

Identifying Major Depressive Disorder with a Sparse Set of fMRI Functional Connectivity Biomarkers

Nikita M. Smirnov

*Baltic Center for Neurotechnology and Artificial Intelligence
Immanuel Kant Baltic Federal University
Kaliningrad, Russia
nmsmirnov@kantiana.ru*

Drozdstoy Stoyanov

*Medical University Plovdiv, Research Institute
Plovdiv, Bulgaria
ORCID: 0000-0002-9975-3680*

Abstract—Differentiating Major Depressive Disorder (MDD) from healthy controls using resting-state fMRI is challenged by the high dimensionality of functional connectivity data. This study addresses this by using forward sequential feature selection to identify a sparse set of discriminative connections from an initial set of over 13,000 features. A K-Nearest Neighbors (KNN) classifier was trained on the selected features. A small subset of 20 functional connections was sufficient to classify MDD patients with an F1 score of 0.83, a significant improvement from 0.57 when using the full feature set. The classification performance was specific to the KNN algorithm using a cosine distance metric, suggesting that the directionality of connectivity patterns is a key distinguishing factor. The results indicate that feature selection is an effective strategy for developing robust fMRI-based diagnostic models for MDD.

Index Terms—Major Depressive Disorder, fMRI, functional connectivity, machine learning, feature selection, biomarkers.

I. INTRODUCTION

Major Depressive Disorder (MDD) is a leading cause of disability worldwide, yet its diagnosis relies predominantly on subjective clinical assessments. The search for objective, data-driven biomarkers is therefore a critical area of psychiatric research. The rapid development of modern computing technologies has enabled significant progress in this and a wide variety of other fields, ranging from physics and electronics [1], [2] to social and human studies [3]–[8].

Resting-state functional magnetic resonance imaging (fMRI) has emerged as a promising tool, allowing for the non-invasive mapping of the brain's intrinsic functional architecture by analyzing spontaneous brain activity [9], [10]. A powerful method for analyzing these data is the construction of functional networks, which model the brain as a graph of regions (nodes) connected by pairwise statistical dependencies (edges) [11]–[19]. This approach has consistently revealed alterations in functional connectivity in individuals with MDD and has shown significant predictive value in classifying patients from healthy controls [20]–[22]. Furthermore, artificial intelligence and machine learning approaches effectively identify biomarkers of these disorders and can be used to train classifiers and medical decision support systems [23]–[26].

However, translating these findings into robust diagnostic tools is challenging due to the high dimensionality of fMRI data (so called curse of dimensionality [27]). With thousands of potential functional connections to consider, machine learning models are susceptible to overfitting and often fail to generalize to new data. Consequently, identifying a small but reliable subset of features that carries the most discriminative information is a crucial step toward clinical application.

This study addresses this challenge directly by employing a machine learning approach to identify a sparse set of functional connectivity biomarkers for MDD. We utilize a forward sequential feature selection algorithm to systematically search for the most informative connections from the vast initial feature space. These selected features are then used to train a K-Nearest Neighbors (KNN) classifier to distinguish between MDD patients and healthy controls.

Our results demonstrate that a remarkably small subset of just 20 functional connections is sufficient to achieve high classification accuracy. Furthermore, we find that this result is highly specific, with performance depending critically on the choice of classification algorithm and distance metric. This work not only highlights the potential of feature selection for building effective diagnostic models but also offers deeper insights into the subtle and complex nature of brain network disruptions in MDD.

II. METHODOLOGY

A. Data Acquisition and Preprocessing

The study included 164 participants: 70 diagnosed with major depressive disorder (MDD) and 94 healthy controls (HC). All participants provided informed consent, and the study received ethical approval. Resting-state fMRI data was acquired on a 3T MRI system. The data underwent standard preprocessing steps, including realignment, co-registration with anatomical images, and normalization to MNI standard space using SPM 12 software. This process yielded voxel-level blood oxygen level dependent (BOLD) signals for each participant.

