**Exploring EEG Features Structure for Neuroscreening: A Study of Dimensionality Reduction Techniques**

Maja Marzec1

1Faculty of Physics UW, Ludwika Pasteura 5, 02-093 Warszawa
*me.marzec3*@*student.uw.edu.pl*

**Keywords**: machine learning, dimensionality reduction, feature engineering, interpretability,

XAI, PCA, ICA, UMAP

1. **Introduction**

Electroencephalography (EEG) is a cost-effective, non-invasive tool for monitoring brain activity, but the resulting signals are high-dimensional and noisy, making clinical interpretation challenging [1]. In computational medicine, particularly in neuroscreening as a rapid diagnostic aid, there is a need for automated models that reduce data complexity while preserving interpretability. This ensures healthcare professionals’ trust in the results and optimizes both computational cost and processing time. Dimensionality reduction (DR) offers a promising solution by extracting compact latent representations from EEG-derived features without compromising clinical transparency.

1. **Description of the problem**

This study investigates how different dimensionality reduction (DR) techniques influence the classification of normal versus pathological EEGs. DR was applied to hand-crafted features - covariance matrices, power spectral densities (PSD), and coherence measures, computed at both the local (6-second frame-level) and global (recording-level, median-aggregated by frames) scales. This dual representation captures overall signal trends as well as short-term variations. Each feature vector consists of 2850 elements per frame/ recording, which means that training effective models on such high-dimensional data typically requires a large number of training samples [2]. Properly applied dimensionality reduction can make it feasible to train models on smaller EEG datasets, which is particularly valuable in biomedical applications where data are scarce and difficult to obtain. The main challenge is to achieve computational efficiency and interpretability while preserving diagnostic accuracy comparable to that obtained with the full original feature set.

1. **Related work**

The dataset, Gradient-Boosted Ensemble (GBE) model, and original features are sourced from [2]. Principal Component Analysis (PCA) is commonly used for dimensionality reduction, while Independent Component Analysis (ICA) is primarily applied in EEG analysis for artifact removal [3]. Uniform Manifold Approximation and Projection (UMAP) is used for visualizing high-dimensional data because it preserves the topological structure when projecting into lower dimensions, such as 2D or 3D. Investigating ICA and UMAP for dimensionality reduction of input features in classification models represents an innovative approach that has not been previously explored.

1. **Solution to the problem**

The implemented framework combines systematic component selection, classification, and interpretability. First, a variance-based Meaningful Components Selection criterion was applied, retaining components until 95% of cumulative explained variance was captured. This ensured DR without neglecting diagnostically relevant information. Next, to enhance clinical transparency, an innovative Selective Activation algorithm was developed. The Selective Activation algorithm back-projects variance-preserving components onto original EEG feature spaces, enabling temporal and spectral mapping that supports physiological interpretation [5]. Finally, classification performance was evaluated using a Gradient-Boosted Ensemble (GBE) model trained on (i) original scaled features, (ii) PCA-reduced features, and (iii) ICA-reduced features.



**Fig1. Workflow chart summarizing the EEG Feature Extraction + DR + Interpretability + ML pipeline.**

1. **Conclusions and future work**

With variance-based component selection, dimensionality reduction [preserved](https://drive.google.com/file/d/19R8aC6dFnYWSndgU1fzK9n6-QcL76gBa/view) GBE classification performance while improving interpretability and generalizability, supporting robust EEG assessment across heterogeneous clinical contexts. These findings confirm that reduced feature spaces can provide both computational efficiency and clinical transparency, making them well suited for neuroscreening applications.

Future work will extend the Selective Activation algorithm into dedicated software to facilitate integration into clinical workflows and to strengthen healthcare professionals’ trust in automated assessment tools. In parallel, exploring unsupervised and neural-network-based models may uncover structure beyond hand-crafted features [6], further enhancing generalizability across heterogeneous datasets and improving applicability in real-world computational medicine.

**Acknowledgements.** I acknowledge the support of the research supervisor, dr hab. Jarosław Żygierewicz (University of Warsaw, Faculty of Physics, Institute of Experimental Physics, Biomedical Physics)

**References**

[1] Singh AK, Krishnan S. Trends in EEG signal feature extraction applications. *Front Artif Intell*. 2023;5:1072801. Published 2023 Jan 25. doi: [10.3389/frai.2022.1072801](https://doi.org/10.3389/frai.2022.1072801)

[2] Poziomska, M., Dovgialo, M., Olbratowski, P., Niedbalski, P., Ogniewski, P., Zych, J., Rogala,

J., and Żygierewicz, J. (2024). Quantity versus Diversity: Influence of Data on Detecting

EEG Pathology with Advanced ML Models. doi:[10.48550/arXiv.2411.17709](http://dx.doi.org/10.48550/arXiv.2411.17709)

[3] Bugli C, Lambert P. Comparison between principal component analysis and independent component analysis in electroencephalograms modelling. *Biom J*. 2007;49(2):312-327. doi: [10.1002/bimj.200510285](https://doi.org/10.1002/bimj.200510285)

[4] Rafieian, B., Hermosilla, P., and Vázquez, P.-P. (2023). Improving Dimensionality Reduction

Projections for Data Visualization. Applied Sciences, 13(17):9967. doi:[10.3390/app13179967](http://dx.doi.org/10.3390/app13179967)

[5] Wong, S., Simmons, A., Rivera-Villicana, J., Barnett, S., Sivathamboo, S., Perucca, P., Kwan,

P., Kuhlmann, L., Vasa, R., and O’Brien, T. J. (2023). EEG based automated seizure detection

– A survey of medical professionals. Epilepsy & Behavior, 149:109518. doi: [10.1016/j.yebeh.2023.109518](https://doi.org/10.1016/j.yebeh.2023.109518)

[6] Marzec, M. *Exploring EEG Features Structure for Neuroscreening: A Study of Dimensionality Reduction Techniques*. BSc Thesis, University of Warsaw, 2025. [<https://drive.google.com/file/d/1lkGzY4agJ_ATS46nA7OnwqCah40d8xu6/view?usp=share_link>]