A300 (Fri-PM)

BrainMT: A HYBRID MAMBA-TRANSFORMER ARCHITECTURE FOR MODELING LONG-RANGE DEPENDENCIES IN FUNCTIONAL MRI DATA

Arunkumar Kannan*, **Martin A. Lindquist**†, **Brian Caffo**†

* Department of Electrical and Computer Engineering, Johns Hopkins University, USA
† Department of Biostatistics, Johns Hopkins University, USA

JOHNS HOPKINS UNIVERSITY

MICCAI 2025
28th International Conference on Medical Image Computing and Computer Assisted Intervention
23-27 September 2025
Daejeon Convention Center
Daejeon, REPUBLIC OF KOREA

BACKGROUND

Motivation: The Brain is Spatio-Temporal

- Functional-MRI (fMRI) captures brain activity across **space + time** [1].
- Most predictive models reduce fMRI to static connectivity matrices – losing rich spatial structure and temporal dynamics.

Challenge: Limits of Current Models

- Correlation-based models: collapse 4D to 2D; parcellation schemes are inconsistent [2].
- Voxel-based transformer models: quadratic cost; limited to short (10-20) frames.

Both approaches fail to model long-range temporal dependencies

Our Solution: BrainMT

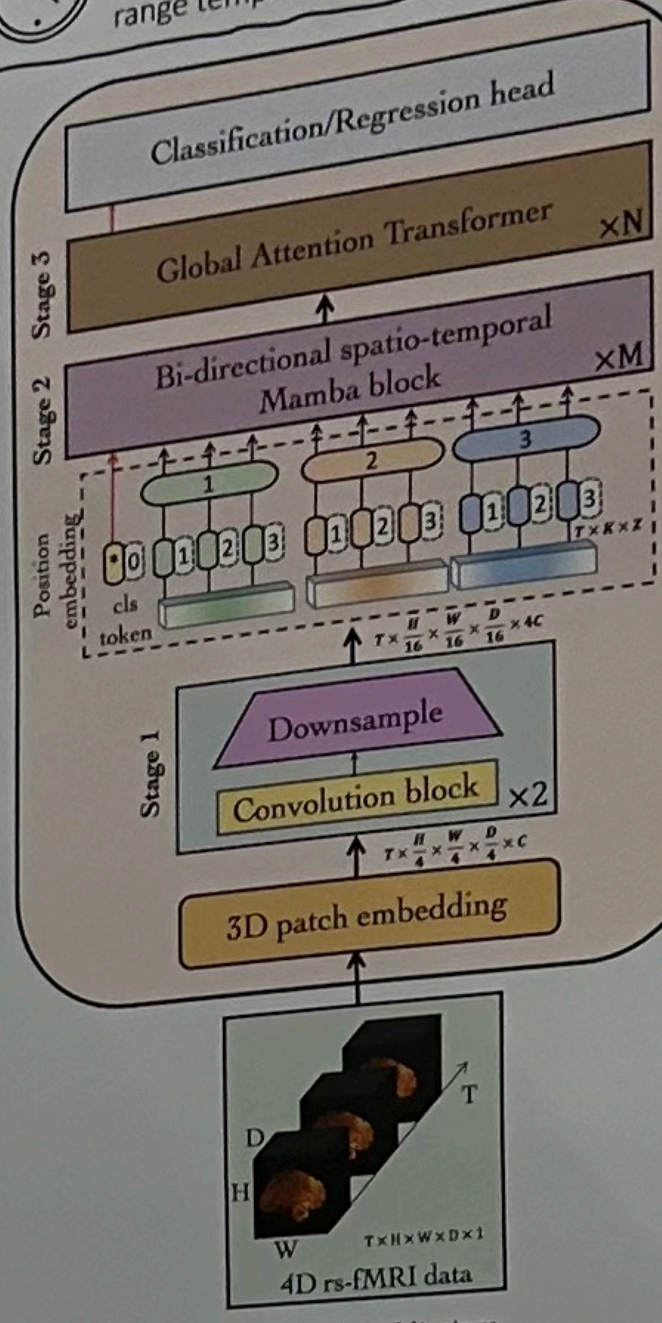
- Hybrid Mamba + Transformer for long-range temporal + global spatial modeling, trained end-to-end on voxel-level 4D fMRI.