B. Feature Extraction

For each subject, functional brain networks were reconstructed from the preprocessed fMRI data. The brain was parcellated into 165 regions of interest based on the AAL3 atlas [28], yielding 13,530 unique pairs of regions. An average BOLD time series was extracted for each region, and the functional connectivity between each pair of regions was estimated by calculating the Pearson correlation coefficient of their respective signals. This resulted in a feature vector of 13,530 correlation values for each participant.

C. Classification and Feature Selection

A K-Nearest Neighbors (KNN) [29] classifier with cosine distance was used to distinguish between MDD patients and healthy controls. To address the high dimensionality of the feature space, a forward sequential feature selection algorithm [30] was employed to identify the 20 most informative features. The selection process was guided by a stratified 5-fold cross-validation strategy [30], using the F1 score as the evaluation metric to ensure the robustness of the chosen features. The resulting 20 most discriminative functional connections were then visualized using the mne-connectivity library [31].

III. RESULTS

The classification performance was evaluated using the F1 score. When using the full feature space of 13,530 correlation values, the KNN classifier achieved an F1 score of 0.57 ± 0.11 . After applying forward sequential feature selection to identify the 20 most informative features, the classifier's performance significantly improved to an F1 score of 0.83 ± 0.07 .

The 20 selected functional connections that provided the highest discriminative power are visualized in Fig. 1.

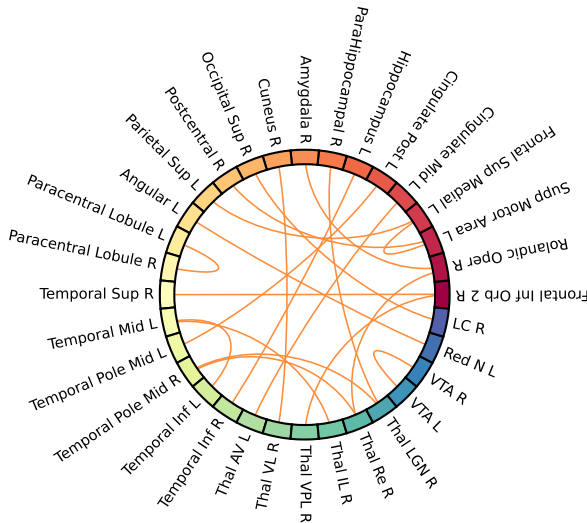


Fig. 1. The 20 most discriminative functional connections identified by the sequential feature selection algorithm. These connections yielded the highest performance for classifying individuals with MDD and healthy controls.

IV. CONCLUSION

This study demonstrates that a small subset of 20 functional connectivity features can effectively distinguish between individuals with MDD and healthy controls, achieving a high F1 score of 0.83 ± 0.07 . This is a significant improvement over the 0.57 ± 0.11 F1 score obtained using the entire set of 13,530 features.

Interestingly, the classification performance is highly specific to the chosen methodology. While the selected features enable accurate classification with a KNN classifier using cosine distance, they do not show significant differences in their mean correlation values between the MDD and healthy groups. Furthermore, employing a different classifier, such as a Random Forest [32], on the same 20 features yields poor results. Similarly, substituting the cosine distance with Euclidean distance in the KNN classifier also leads to a significant drop in performance.

These findings suggest that the feature space, particularly when samples are represented as normalized vectors, has a complex and non-linear boundary separating the two groups. The success of the cosine distance metric indicates that the direction of the feature vectors, rather than their magnitude, is the critical differentiating factor. This highlights the intricate nature of brain connectivity alterations in MDD, which may not be discernible through simple statistical comparisons or conventional classification models.

ACKNOWLEDGMENT

This work was supported by the Brain Program of the IDEAS Research Center. The author would like to thank the members of the Baltic Center of Neurotechnology and Artificial Intelligence for their valuable feedback and discussions.

