

# 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

## 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.

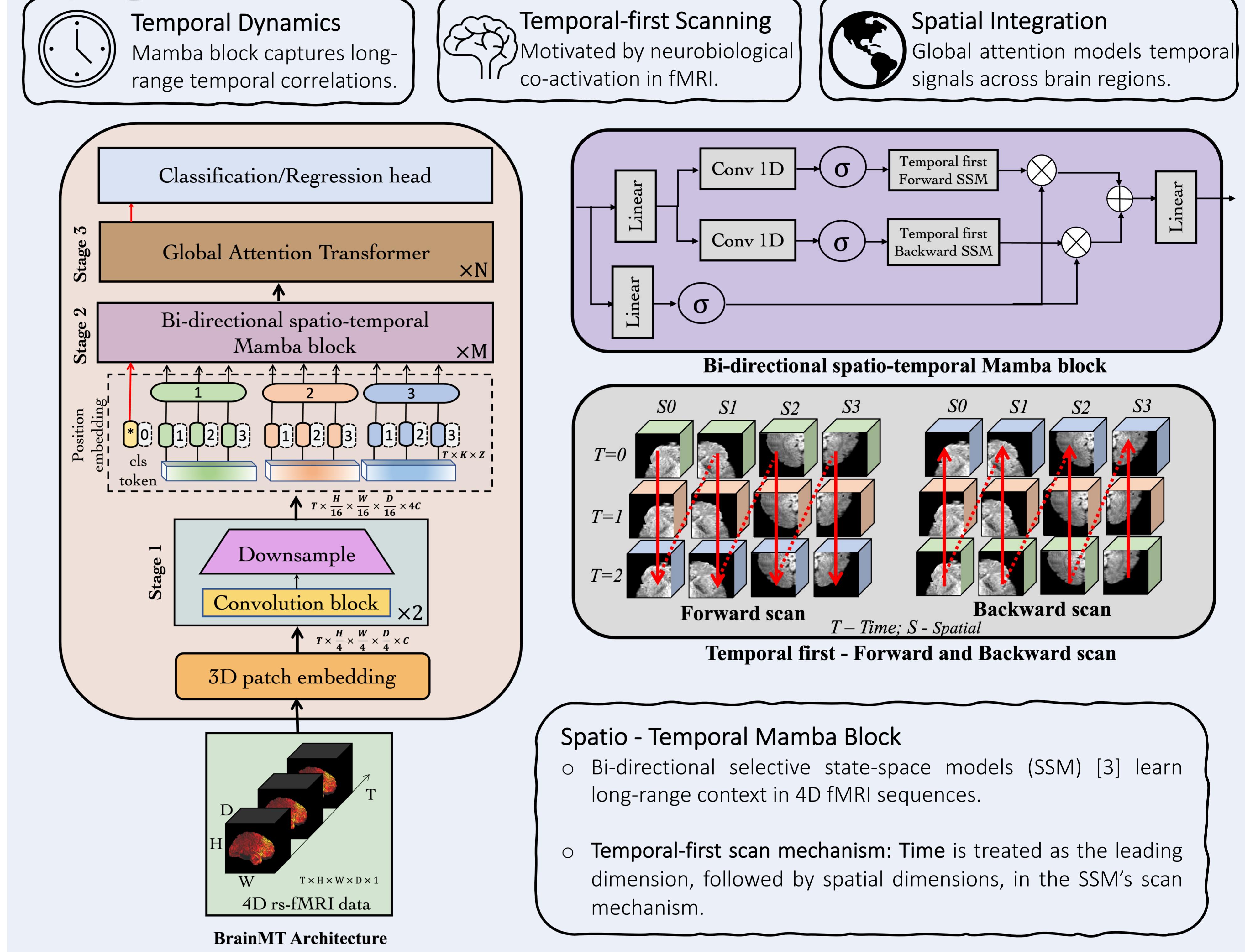
**Challenge: Limits of Current Models**

- 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



## RESULTS

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

Method	HCP			UKBioBank		
	MSE	MAE	R	MSE	MAE	R
XG-Boost	1.004	0.22	0.831	0.15	0.14	0.05
BrainNetCNN	0.981	0.21	0.799	0.10	0.21	0.02
BrainGNN	0.946	0.02	0.791	0.10	0.28	0.04
BrainNetTF	0.998	0.17	0.820	0.03	0.18	0.10
TFF	0.957	0.10	0.798	0.02	0.27	0.08
SwiFT	0.914	0.11	0.790	0.04	0.32	0.02
<b>BrainMT</b>	<b>0.835</b>	<b>0.02</b>	<b>0.741</b>	<b>0.01</b>	<b>0.41</b>	<b>0.03</b>

Method	HCP			UKBioBank		
	Acc.	B.Acc	AUROC	Acc.	B.Acc	AUROC
XG-Boost	68.43	2.37	67.86	3.35	73.2	2.43
BrainNetCNN	76.94	3.29	75.41	2.31	82.3	2.35
BrainGNN	84.73	1.22	84.26	0.98	90.1	0.73
BrainNetTF	82.87	2.19	81.55	3.16	89.3	2.09
TFF	92.94	1.17	92.40	1.19	97.1	2.08
SwiFT	93.06	1.08	92.61	1.26	97.5	1.95
<b>BrainMT</b>	<b>96.28</b>	<b>0.02</b>	<b>96.15</b>	<b>0.03</b>	<b>98.7</b>	<b>0.06</b>

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

## 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.

## 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.

## 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.

