Boltzmann-Inspired Model for fMRI Time-Series Classification

Joseph Van Duyn[1] JVANDUYN@EMORY.EDU
Pavel Popov[2,4] PPOPOV@GSU.EDU
Armin Iraji[2,4] AIRAJI@GSU.EDU
Zening Fu[2,4] ZFU@GSU.EDU
Vince D. Calhoun[1,2,3,4] VCALHOUN@GSU.EDU
Sergey M. Plis[2,4] SPLIS@GSU.EDU
Alex Fedorov[1] AVFEDOR@EMORY.EDU

[1] *Emory University, Atlanta, GA, USA*
[2] *Georgia State University, Atlanta, GA, USA*
[3] *Georgia Institute of Technology, Atlanta, GA, USA*
[4] *Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Atlanta, GA, USA*

Abstract

Accurate classification of psychiatric and neurological disorders using resting-state fMRI data is critical for understanding the mechanisms underlying brain diseases, yet existing approaches often rely on complex architectures that are difficult to interpret. In this work, we introduce a Boltzmann-inspired Model (BiM) for fMRI classification that leverages bilinear interactions between features within a lightweight mean aggregation framework. Our model captures higher-order dependencies among brain regions. We evaluate the proposed model across multiple brain disorders data and parcellations based on Independent Component Analysis (ICA) and Regions of Interest (ROIs). The performance of the proposed model is compared with state-of-the-art time series and functional network connectivity models on the classification of diseases and disorders. The proposed model demonstrates competitive or superior Area Under the Receiver Operating Characteristic Curve (ROC AUC) performance, particularly using ROI-based features. These results suggest that relatively simple, bilinear architectures can match or surpass more complex models.

Keywords: resting-state fMRI, machine learning, deep learning, brain disorders, predictive neuroimaging

Data and Code Availability We used resting-state fMRI images collected from FBIRN (Function Biomedical Informatics Research Network) (Keator et al., 2016), COBRE (Center of Biomedical Research Excellence) (Cetin et al., 2014), BSNIP (Bipolar and Schizophrenia Network for Intermediate Phenotypes) (Tamminga et al., 2014), ABIDE (Autism Brain Imaging Data Exchange, release 1.0) (Di Martino et al., 2014), OASIS (Open Access Series of Imaging Studies, release 3.0) (Rubin et al., 1998), ADNI (Alzheimer's Disease Neuroimaging Initiative) (Petersen et al., 2010), HCP (Human Connectome Project, 1200 subjects release) (Van Essen et al., 2013), and UK Biobank.

Institutional Review Board (IRB) This study involved secondary analysis of publicly available, de-identified data and did not require IRB approval.

1. Introduction

Recent advances in machine learning for functional MRI (fMRI) have opened new opportunities for classifying psychiatric disorders and neurological diseases, as well as for identifying clinically relevant biomarkers (Liu et al., 2021; de Filippis et al., 2019; Warren and Moustafa, 2023). By capturing spatiotemporal patterns of neural activity, fMRI provides a rich, high-dimensional view of brain function. However, translating these data into accurate predictions remains challenging due to the complexity of feature interactions, variability across datasets, and the difficulty of interpreting fMRI-based models.

A critical step in fMRI analysis is how brain regions are represented. One common approach is anatom-

ical parcellation, where the brain is segmented into regions of interest (ROIs) (Desikan et al., 2006; Schaefer et al., 2018). Alternatively, data-driven methods such as independent component analysis (ICA) can define parcels based on functional organization (Du et al., 2020). These parcellations allow researchers to derive functional connectivity networks that capture distributed activity patterns across the brain. Once a spatial representation is chosen, modeling strategies vary: functional network connectivity (FNC) methods quantify correlations between regions, whereas time-series models operate directly on regional activity signals over time. Both representations can serve as the basis for predictive models of psychiatric disorders and neurological diseases.

In this work, we introduce **BiM**, a novel **B**oltzmann-**i**nspired **M**odel for fMRI time-series classification. Building upon the meanMLP architecture (Popov et al., 2024), BiM incorporates bilinear energy interactions between features. The key idea is to capture higher-order dependencies among brain regions, in a manner similar to functional connectivity, while still leveraging the temporal richness of time-series representations. In doing so, BiM seeks to combine the strengths of both worlds: the direct modeling of temporal dynamics and the network-level structure of functional connectivity. We evaluate our approach across multiple benchmark datasets and show that it consistently matches or significantly outperforms state-of-the-art time-series models such as meanMLP (Popov et al., 2024), BolT (Bedel et al., 2023), and Mutual Information Local to Context (MILC) (Mahmood et al., 2020), as well as functional connectivity models including the BrainNetwork Transformer (BNT) (Kan et al., 2022) and Logistic Regression from functional network connectivity (LR) (Popov et al., 2024).