REFERENCES

- [1] A. E. Dubinov, A. G. Petrik, S. A. Kurkin, N. S. Frolov, A. A. Koronovskii, and A. E. Hramov, "Virpertron: A novel approach for a virtual cathode oscillator design," *Physics of Plasmas*, vol. 24, no. 7, 2017.
- [2] E. Egorov, A. Koronovskii, S. Kurkin, and A. Hramov, "Formation and nonlinear dynamics of the squeezed state of a helical electron beam with additional deceleration," *Plasma Physics Reports*, vol. 39, no. 11, pp. 925–935, 2013.
- [3] M. V. Khramova, A. K. Kuc, V. A. Maksimenko, N. S. Frolov, V. V. Grubov, S. A. Kurkin, A. N. Pisarchik, N. N. Shusharina, A. A. Fedorov, and A. E. Hramov, "Monitoring the cortical activity of children and adults during cognitive task completion," *Sensors*, vol. 21, no. 18, p. 6021, 2021.
- [4] T. V. Bukina, M. V. Khramova, and S. A. Kurkin, "Modern research on primary school children brain functioning in the learning process: Review," *Izvestiya VUZ. Applied Nonlinear Dynamics*, vol. 29, no. 3, pp. 449–456, 2021.
- [5] S. Kurkina, V. Grubova, V. Maksimenkoa, E. Pitsika, and M. Khramovab, "System for monitoring and adjusting the learning process of primary schoolchildren based on the eeg data analysis," *Information and Control*, no. 5, pp. 50–61, 2020.
- [6] M. V. Khramova, T. V. Bukina, N. M. Smirnov, S. A. Kurkin, and A. E. Hramov, "Prevalence of neuromyths among students and pre-service teachers," *Humanities and Social Sciences Communications*, vol. 10, no. 1, pp. 1–14, 2023.
- [7] V. V. Grubov, M. V. Khramova, S. Goman, A. A. Badarin, S. A. Kurkin, D. A. Andrikov, E. Pitsik, V. Antipov, E. Petushok, N. Brusinskii *et al.*, "Open-loop neuroadaptive system for enhancing student's cognitive abilities in learning," *IEEE Access*, vol. 12, pp. 49 034–49 049, 2024.

