Field Power (GFP) peak extraction, spatial clustering, and temporal label assignment that underpins contemporary microstate research. Four canonical classes (usually referred to as A–D) were identified through large-scale normative studies [9, 10] and remain the standard reference. Modified K-Means and Atomize and Agglomerate Hierarchical Clustering are the predominant clustering algorithms [11], both operating in electrode space with hard assignments, whereby each time point is assigned to exactly one microstate. These methods achieve polarity invariance through their distance metric, typically the absolute spatial correlation, rather than through the representation itself; a scalp topography and its sign-inverted counterpart are treated as representing the same microstate during assignment, meaning the algorithm handles polarity implicitly without explicitly encoding this invariance into the model. Clustering quality is conventionally assessed via Global Explained Variance (GEV), which measures how well the assigned microstate templates account for the observed scalp topography [1]. The standard analysis pipeline applies Modified K-Means (ModKMeans), which fits prototype scalp voltage distributions by clustering multichannel EEG voltage vectors sampled at GFP peaks, where spatial configurations are most stable. The resulting microstate templates are then projected back onto the full continuous EEG recording, assigning every time point to its most correlated microstate map, thereby capturing the temporal extent of each microstate beyond the isolated GFP peak. While four canonical microstate classes are widely reported in the literature, this number reflects an empirical convention rather than a constraint of the algorithm itself [11, 12].

*Deep learning for EEG microstates.* Deep learning architectures have been explored across multiple aspects of EEG microstate research. Thukral et al. [7] trained a convolutional autoencoder on EEG topographic maps and clustered the latent space with ModKMeans, improving silhouette and Davies-Bouldin scores over direct electrode-space clustering. Zhao et al. [5] framed microstate identification as supervised sequence-to-sequence classification, achieving 74.26% accuracy. Sikka et al. [6] developed an LSTM-based autoencoder to capture complex temporal dependencies in EEG data, using the learned representations to model microstate transitions across the continuous recording. These approaches rely on post-hoc clustering procedures applied outside the network, and without a generative decoder, the internal representations cannot be reconstructed, leaving the discovered microstate templates as a black box. Additionally, none of them enforces polarity invariance during training. The absence of polarity invariance risks creating redundant clusters that split a single brain state along an irrelevant sign axis.

*Deep clustering and variational methods.* Deep clustering methods that jointly optimise representation learning and clustering within a single network offer a more coherent alternative to post-hoc clustering, where representation learning and clustering are treated as separate tasks. Deep Embedded Clustering

(DEC) [13] jointly optimises representation learning and clustering but lacks a decoder, meaning it only optimises the clustering objective, leaving no room to visualise or inspect the scalp topographies. Its improved variant, Improved Deep Embedded Clustering (IDEC) [14], addresses this by adding a reconstruction loss and a decoder, but the decoder is deterministic rather than generative, preventing probabilistic assignment of EEG segments to microstate templates. Variational Deep Embedding (VaDE) [8] addresses this by embedding a Gaussian Mixture Model (GMM) prior, a probabilistic model that assumes data is generated from a mixture of $K$ Gaussian distributions, into the VAE framework, optimising a mixture-augmented evidence lower bound that trains encoder, decoder, and GMM parameters end-to-end. This is distinct from two-stage pipelines that cluster a frozen latent space post hoc, because the clustering objective shapes latent geometry during learning. Simple, Scalable, and Stable Variational Deep Clustering (S3VDC) extended VaDE with interleaved Kullback-Leibler divergence (KL) regularisation during pretraining to prevent posterior collapse, a failure mode where the approximate posterior ignores the latent code and the decoder generates from the prior alone [15, 16]. Fu et al. [15] showed that cyclical annealing of the KL weight mitigates this collapse by allowing the encoder to learn informative representations before the full KL penalty is applied.