METHODS

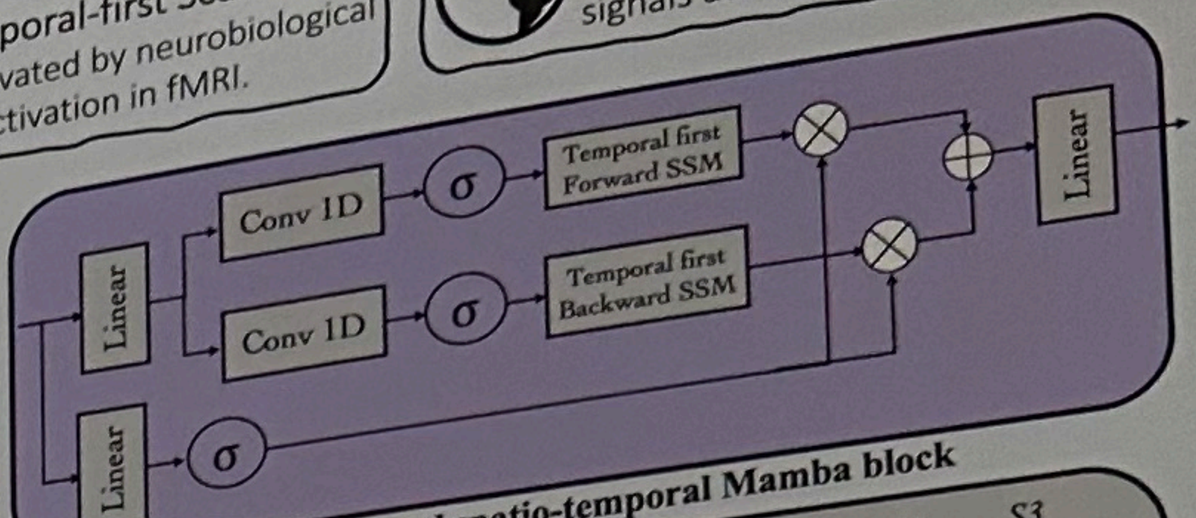
Temporal Dynamics
Mamba block captures long-range temporal correlations.

Temporal-first Scanning
Motivated by neurobiological co-activation in fMRI.

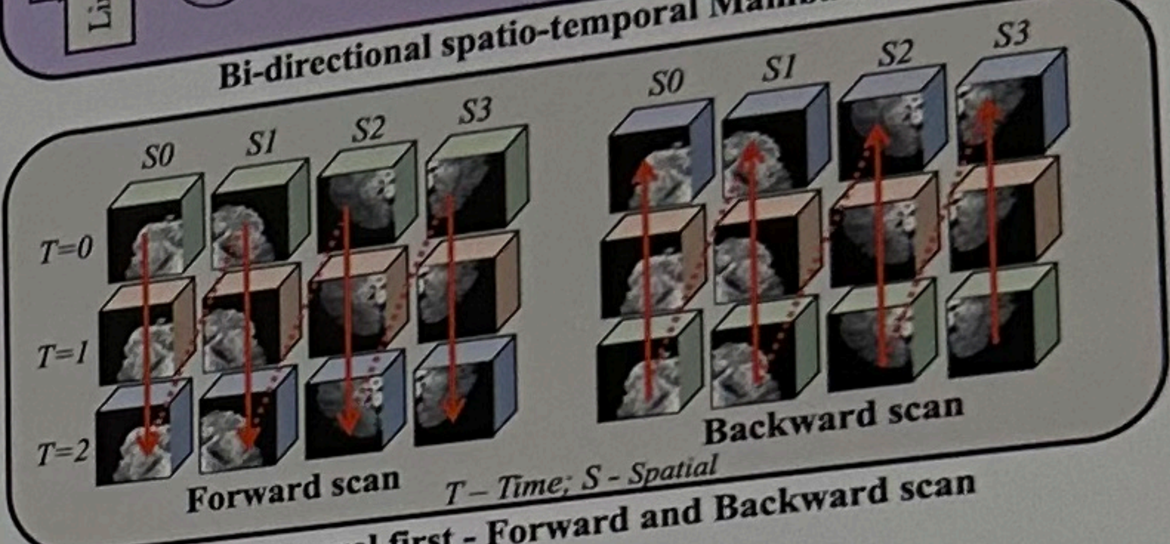
Spatial Integration
Global attention models temporal signals across brain regions.



BrainMT Architecture



Bi-directional spatio-temporal Mamba block



Temporal first - Forward and Backward scan

Spatio - Temporal Mamba Block

- Bi-directional selective state-space models (SSM) [3] learn long-range context in 4D fMRI sequences.
- **Temporal-first scan mechanism:** Time is treated as the leading dimension, followed by spatial dimensions, in the SSM's scan mechanism.