- [8] S. Kurkin, S. Gordileeva, A. Savosenkov, N. Grigorev, N. Smirnov, V. V. Grubov, A. Udoratina, V. Maksimenko, V. Kazantsev, and A. E. Hramov, "Transcranial magnetic stimulation of the dorsolateral prefrontal cortex increases posterior theta rhythm and reduces latency of motor imagery," *Sensors*, vol. 23, no. 10, p. 4661, 2023.
- [9] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature reviews neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.
- [10] K. J. Friston, "Functional and effective connectivity: a review," *Brain connectivity*, vol. 1, no. 1, pp. 13–36, 2011.
- [11] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.
- [12] A. E. Hramov, N. S. Frolov, V. A. Maksimenko, S. A. Kurkin, V. B. Kazantsev, and A. N. Pisarchik, "Functional networks of the brain: from connectivity restoration to dynamic integration," *Physics-Uspexhi*, vol. 64, no. 6, p. 584, 2021.
- [13] S. A. Kurkin, N. M. Smirnov, R. Paunova, S. Kandilarova, D. Stoyanov, L. Mayorova, and A. E. Hramov, "Beyond pairwise interactions: higher-order q-analysis of fmri-based brain functional networks in patients with major depressive disorder," *IEEE Access*, vol. 12, pp. 197 168–197 186, 2024.
- [14] A. Badarin, A. Andreev, V. Klinshov, V. Antipov, and A. E. Hramov, "Hidden data recovery using reservoir computing: Adaptive network model and experimental brain signals," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 34, no. 10, 2024.
- [15] V. S. Khorev, S. A. Kurkin, G. Zlateva, R. Paunova, S. Kandilarova, M. Maes, D. Stoyanov, and A. E. Hramov, "Disruptions in segregation mechanisms in fmri-based brain functional network predict the major depressive disorder condition," *Chaos, Solitons & Fractals*, vol. 188, p. 115566, 2024.
- [16] R. Islam, A. V. Andreev, N. N. Shusharina, and A. E. Hramov, "Explainable machine learning methods for classification of brain states during visual perception," *Mathematics*, vol. 10, no. 15, p. 2819, 2022.
- [17] A. Badarin, V. Klinshov, P. Smelov, and A. Andreev, "Reservoir computing reconstructs blood-oxygen-level-dependent signals: whole-brain modeling study," *The European Physical Journal Special Topics*, pp. 1–8, 2025.
- [18] N. S. Frolov, V. A. Maksimenko, M. V. Khramova, A. N. Pisarchik, and A. E. Hramov, "Dynamics of functional connectivity in multilayer cortical brain network during sensory information processing," *The European Physical Journal Special Topics*, vol. 228, no. 11, pp. 2381–2389, 2019.
- [19] A. Badarin, V. Antipov, V. Grubov, A. Andreev, E. Pitsik, S. Kurkin, and A. Hramov, "Brain compensatory mechanisms during the prolonged cognitive task: fnirs and eye-tracking study," *IEEE Transactions on Cognitive and Developmental Systems*, 2024.
- [20] A. V. Andreev, S. A. Kurkin, D. Stoyanov, A. A. Badarin, R. Paunova, and A. E. Hramov, "Toward interpretability of machine learning methods for the classification of patients with major depressive disorder based on functional network measures," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 33, no. 6, 2023.
- [21] D. Stoyanov, V. Khorev, R. Paunova, S. Kandilarova, D. Simeonova, A. Badarin, A. Hramov, and S. Kurkin, "Resting-state functional connectivity impairment in patients with major depressive episode," *International Journal of Environmental Research and Public Health*, vol. 19, no. 21, p. 14045, 2022.
- [22] A. N. Pisarchik, A. V. Andreev, S. A. Kurkin, D. Stoyanov, A. A. Badarin, R. Paunova, and A. E. Hramov, "Topology switching during window thresholding fmri-based functional networks of patients with major depressive disorder: Consensus network approach," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 33, no. 9, 2023.
- [23] O. E. Karpov, M. S. Khoymov, V. A. Maksimenko, V. V. Grubov, N. Utyashev, D. A. Andrikov, S. A. Kurkin, and A. E. Hramov, "Evaluation of unsupervised anomaly detection techniques in labelling epileptic seizures on human eeg," *Applied Sciences*, vol. 13, no. 9, p. 5655, 2023.
- [24] S. Kurkin, E. Pitsik, and N. Frolov, "Artificial intelligence systems for classifying eeg responses to imaginary and real movements of operators," in *Saratov Fall Meeting 2018: Computations and Data Analysis: from Nanoscale Tools to Brain Functions*, vol. 11067. SPIE, 2019, pp. 42–47.
- [25] P. Chholak, A. N. Pisarchik, S. A. Kurkin, V. A. Maksimenko, and A. E. Hramov, "Phase-amplitude coupling between mu-and gamma-waves to carry motor commands," in *2019 3rd School on Dynamics of Complex Networks and their Application in Intellectual Robotics (DCNAIR)*. IEEE, 2019, pp. 39–45.
- [26] V. V. Grubov, S. I. Nazarikov, S. A. Kurkin, N. P. Utyashev, D. A. Andrikov, O. E. Karpov, and A. E. Hramov, "Two-stage approach with combination of outlier detection method and deep learning enhances automatic epileptic seizure detection," *IEEE Access*, vol. 12, pp. 122168–122182, 2024.
- [27] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.
- [28] E. T. Rolls, C.-C. Huang, C.-P. Lin, J. Feng, and M. Joliot, "Automated anatomical labelling atlas 3," *Neuroimage*, vol. 206, p. 116189, Feb. 2020.
- [29] E. Fix, "Discriminatory analysis: nonparametric discrimination, consistency properties," vol. 1, 1985.
- [30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [31] E. Larson, A. Gramfort, and D. A. e. a. Engemann, "MNE-python," 2025.
- [32] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.