*Explainability perspective.* From an explainability perspective, the distinction between intrinsic and post-hoc methods is central to this work [17]. Post-hoc approaches such as LIME [18], which explains individual predictions by fitting a simple interpretable model locally, and SHAP [19], which assigns each input feature a contribution value based on cooperative game theory, explain an already-trained model by approximating its decision boundary locally, but require a separate attribution step and do not guarantee faithfulness to the model's internal reasoning. Intrinsic explainability, by contrast, is built into the model architecture: the explanation is the model's own output rather than an external approximation. In the context of generative models, VAE decoders provide intrinsic explainability by mapping latent codes to data space [20]. Systematic architecture evaluation complements this by revealing how design choices affect model behaviour, contributing to model transparency in the sense of Lipton's taxonomy [21].

## 3. Methodology

This study adopts a secondary research approach, utilising the existing LEMON resting-state EEG dataset to construct and empirically evaluate a novel Conv-VAE-GMM architecture.[1] The hypothesis is: if a VaDE-based Conv-VaDE model with a GMM prior is trained on resting-state EEG data and subjected to a systematic four-dimensional architecture search over cluster count, latent dimensionality, network depth, and channel width, then moderately deep architectures with compact channel widths will consistently dominate the search space, yielding the most stable and interpretable microstate templates as evidenced by higher silhouette scores, lower Davies-Bouldin indices, higher Calinski-Harabasz indices, and higher GEV, while the generative decoder will enable direct reconstruction of each GMM component centre into a verifiable scalp topography.

*Data Understanding and Preparation.* The LEMON dataset [22] comprises eyes-closed resting-state EEG at $f_s = 250$ Hz, 61 channels, 10–10 montage. Ten participants were selected from the dataset. A common average reference and zero-phase FIR bandpass filter (2–20 Hz) preserves temporal alignment and prevents phase-induced microstate boundary blurring. GFP peaks are extracted via Pycrostates [23] (min. inter-peak distance: 3 samples) and projected onto $40 \times 40$ topographic images via azimuthal equidistant projection and cubic interpolation [24]. Z-score normalisation with $\pm 5\sigma$ clipping is computed on the training split only: $x_{\text{norm}} = \text{clip}((x - \bar{x}_{\text{train}})/\sigma_{\text{train}}, -5, +5)$. A 90/10 train/evaluation split maximises training data; the evaluation set is left unaugmented.

*Model Architecture.* The encoder maps $\mathbf{x} \in \mathbb{R}^{1\times 40\times 40}$ through four `Conv2d` layers ($\texttt{ndf} \to 2\texttt{ndf} \to 4\texttt{ndf} \to 8\texttt{ndf}$) with batch normalisation, LeakyReLU ($\alpha = 0.2$), and dropout ($p = 0.2$), then adaptive average pooling and FC layers producing $(\boldsymbol{\mu}_\phi, \log \boldsymbol{\sigma}^2_\phi) \in \mathbb{R}^{d_z}$ via reparameterisation [25]. The decoder mirrors this with `ConvTranspose2d` and a linear final activation. The GMM prior [8]

[1] Code and sweep results are available at https://github.com/fayisode/microstate-architecture-search.