2. Methods

To capture pairwise interactions between ROIs, we drew inspiration from the *Boltzmann machine* (Sherrington and Kirkpatrick, 1975), a network where units (on/off variables) interact through weighted connections that define an energy landscape. States with lower energy are more likely under the Boltzmann distribution:

$$P(\mathbf{v},\mathbf{h}) = \frac{1}{Z}\exp(-E(\mathbf{v},\mathbf{h}))\,, \tag{1}$$

with partition function

$$Z = \sum_{\mathbf{v},\mathbf{h}} \exp(-E(\mathbf{v},\mathbf{h}))\,.$$

Here, $\mathbf{v}$ denotes the *visible units*, which correspond to observed features (e.g., ROI or ICA time series), and $\mathbf{h}$ denotes the *hidden units*, which capture latent structure in the data. The function $E(\mathbf{v},\mathbf{h})$ is the energy of a configuration $(\mathbf{v},\mathbf{h})$, and Z is the normalization constant (partition function) ensuring that probabilities sum to one.

The energy function is

$$E = -\Big(\sum_{i<j} w_{ij}s_i s_j + \sum_i \theta_i s_i\Big),$$

where w_{ij} captures pairwise interactions between units, and θ_i biases individual units. Here, $s_i \in \{0,1\}$ denotes the binary state (active/inactive) of unit i. Minimizing this energy favors the most probable configurations and could provide a principled way to model dependencies in high-dimensional fMRI data, similar to functional connectivity.

2.1. Boltzmann-inspired model (BiM)

Based on the physical intuition of energy-based systems, we propose the **Boltzmann-inspired model (BiM)**. BiM builds on bilinear energy interactions between features with hidden state size d (as tunable hyperparameters).

Let a single fMRI time series sample be $\{x, y\}$, where $x \in \mathbb{R}^{T\times k}$ and $y \in \mathbb{R}^C$. Here, T denotes the number of time points, k the input feature dimension (e.g., number of ROIs or ICs), and C the number of classes. In the implementation, each time point $x_t \in \mathbb{R}^k$ is processed independently to form

$$e_t = f\big(W_b(x_t, x_t)\big) + f\big(W_l x_t\big), \tag{2}$$

where $W_b \in \mathbb{R}^{d\times k\times k}$ is a bilinear transformation capturing pairwise interactions among features *within* x_t, $W_l \in \mathbb{R}^{d\times k}$ is a linear term, and $f(\cdot)$ denotes the composition of layer normalization, ReLU, and dropout. The sequence $\{e_t\}_{t=1}^T$ is then aggregated by mean pooling,

$$e_{mean} = \frac{1}{T}\sum_{t=1}^{T} e_t, \tag{3}$$

to obtain a fixed-length representation, which is passed through a linear classifier

$$h = W_c e_{mean} + b_c, \quad W_c \in \mathbb{R}^{C\times d},\ b_c \in \mathbb{R}^C, \tag{4}$$

Table 1: Median (IQR) ROC AUC for ICA FNC and Time Series Models. Bold indicates the best model per dataset for time-series based models, while underline indicates the best model per dataset for FNC based models.

Model	ABIDE	ADNI	BSNIP	COBRE	FBIRN	HCP	OASIS
Time Series Models							
MILC	0.669 (0.046)	0.730 (0.068)	0.695 (0.068)	**0.742** (0.136)	0.798 (0.079)	0.897 (0.038)	0.742 (0.042)
BolT	0.733 (0.062)	0.630 (0.115)	0.765 (0.056)	0.739 (0.093)	0.806 (0.068)	**0.934** (0.035)	0.756 (0.069)
meanMLP	**0.748** (0.041)	0.739 (0.058)	0.795 (0.062)	0.735 (0.104)	0.851 (0.045)	0.903 (0.039)	**0.821** (0.078)
BiM	0.740 (0.038)	**0.741** (0.066)	**0.808** (0.046)	0.724 (0.100)	**0.866** (0.031)	0.924 (0.030)	0.817 (0.084)
FNC Models							
BNT	0.719 (0.064)	0.750 (0.053)	0.757 (0.036)	0.739 (0.153)	0.842 (0.052)	0.888 (0.038)	0.780 (0.093)
LR	0.718 (0.044)	0.743 (0.115)	0.746 (0.033)	0.724 (0.082)	0.867 (0.032)	0.902 (0.035)	0.768 (0.080)

