Comparative Evaluation of Adjacency Matrix Applications for EEG Signal Classification Tasks

Jonas Mindaugas Rimšelis[1, 2], Gajane Mikalkėnienė[1, 2], Jolita Bernatavičienė[1]

[1]Institute of Data Science and Digital Technologies, Vilnius University
[2]Republican Vilnius Psychiatric Hospital

Institute of Data Science and Digital Technologies

Introduction

Major depressive disorder (MDD) is characterized by abnormal resting-state functional brain connectivity. EEG-based functional connectivity, especially when paired with machine learning, offers promising advancements in MDD diagnostics (Li et al. 2022). Represented as adjacency matrices derived from correlations between EEG electrodes, functional connectivity provides a graph-based framework for analysis. Comparative evaluation of various metrics is critical, as each captures distinct aspects of brain connectivity (Wang et al. 2024). This study utilizes SVM, Random Forests, and XGBoost to assess the classification performance of functional connectivity metrics — Pearson's correlation (Corr), phase-locked value (PLV), phase lag index (PLI), and the imaginary part of coherence (iCoh) — in distinguishing brain resting states.

Datasets

For the comparative analysis of adjacency metrics, the following datasets were used:

1. **EEG Motor Movement/Imagery dataset (*MMI*)**: EEG signals were recorded from 100 subjects per label (closed or open eyes) using 64 electrodes at a 160 Hz sampling rate.
2. **Republican Vilnius Psychiatric Hospital dataset (*RVPH*)**: EEG signals were recorded from 100 subjects per label (healthy control (HC) or MDD) using 20 electrodes at a 256 Hz sampling rate.

Aim

This study aims to evaluate the ability of adjacency metrics to differentiate resting-state brain functional connectivity.

Preprocessing

For both datasets, identical preprocessing steps were applied:

1. **Notch filter**: Removes 50 Hz power line noise.
2. **Bandpass filter**: Extracts frequencies corresponding to different brain wave types.
3. **Z-score normalization**: Applied per electrode.

All types	Delta	Theta	Alpha	Beta
1-40 Hz	0.5-4 Hz	4-8 Hz	8-12 Hz	12-30 Hz

Table: Brain wave type with corresponding frequencies.

Adjacency matrices

For each subject, weighted normalized adjacency matrices were calculated across different frequency ranges using the Corr, PLV, PLI, and iCoh metrics. A threshold of 0 was applied to binarize these matrices. Both binarized and weighted matrices were then aggregated across brain wave types and different adjacency metrics.