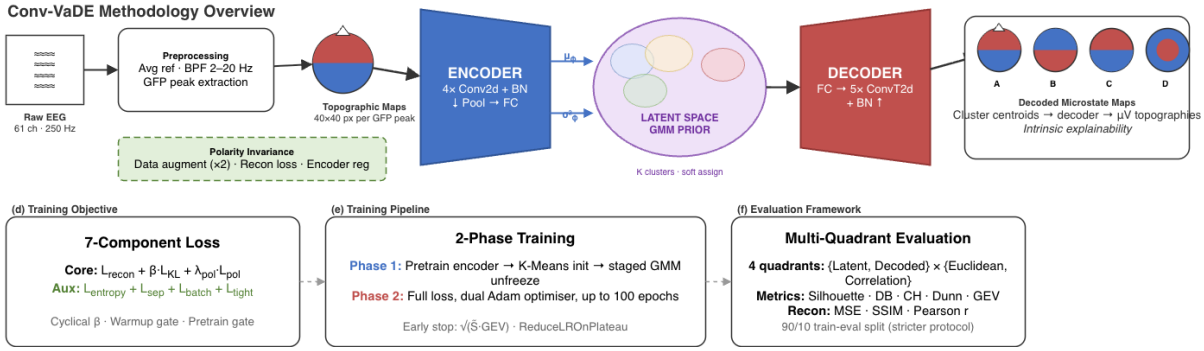



**Figure 1:** Conv-VaDE processes EEG recordings into topographic maps and clusters them using a convolutional encoder with a GMM-structured latent space, trained via a seven-component objective and evaluated across multiple representation and distance metrics.

$p(\mathbf{z}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_k^{(c)}, \mathrm{diag}(\boldsymbol{\sigma}_k^{2(c)}))$ is optimised end-to-end. The search varies $K \in \{3, \ldots, 20\}$, $d_z \in \{16, 32, 64\}$, depth $\in \{2, 3, 4\}$, $\mathtt{ndf} \in \{32, 64, 128\}$, yielding 486 unique configurations evaluated across 10 subjects.

*Polarity Invariance and Training Objective.* Polarity invariance ($\mathbf{x}$ and $-\mathbf{x}$ are the same state) is enforced at three levels: (1) sign-flip augmentation of the training set, with GMM initialisation on the unaugmented subset; (2) polarity-invariant loss $\mathcal{L}_{\text{recon}} = \min(\text{MSE}(\hat{\mathbf{x}}, \mathbf{x}), \text{MSE}(\hat{\mathbf{x}}, -\mathbf{x})) \cdot D_x$, $D_x = 1600$; (3) encoder regularisation $\mathcal{L}_{\text{pol}} = \text{MSE}(\boldsymbol{\mu}_\phi(\mathbf{x}), \boldsymbol{\mu}_\phi(-\mathbf{x}))$, $\lambda_{\text{pol}} = 0.1$. The total loss is $\mathcal{L} = \mathcal{L}_{\text{recon}} + \beta\mathcal{L}_{\text{KL}} + \lambda_e\mathcal{L}_{\text{entropy}} + \lambda_s\mathcal{L}_{\text{sep}} + \lambda_b\mathcal{L}_{\text{batch}} + \lambda_t\mathcal{L}_{\text{tight}} + \lambda_{\text{pol}}\mathcal{L}_{\text{pol}}$, where $\mathcal{L}_{\text{entropy}}$ ($\lambda_e = 0.3/\log K$) prevents weight collapse; $\mathcal{L}_{\text{sep}}$ ($\lambda_s = 50.0$) penalises topographic redundancy via mean pairwise $|\text{Pearson}\, r|$; $\mathcal{L}_{\text{batch}}$ ($\lambda_b = 5.0$) enforces uniform cluster usage; $\mathcal{L}_{\text{tight}}$ ($\lambda_t = 0.2$) pulls samples toward assigned cluster priors [16]. Cyclical $\beta$-annealing ($\beta_{\max} = 0.1$, 4 cycles) [15] modulates KL pressure; auxiliary losses are zeroed for 3 epochs then linearly ramped over 10.

*Training Procedure.* Pretraining runs 200 steps ($\beta = 10^{-3}$, no auxiliary losses), followed by GMM initialisation via bisecting K-Means on unaugmented latent means (flat for $K \leq 4$, iterative bisection for $K > 4$), prior frozen for 5 epochs, dead clusters ($< 1\%$) reinitialised from the largest surviving cluster. Main training runs up to 100 epochs with dual Adam optimisers (encoder/decoder: $\eta = 10^{-3}$, $\lambda_{L2} = 10^{-5}$; GMM: $\eta = 5 \times 10^{-4}$), gradient clipping at 5.0 and 1.0, and early stopping on $\sqrt{\tilde{S}} \cdot \text{GEV}$ ($\tilde{S} = (\text{sil} + 1)/2$), with ReduceLROnPlateau (patience $= 10$).