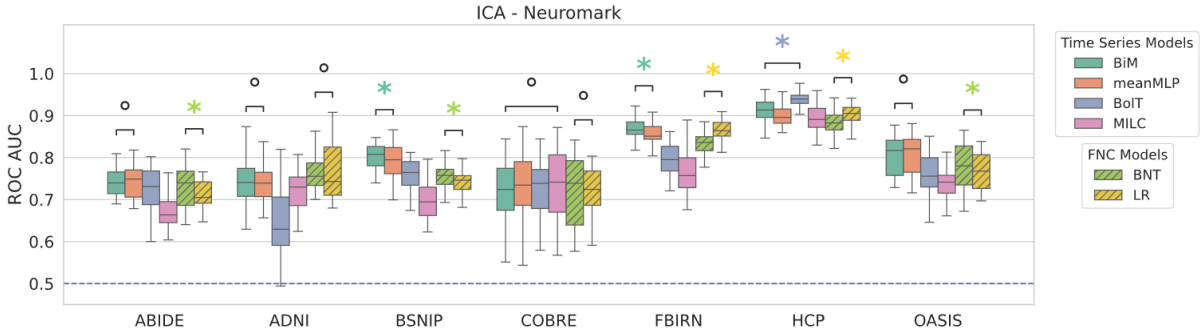

Figure 1: Model performance comparison on seven ICA-based datasets. The boxplots show the distribution of test ROC AUC scores across 50 runs. Asterisks denote a statistically significant improvement ($p < 0.05$, Wilcoxon signed-rank test) of one model over another. BiM demonstrates competitive performance, achieving significant improvements in the BSNIP and FBIRN datasets over the meanMLP model.

with predictions given by $\hat{y} = \text{Softmax}(h)$.

This formulation enables the model to capture nonlinear, pairwise dependencies among fMRI features at each time point. Due to the mean aggregation across $\{e_t\}_{t=1}^{T}$, the resulting representation is permutation invariant with respect to the time dimension, thus discarding temporal order while retaining distributional information.

2.2. Experimental setup

We tuned hyperparameters (HPs) for each model, including BiM's hidden size, dimensions, learning rate, and dropout. Using a 3% subset of the UK Biobank (UKB-S) data reserved for tuning, we ran 400 random trials per model, each evaluated with stratified 5-fold cross-validation. The configuration with the highest ROC AUC was selected. The experimental setup, the preprocessing and datasets details are the same as in meanMLP paper (Popov et al., 2024).

Models BiM, meanMLP (Popov et al., 2024), BolT (Bedel et al., 2023), BrainNetwork Transformer (BNT) (Kan et al., 2022), and Logistic Regression from functional network connectivity (LR) (Popov et al., 2024) were then trained and tested on twelve datasets using stratified 5-fold cross-validation with a 64/16/20 train/validation/test split, consistent across models. Mutual Information Local to Context (MILC) (Mahmood et al., 2020) was trained and tested on ten datasets, excluding the UK Biobank datasets (UKB-S and UKB-SA) due to their large size. Model checkpoints with the lowest validation loss were evaluated on the test set. To reduce variability from random seeds and splits, each fold was repeated ten times, yielding 50 independent test scores per model–dataset pair. For the UK Biobank datasets (UKB-S and UKB-SA), we performed 3-fold cross-validation instead of 5-fold, resulting in 30 independent test scores per model–dataset pair.

3. Results

Using the experimental setup described in the previous section, we trained all models for classification, comparing performance on both ICA- and ROI-based

Table 2: Median (IQR) ROC AUC for UK Biobank (UKB-S and UKB-SA).