RESULTS

Datasets & Tasks: Evaluated on resting-state fMRI from 6000 UKBioBank and 1075 HCP participants for cognitive score regression and sex classification.

Method	HCP MSE		HCP MAE		HCP R		UKBioBank MSE		UKBioBank MAE		UKBioBank R	
XG-Boost	1.004	0.22	0.831	0.15	0.14	0.05	1.049	0.19	0.811	0.14	0.01	0.03
BrainNetCNN	0.981	0.21	0.799	0.10	0.21	0.02	1.003	0.23	0.801	0.16	0.01	0.01
BrainGNN	0.946	0.02	0.791	0.10	0.28	0.04	0.995	0.13	0.794	0.03	0.06	0.10
BrainNetTF	0.998	0.17	0.820	0.03	0.18	0.10	0.999	0.01	0.798	0.15	0.04	0.05
TFF	0.957	0.10	0.798	0.02	0.27	0.08	0.998	0.02	0.795	0.11	0.04	0.03
SwiFT	0.914	0.11	0.790	0.04	0.32	0.02	0.994	0.16	0.791	0.03	0.07	0.01
BrainMT	0.835	0.02	0.741	0.01	0.41	0.03	0.932	0.03	0.773	0.01	0.24	0.02

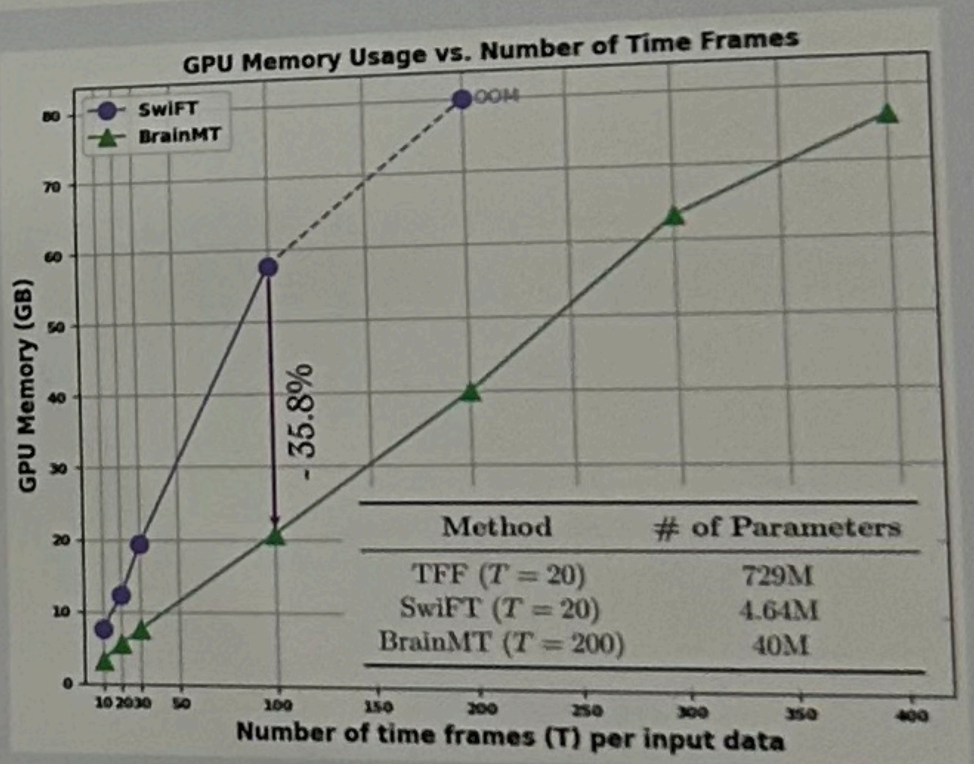
Method	HCP Acc.		HCP B.Acc		HCP AUROC		UKBioBank Acc.		UKBioBank B.Acc		UKBioBank AUROC	
XG-Boost	68.43	2.37	67.86	3.35	73.2	2.43	79.15	1.38	78.27	1.37	86.3	0.42
BrainNetCNN	76.94	3.29	75.41	2.31	82.3	2.35	85.78	0.41	84.82	0.38	92.4	0.32
BrainGNN	84.73	1.22	84.26	0.98	90.1	0.73	89.85	1.28	89.71	0.21	95.8	1.14
BrainNetTF	82.87	2.19	81.55	3.16	89.3	2.09	87.92	1.23	87.38	1.15	95.1	0.71
TFF	92.94	1.17	92.40	1.19	97.1	2.08	96.11	0.13	95.46	0.28	99.3	0.12
SwiFT	93.06	1.08	92.61	1.26	97.5	1.95	97.45	0.11	97.69	0.05	99.4	0.13
BrainMT	96.28	0.02	96.15	0.03	98.7	0.06	97.91	0.09	97.77	0.08	99.2	0.04

Performance: BrainMT outperforms correlation-based and voxel-based baselines, achieving lowest error in regression and highest accuracy in classification tasks.