*Evaluation.* Clustering quality is assessed in the latent space via Silhouette, Davies–Bouldin, Calinski–Harabasz, and Dunn indices. Reconstruction quality is measured with MSE, SSIM (range $= 10.0$), and spatial correlation. Topographic fidelity is evaluated via GEV computed through backfitting: $\text{GEV} = \sum_t \text{GFP}^2(t)\rho^2(t) / \sum_t \text{GFP}^2(t)$. For each $K$, the best configuration is selected by the highest mean GEV averaged across all 10 subjects.

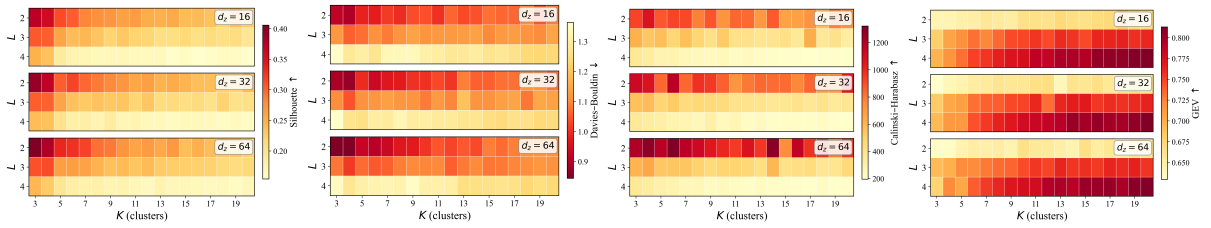
## 4. Results and Discussion

Table 1 reports the best configuration per $K$ from the architecture sweep, selected by highest mean GEV across 10 subjects. Figures 2 and 3 visualise Q1 clustering metric landscapes and architecture parameter effects across the sweep space. Figures 4 and 5 demonstrate the explainability of Conv-VaDE through decoded cluster centroids, latent space structure, temporal clustering, and cluster distribution analysis. At $K = 4$, the best architecture ($d_z = 16$, $L = 4$, $n_f = 32$) achieves a silhouette of $0.229 \pm 0.039$ and a Davies–Bouldin index of $1.28 \pm 0.13$, indicating well-separated and compact clusters in the learned latent space, with a global explained variance of $0.730 \pm 0.072$ confirming strong topographic representational fidelity in electrode space.

*Explainability of Decoded Centroids.* The core explainability claim of Conv-VaDE is that GMM cluster centres can be decoded into inspectable scalp topographies. Figure 4 shows the four decoded centroids

**Table 1**
Architecture sweep: best configuration per $K$ by GEV (held-out 10% split, $n = 10$ subjects). Q1 Sil = latent Euclidean silhouette.

| $K$ | $d_z$ | Depth | ndf | Q1 Sil ↑ | GEV ↑ | $K$ | $d_z$ | Depth | ndf | Q1 Sil ↑ | GEV ↑ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 16 | 4 | 32 | *0.234* | 0.710 | 12 | 64 | 4 | 32 | 0.162 | 0.803 |
| 4 | 16 | 4 | 32 | 0.229 | 0.730 | 13 | 16 | 4 | 32 | 0.172 | 0.806 |
| 5 | 16 | 4 | 32 | 0.195 | 0.741 | 14 | 32 | 4 | 32 | 0.160 | 0.814 |
| 6 | 16 | 4 | 32 | 0.165 | 0.760 | 15 | 16 | 4 | 32 | 0.161 | 0.810 |
| 7 | 16 | 4 | 32 | 0.171 | 0.771 | 16 | 16 | 4 | 32 | 0.150 | 0.817 |
| 8 | 16 | 4 | 32 | 0.167 | 0.777 | 17 | 64 | 4 | 32 | 0.164 | 0.813 |
| 9 | 32 | 4 | 32 | 0.186 | 0.786 | 18 | 16 | 4 | 32 | 0.161 | 0.817 |
| 10 | 16 | 4 | 32 | 0.162 | 0.792 | 19 | 16 | 4 | 32 | 0.165 | 0.817 |
| 11 | 16 | 4 | 32 | 0.158 | 0.798 | 20 | 16 | 4 | 32 | 0.157 | *0.821* |




(a) Silhouette (b) Davies–Bouldin (c) Calinski–Harabasz (d) GEV

**Figure 2:** Q1 (latent Euclidean) clustering metric landscapes across the $K \times L \times d_z$ sweep space. Each heatmap panel is faceted by $d_z$, with rows showing depth $L$ and columns showing cluster count $K$; $n_f$ is averaged. Colour intensity encodes metric magnitude.