Model	UKB-S	UKB-SA
Time Series Models		
BolT	**0.969** (0.002)	**0.832** (0.003)
meanMLP	0.968 (0.003)	0.821 (0.002)
BiM	**0.969** (0.002)	0.825 (0.001)
FNC Models		
BNT	0.965 (0.002)	0.824 (0.001)
LR	0.964 (0.003)	0.785 (0.003)

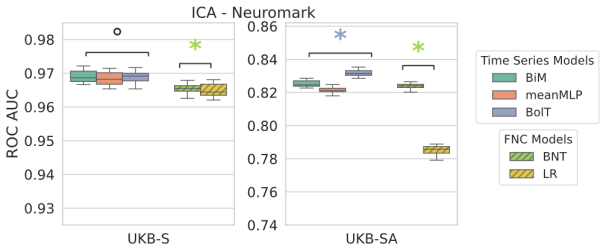

Figure 2: Model performance on UK Biobank (UKB-S) and UK Biobank (Age Bins) (UKB-SA). Boxplots show test ROC AUC across 30 runs. Asterisks denote statistically significant improvement ($p < 0.05$, Wilcoxon signed-rank test). All datasets are comparable for UKB-S, while BolT is the top-performing time-series model for UKB-SA.

Table 3: Median (IQR) ROC AUC for ROI datasets.

Model	ABIDE	FBIRN	HCP
Time Series Models			
MILC	0.670 (0.043)	0.828 (0.044)	0.901 (0.035)
BolT	0.737 (0.044)	0.820 (0.072)	0.921 (0.039)
meanMLP	**0.747** (0.034)	0.852 (0.045)	0.913 (0.038)
BiM	0.741 (0.023)	**0.865** (0.034)	**0.930** (0.028)
FNC Models			
BNT	0.706 (0.045)	0.859 (0.078)	0.890 (0.036)
LR	0.723 (0.033)	0.871 (0.031)	0.900 (0.037)

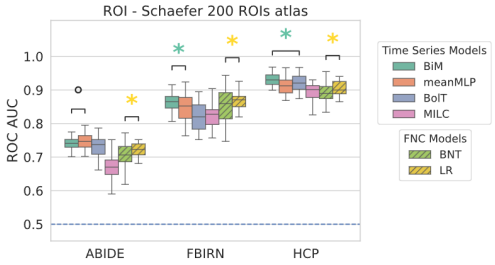

Figure 3: Model performance on ROI-based datasets. Boxplots show test ROC AUC across 50 runs. Asterisks denote statistically significant improvement ($p < 0.05$, Wilcoxon signed-rank test). BiM is consistently strong, surpassing meanMLP in FBIRN and HCP.

datasets. The tables 1, 2, and 3 report the median ROC AUC along with the interquartile range (IQR) for each model and dataset. For statistical comparison, asterisks denote significant differences between models according to Wilcoxon signed-rank tests ($p < 0.05$). Figure 1 shows these results as boxplots, visualizing the distribution of test ROC AUC scores across 50 runs for each dataset and model.

BiM demonstrates competitive performance on ICA datasets (see Figure 1 and Table 1). It achieves statistically significant improvements over competing models in BSNIP and FBIRN, as indicated by the asterisk annotations in Figure 1. While BolT achieves the best performance in HCP, BiM exhibits greater overall reliability. The meanMLP achieves the highest median ROC AUC in ABIDE and OASIS, although its advantage over BiM is not statistically significant under the Wilcoxon test.

On the UK Biobank (UKB-S) datasets (see Figure 2 and Table 2), all of the models perform similarly. However, when considering the UK Biobank (Age Bins) dataset (UKB-SA), which incorporates age-stratified subgroups, BolT performs significantly better than BiM and meanMLP. As reported in (Popov et al., 2024), meanMLP outperforms MILC in ROC AUC on both UKB-S and UKB-SA, indicating that MILC is not a top-performing model in either dataset; consequently, we did not run it. For FNC models, BNT outperforms LR on these datasets.

On the ROI datasets (see Figure 3 and Table 3), BiM is also consistently strong. It demonstrates statistically significant improvement over meanMLP in FBIRN and HCP.

While LR underperforms on ICA-based features, particularly relative to BNT, it achieves competitive or superior ROC AUC values on the ROI datasets, surpassing BNT in ABIDE, FBIRN, and HCP.

4. Conclusion

BiM achieved competitive or superior performance compared to state-of-the-art models. It performed robustly across both ROI-based and ICA-derived features, and rivaled meanMLP on UK Biobank datasets. The model's simplicity contributes to its efficiency and reduced risk of overfitting, making it a strong baseline for neuroimaging classification. Fu-

ture work could explore new aggregation schemes (e.g., attention or hierarchical pooling) to capture temporal dependencies. Future work could further investigate the interpretability of BiM for biomarker discovery in brain disorders and diseases. Overall, these findings show that lightweight models with bilinear interactions can serve as an effective predictive model for neuroimaging time-series analysis.