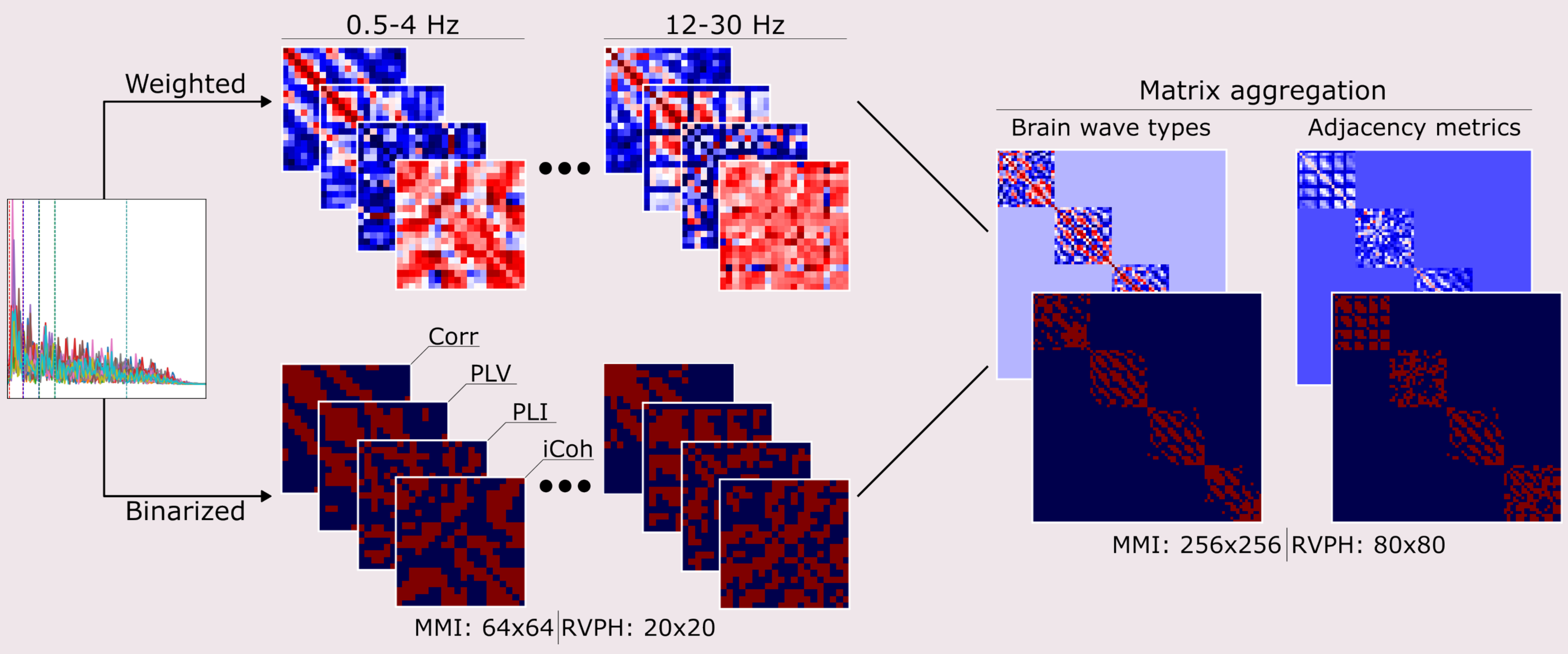

Figure: Framework of adjacency matrices calculation

Feature selection

Feature selection was performed using Lasso regularization and Principal Component Analysis (PCA) for each adjacency matrix aggregated by brain wave type and adjacency metric. These methods were applied to flattened matrices, where each variable represents the relationship between two electrodes. The selected features were validated through bootstrapping, with SVM, Random Forest, and XGBoost classifiers trained on resampled training subsets and tested on out-of-sample data.

Figure: Framework of feature selection

Model optimization

The SVM, Random Forest, and XGBoost classifiers were optimized using hyperparameter tuning through grid search and 5-fold cross-validation on the selected feature training sets. Each classifier's performance was evaluated on the test subset, reporting accuracy and AUC. For the classifier with the highest test accuracy, ROC curves were plotted to further assess performance.