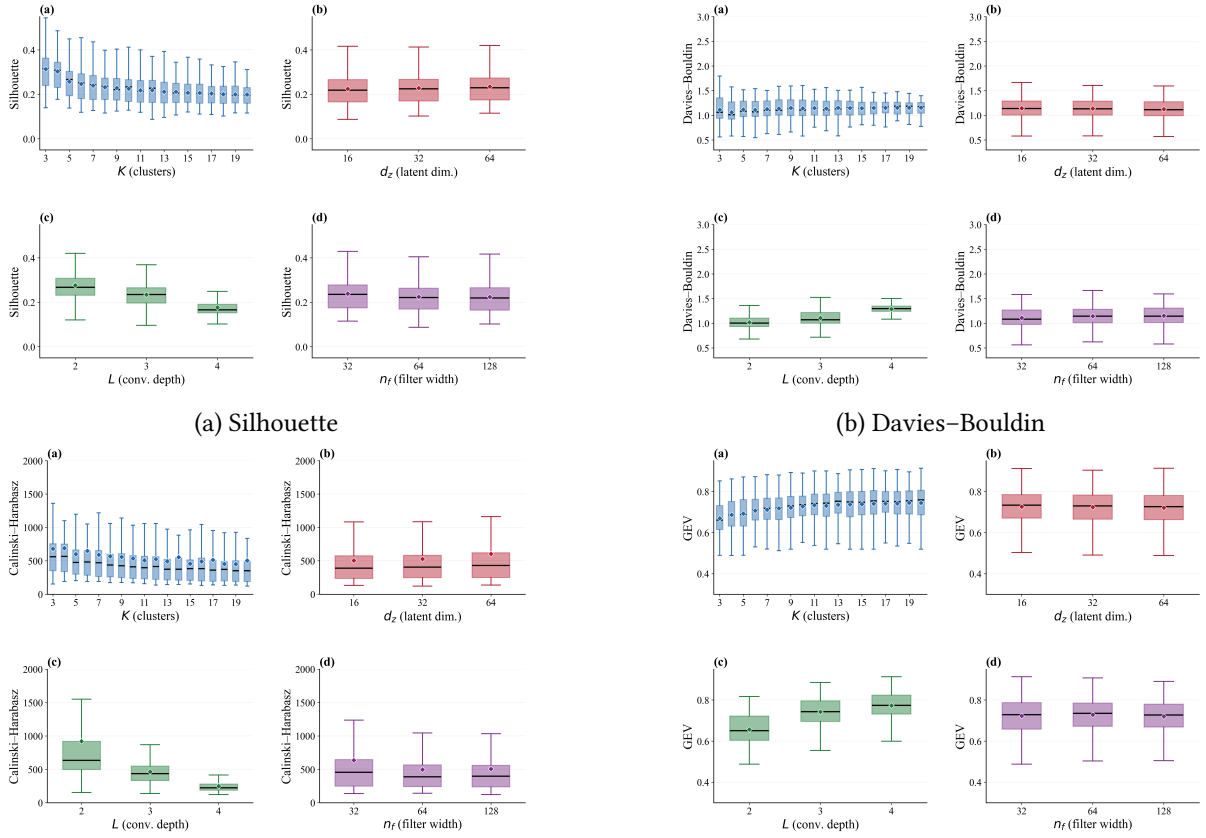
at $K = 4$, each rendered as a circular scalp map in the original $40 \times 40$ pixel grid. The topographies reveal distinct spatial patterns consistent with known microstate classes: anterior-posterior gradients and lateral asymmetries are clearly visible, confirming that the decoder has learned a meaningful mapping from latent cluster centres to physiologically plausible scalp distributions.