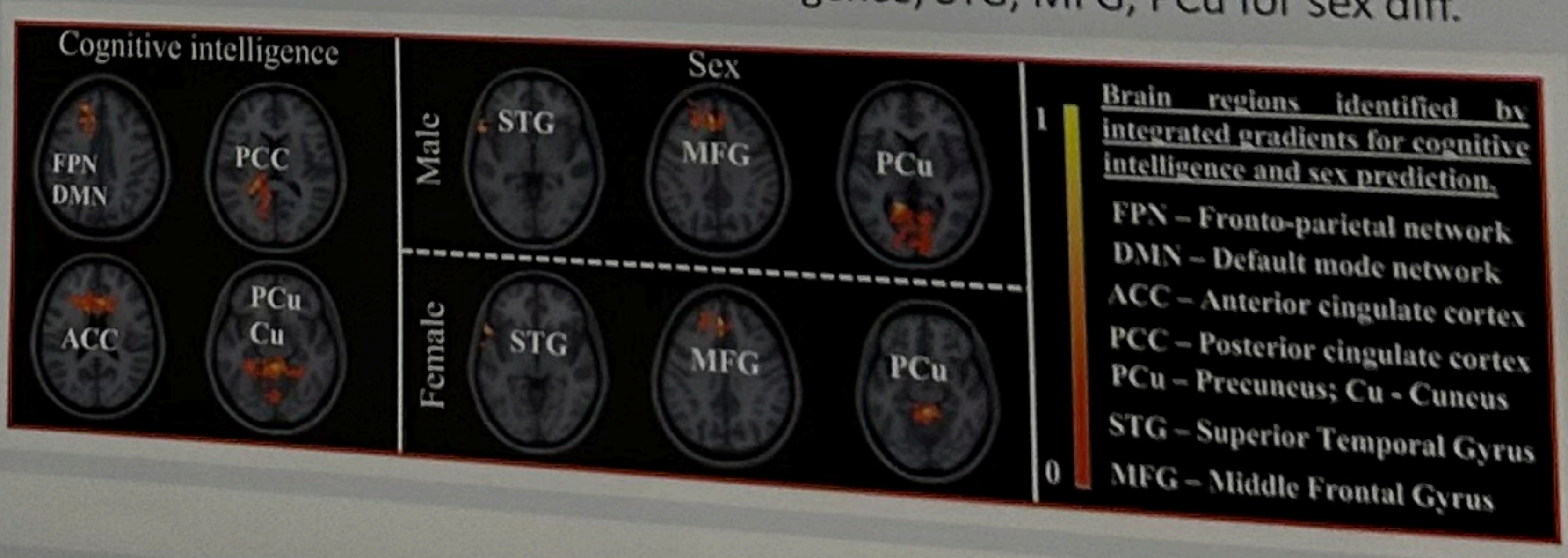
Computational Performance

BrainMT scales efficiently to long fMRI sequences.

Models up to 400 TRs per subject on a single NVIDIA A100 GPU.



Interpretability: Integrated Gradients [4] highlight biologically consistent regions - FPN, DMN, ACC, PCC, PCu and Cu for cognitive intelligence; STG, MFG, PCu for sex diff.



CONCLUSIONS

✓ Key Insight: Modeling voxel-level fMRI data with a hybrid Mamba + Transformer design unlocks richer spatiotemporal patterns than traditional parcellation or short-window approaches.

✓ Impact: BrainMT shows strong generalization across large cohorts for diverse tasks such as classification and regression.

✓ Next Steps: Extend to self-supervised pretraining and clinical tasks (e.g., pain and neurological disorders) to improve robustness and translational value.

Acknowledgements

This work was supported by the National Institutes of Health grants R01 EB029977 (PI Caffo) and R01 EB026549 (PI Lindquist) from the National Institute of Biomedical Imaging and Bioengineering. This work used data from the Human Connectome Project - WU-Minn Consortium, as well as from UK Biobank (Project ID 33278), a major biomedical database.

REFERENCES

[1] Lindquist et al., Annual Review of Statistics, 2025.
[2] Abraham et al., NeuroImage, 2017.
[3] Zhu et al., ECCV, 2024.
[4] Sundararajan et al., ICML, 2017.