Results for EEG Motor Movement / Imagery dataset

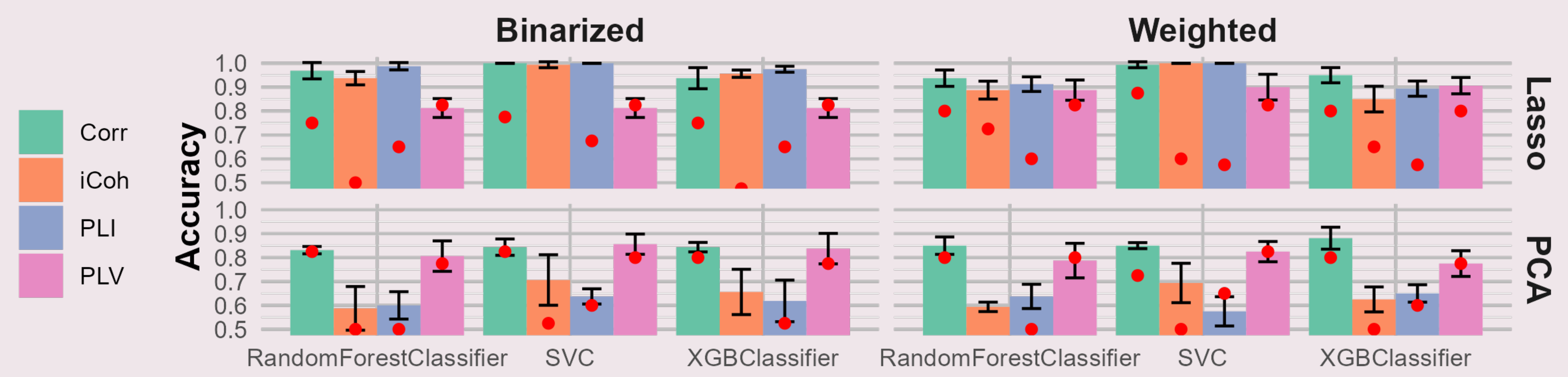

Figure: Classification results of MMI adjacency matrices with selected features aggregated on brain wave types. Red dot shows test accuracy.

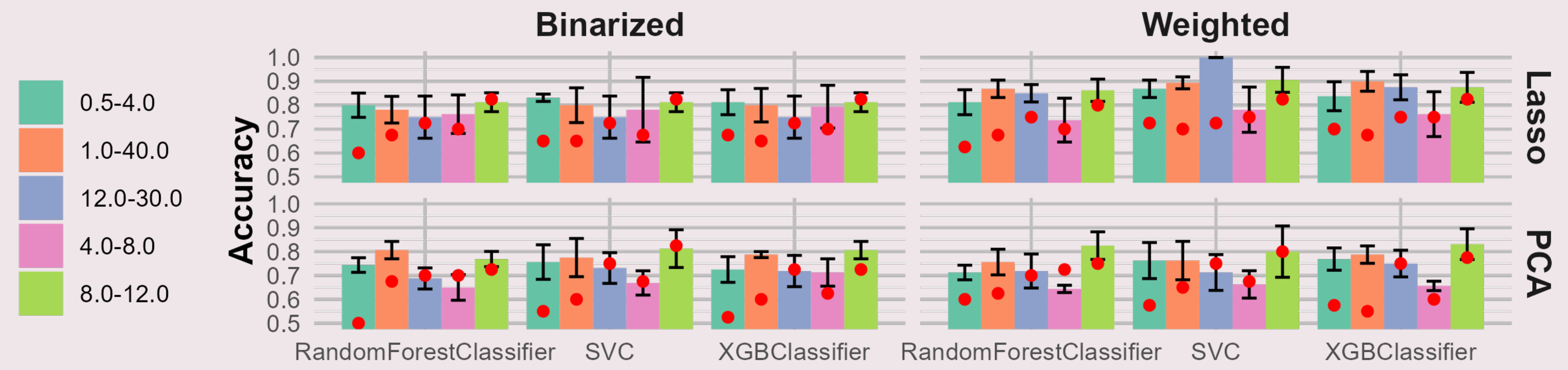

Figure: Classification results of MMI adjacency matrices with selected features aggregated on adjacency metrics. Red dot shows test accuracy.