Figure 5 provides further explainability evidence. The centroid correlation matrix (a) shows moderate-to-high absolute correlations ($|r| = 0.54$–$0.92$), indicating shared global topographic structure while signed differences capture polarity contrasts. The PCA projection (b) confirms that the GMM prior imposes separable cluster structure in the first two principal components. The temporal clustering (c) reveals characteristic rapid alternation between states consistent with the 60–120 ms durations reported in the literature [1]. The approximately balanced cluster distribution (d) confirms that batch entropy regularisation prevents mode collapse.

*Discussion.* The multi-quadrant framework reveals an asymmetry a single metric would obscure. ModKMeans templates are native electrode vectors optimised for spatial correlation, giving a structural advantage on GEV. Conv-VaDE decoded centroids must transit from the $40 \times 40$ pixel grid back to 61 electrodes before GEV is computed, attenuating the correlation that GEV weights quadratically. Comparing Q4 places both methods on equal footing in decoded topographic space; Q1 and Q2 characterise latent quality that ModKMeans cannot provide.

The polarity scheme addresses a failure visible at high $K$: without it, the GMM allocates separate components to each polarity, halving the useful cluster count. Q1 silhouette generally declines from 0.234 at $K = 3$ to 0.157 at $K = 20$ (Table 1), though the decline is not strictly monotonic, indicating that latent cluster separation degrades as the number of states increases. The polarity scheme eliminates mirror-image duplicates but does not prevent semantic splitting of related brain states.

The architecture sweep fills a gap absent from prior work: optimal $d_z$, depth, and channel width for topographic EEG data cannot be assumed from image benchmarks. Depth $L = 4$ dominates all 18




(a) Silhouette

(b) Davies–Bouldin

(c) Calinski–Harabasz

(d) GEV

**Figure 3:** Architecture parameter effects on Q1 (latent Euclidean) clustering metrics. Each panel shows the marginal effect of one sweep dimension ($d_z$, $L$, $n_f$) on a given metric, aggregated across all $K$ values.

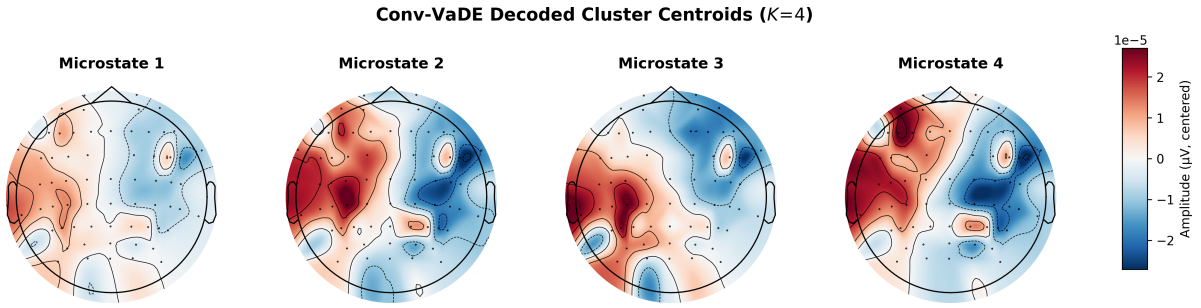
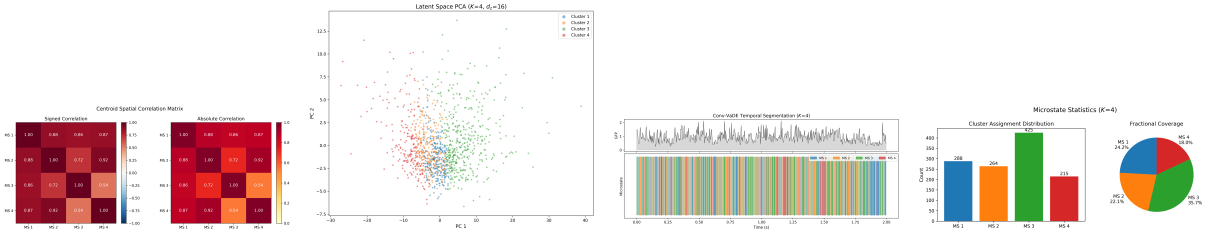