Acknowledgements

A.F. and J.V.D. were supported by the Nell Hodgson Woodruff School of Nursing at Emory University, and by philanthropic funds donated to the Goizueta ADRC, and by the NIH National Institute on Aging (P30AG066511). S.M.P., V.D.C. and P.P. were supported by NIH grants R01MH129047 and 2R01EB006841, and NSF grant 2112455.

References

Hasan A Bedel, Irmak Sivgin, Onat Dalmaz, Salman UH Dar, and Tolga Çukur. Bolt: Fused window transformers for fmri time series analysis. *Medical image analysis*, 88:102841, 2023.

Mustafa S Cetin, Fletcher Christensen, Christopher C Abbott, Julia M Stephen, Andrew R Mayer, José M Cañive, Juan R Bustillo, Godfrey D Pearlson, and Vince D Calhoun. Thalamus and posterior temporal lobe show greater inter-network connectivity at rest and across sensory paradigms in schizophrenia. *Neuroimage*, 97:117–126, 2014.

Renato de Filippis, Elvira Anna Carbone, Raffaele Gaetano, Antonella Bruni, Valentina Pugliese, Cristina Segura-Garcia, and Pasquale De Fazio. Machine learning techniques in a structural and functional mri diagnostic approach in schizophrenia: a systematic review. *Neuropsychiatric disease and treatment*, pages 1605–1627, 2019.

Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

Yuhui Du, Zening Fu, Jing Sui, Shuang Gao, Ying Xing, Dongdong Lin, Mustafa Salman, Anees Abrol, Md Abdur Rahaman, Jiayu Chen, et al. Neuromark: An automated and adaptive ica based pipeline to identify reproducible fmri markers of brain disorders. *NeuroImage: Clinical*, 28:102375, 2020.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer. *Advances in Neural Information Processing Systems*, 35:25586–25599, 2022.

David B Keator, Theo GM Van Erp, Jessica A Turner, Gary H Glover, Bryon A Mueller, Thomas T Liu, James T Voyvodic, Jerod Rasmussen, Vince D Calhoun, Hyo Jong Lee, et al. The function biomedical informatics research network data repository. *Neuroimage*, 124:1074–1079, 2016.

Meijie Liu, Baojuan Li, and Dewen Hu. Autism spectrum disorder studies using fmri data and machine learning: a review. *Frontiers in neuroscience*, 15: 697870, 2021.

Usman Mahmood, Md Mahfuzur Rahman, Alex Fedorov, Noah Lewis, Zening Fu, Vince D Calhoun, and Sergey M Plis. Whole milc: generalizing learned dynamics across tasks, datasets, and populations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 407–417. Springer, 2020.

Ronald Carl Petersen, Paul S Aisen, Laurel A Beckett, Michael C Donohue, Anthony Collins Gamst, Danielle J Harvey, CR Jack Jr, William J Jagust, Leslie M Shaw, Arthur W Toga, et al. Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology*, 74(3):201–209, 2010.

Pavel Popov, Usman Mahmood, Zening Fu, Carl Yang, Vince Calhoun, and Sergey Plis. A simple but tough-to-beat baseline for fmri time-series classification. *NeuroImage*, 303:120909, 2024.

Eugene H Rubin, Martha Storandt, J Philip Miller, Dorothy A Kinscherf, Elizabeth A Grant, John C

Morris, and Leonard Berg. A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Archives of neurology*, 55(3): 395–401, 1998.

Alexander Schaefer, Ru Kong, Evan M Gordon, Timothy O Laumann, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, and BT Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, 28(9):3095–3114, 2018.

David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35 (26):1792, 1975.

Carol A Tamminga, Godfrey Pearlson, Matcheri Keshavan, John Sweeney, Brett Clementz, and Gunvant Thaker. Bipolar and schizophrenia network for intermediate phenotypes: outcomes across the psychosis continuum. *Schizophrenia bulletin*, 40 (Suppl_2):S131–S137, 2014.

David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.

Samuel L Warren and Ahmed A Moustafa. Functional magnetic resonance imaging, deep learning, and alzheimer's disease: A systematic review. *Journal of Neuroimaging*, 33(1):5–18, 2023.