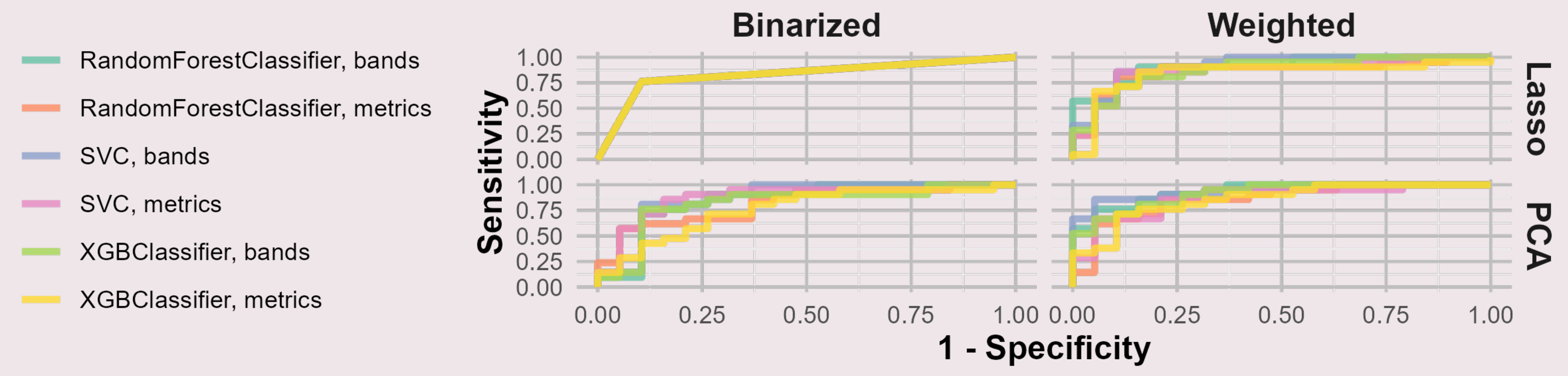

Figure: ROC curves of MMI classifiers with highest test accuracies.

Results for Republican Vilnius Psychiatric Hospital dataset

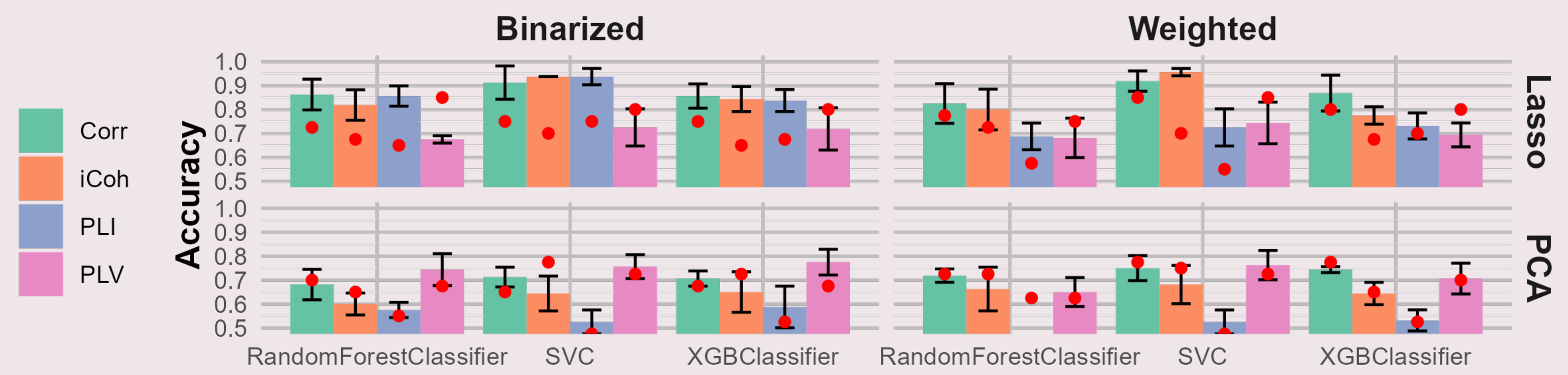

Figure: Classification results of RVPH adjacency matrices with selected features aggregated on brain wave types. Red dot shows test accuracy.