**Figure 4:** Decoded GMM cluster centroids at $K=4$ ($d_z=16$, $L=4$, $n_f=32$, subject 010012). Each panel shows the decoder output $\hat{\mathbf{x}}_k = g_\theta(\boldsymbol{\mu}_k^{(c)})$ for cluster $k$, rendered as a circular scalp topography. Colour encodes z-scored amplitude (RdBu_r).

top-ranked configurations across every $K$ value. Latent dimensionality shows low sensitivity: $d_z=16$ dominates in 14 of 18 best configurations across the full $K$ range, with $d_z \in \{32, 64\}$ appearing only at $K \in \{9, 12, 14, 17\}$, suggesting that the $40\times40$ topographic input is sufficiently represented by a compact latent space regardless of cluster count. Channel width $n_f = 32$ suffices in all 18 best configurations, confirming that the $40\times40$ topographic input does not benefit from wider convolutional filters.




(a) Centroid correlation (b) Latent PCA (c) Segmentation (d) Cluster distribution

**Figure 5:** Explainability analysis at $K = 4$ (subject 010012). (a) Signed Pearson correlation between decoded centroids. (b) PCA of the 16D latent space coloured by GMM assignment. (c) GFP trace and microstate ribbon showing temporal clustering. (d) Cluster sample counts and fractional coverage.

## 5. Conclusion

This paper presented Conv-VaDE, a convolutional variational deep embedding model for explainable EEG microstate analysis that jointly learns topographic reconstruction and probabilistic soft clustering. Decoded GMM cluster centres produce inspectable scalp topographies without post-hoc processing (Figure 4), and the structured latent space enables direct visualisation of cluster geometry (Figure 5). Three contributions were introduced: a three-level polarity invariance scheme eliminating redundant clusters; a seven-component objective preventing mode collapse across $K \in [3, 20]$; and a four-dimensional architecture sweep over 486 configurations evaluated through a multi-quadrant metric framework on 10 subjects. Population-level generalisation requires validation on larger cohorts. Future work targets latent-space backfitting to close the GEV gap, multi-subject generalisation, and temporal transition modelling. The source code, trained model configurations, and sweep results are publicly available at https://github.com/fayisode/microstate-architecture-search.

## Acknowledgments

This research was conducted at the Artificial Intelligence and Cognitive Load Research Lab, University College Cork, and supported by Research Ireland - Taighde Éireann, under Grant Nos. 18/CRT/6223 and 12/RC/2289-P2, co-funded by the European Regional Development Fund. 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.

## Declaration on Generative AI

The authors used Grammarly to help with grammar check, spelling correction, and stylistic clarity during the preparation of this manuscript.

## References

[1] 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.

[2] D. Lehmann, R. D. Pascual-Marqui, C. Michel, EEG microstates, Scholarpedia 4 (2009) 7632. Revision #88985.

[3] L. Bréchet, C. M. Michel, EEG microstates in altered states of consciousness, Frontiers in Psychology 13 (2022) 856697.

[4] M. Murphy, A. E. Whitton, S. Deccy, M. L. Ironside, A. Rutherford, M. Beltzer, M. Sacchet, D. A. Pizzagalli, Abnormalities in electroencephalographic microstates are state and trait markers of major depressive disorder, Neuropsychopharmacology 45 (2020) 2030–2037.