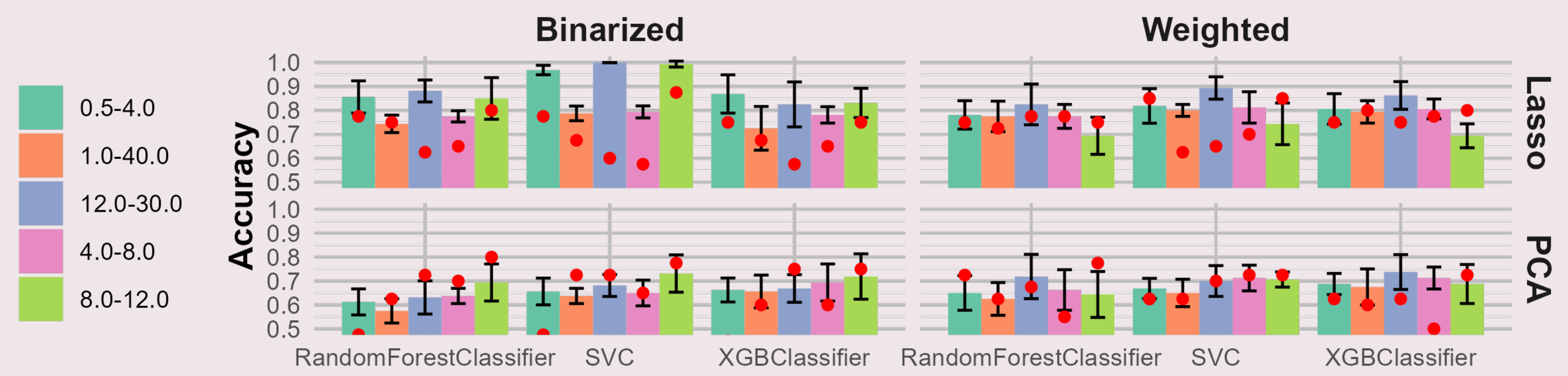

Figure: Classification results of RVPH adjacency matrices with selected features aggregated on adjacency metrics. Red dot shows test accuracy.

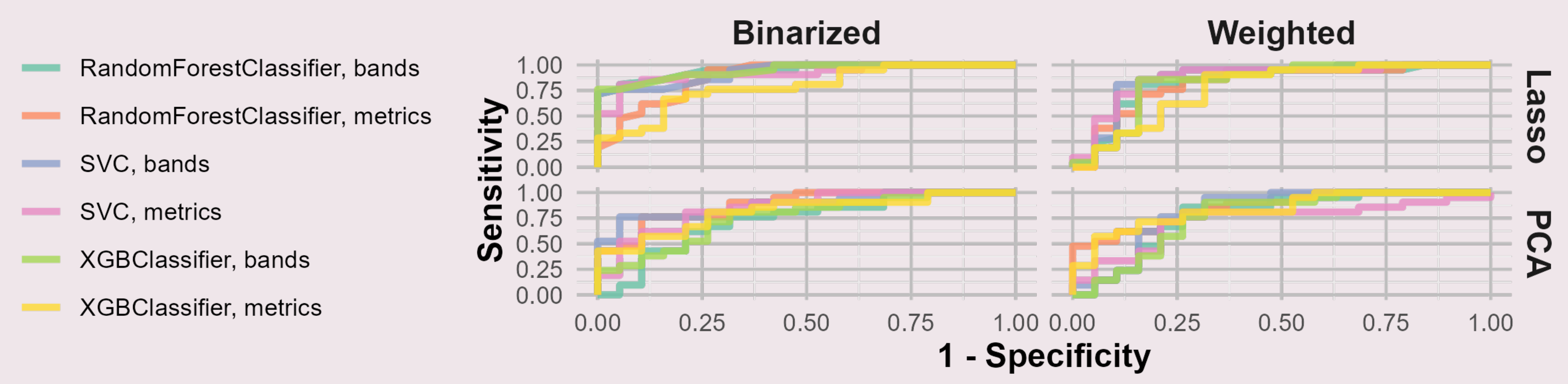

Figure: ROC curves of RVPH classifiers with highest test accuracies.

Conclusions

1. For the MMI dataset, weighted Pearson's correlation (Corr) matrices aggregated across frequency bands with Lasso-selected features resulted in the highest accuracy on test data (87.%, CV accuracy - 99.4%), with an AUC of 91.5%.
2. For the RVPH dataset, the highest accuracy was observed with 8.0-12.0 Hz binarized adjacency matrices aggregated across metrics with Lasso-selected features, yielding 87.5% accuracy (CV accuracy - 99.4%) and an AUC of 91.5%.
3. In all experiments, Lasso regularization consistently outperformed PCA for feature selection. This is likely due to the preservation of the graph structure in the selected features.

Acknowledgments

The conference participation is funded by EPAM.

References

Li, Yujie et al. (2022). "A Novel EEG-based Major Depressive Disorder Detection Framework with Two-stage Feature Selection". In: *BMC Medical Informatics and Decision Making* 22.1, pp. 1–13.

Wang, Yingtan et al. (2024). "Alterations in electroencephalographic functional connectivity in individuals with major depressive disorder: a resting-state electroencephalogram study". In: *Frontiers in Neuroscience* 18. ISSN: 1662-453X.