[5] Q. Zhao, K. Cui, L. Zhang, Z. Wu, H. Jiang, M. Zhao, B. Hu, Identifying and predicting EEG microstates with sequence-to-sequence deep learning models for online applications, Journal of Neural Engineering 22 (2025) 036040.
[6] A. Sikka, H. Jamalabadi, M. Krylova, S. Alizadeh, J. N. van der Meer, L. Danyeli, M. Deliano, P. Vicheva, T. Hahn, T. Koenig, D. R. Bathula, M. Walter, Investigating the temporal dynamics of electroencephalogram (EEG) microstates using recurrent neural networks, Human Brain Mapping 41 (2020) 2334–2346.
[7] S. Thukral, B. Raufi, B. Božic, Convolutional autoencoder-based dimensionality reduction for eeg microstate analysis, in: IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, 2025, pp. 69–82.
[8] Z. Jiang, Y. Zheng, H. Tan, B. Tang, H. Zhou, Variational deep embedding: An unsupervised and generative approach to clustering, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI), 2017, pp. 1965–1972.
[9] T. Koenig, D. Lehmann, M. C. Merlo, K. Kochi, D. Hell, M. Koukkou, A deviant EEG brain microstate in acute, neuroleptic-naive schizophrenics at rest, European Archives of Psychiatry and Clinical Neuroscience 249 (1999) 205–211.
[10] T. Koenig, L. Prichep, D. Lehmann, P. V. Sosa, E. Braeker, H. Kleinlogel, R. Isenhart, E. John, Millisecond by millisecond, year by year: Normative eeg microstates and developmental stages, NeuroImage 16 (2002) 41–48.
[11] A. T. Poulsen, A. Pedroni, N. Langer, L. K. Hansen, Microstate eeglab toolbox: An introductory guide, bioRxiv (2018).
[12] F. von Wegner, P. Knaut, H. Laufs, EEG microstate sequences from different clustering algorithms are information-theoretically invariant, Frontiers in Computational Neuroscience 12 (2018) 70.
[13] J. Xie, R. Girshick, A. Farhadi, Unsupervised deep embedding for clustering analysis, in: International Conference on Machine Learning, 2016, pp. 478–487.
[14] X. Guo, L. Gao, X. Liu, J. Yin, Improved deep embedded clustering with local structure preservation, in: International Joint Conference on Artificial Intelligence, 2017, pp. 1753–1759.
[15] H. Fu, C. Li, X. Liu, J. Gao, A. Celikyilmaz, L. Carin, Cyclical annealing schedule: A simple approach to mitigating KL vanishing, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 2019, pp. 240–250.
[16] L. Cao, S. Asadi, W. Zhu, C. Schmidli, M. Sjöberg, Simple, scalable, and stable variational deep clustering, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2020, pp. 108–124.
[17] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence 1 (2019) 206–215.
[18] M. T. Ribeiro, S. Singh, C. Guestrin, "why should i trust you?": Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.
[19] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems, volume 30, 2017.
[20] P. Cristovao, H. Nakada, Y. Tanimura, H. Asoh, Generating in-between images through learned latent space representation using variational autoencoders, IEEE Access 8 (2020) 149456–149467.
[21] Z. C. Lipton, The mythos of model interpretability, Queue 16 (2018) 31–57.
[22] A. Babayan, M. Erbey, D. Kumral, J. D. Reinelt, A. M. Reiter, J. Röbbig, H. L. Schaare, M. Uhlig, A. Anwander, P.-L. Bazin, et al., A mind-brain-body dataset of mri, eeg, cognition, emotion, and peripheral physiology in young and old adults, Scientific Data 6 (2019) 1–21.
[23] V. Férat, M. Scheltienne, D. Brunet, T. Ros, C. Michel, Pycrostates: a python library to study eeg microstates, Journal of Open Source Software 7 (2022) 4564.
[24] T. Ahmed, L. Longo, Examining the size of the latent space of convolutional variational autoencoders trained with spectral topographic maps of eeg frequency bands, IEEE Access 10 (2022) 107575–107586.
[25] D. P. Kingma, M. Welling, Auto-encoding variational bayes, 2022. arXiv:1312.6114.