NeuroBOLT: Resting-state EEG-to-fMRI Synthesis with Multi-dimensional Feature Mapping

Yamin Li [1] Ange Lou [1] Ziyuan Xu [1] Shengchao Zhang [1] Shiyu Wang [1]
Dario J. Englot [2] Soheil Kolouri [1] Daniel Moyer [1] Roza G. Bayrak [1] Catie Chang [1]
[1] Vanderbilt University [2] Vanderbilt University Medical Center

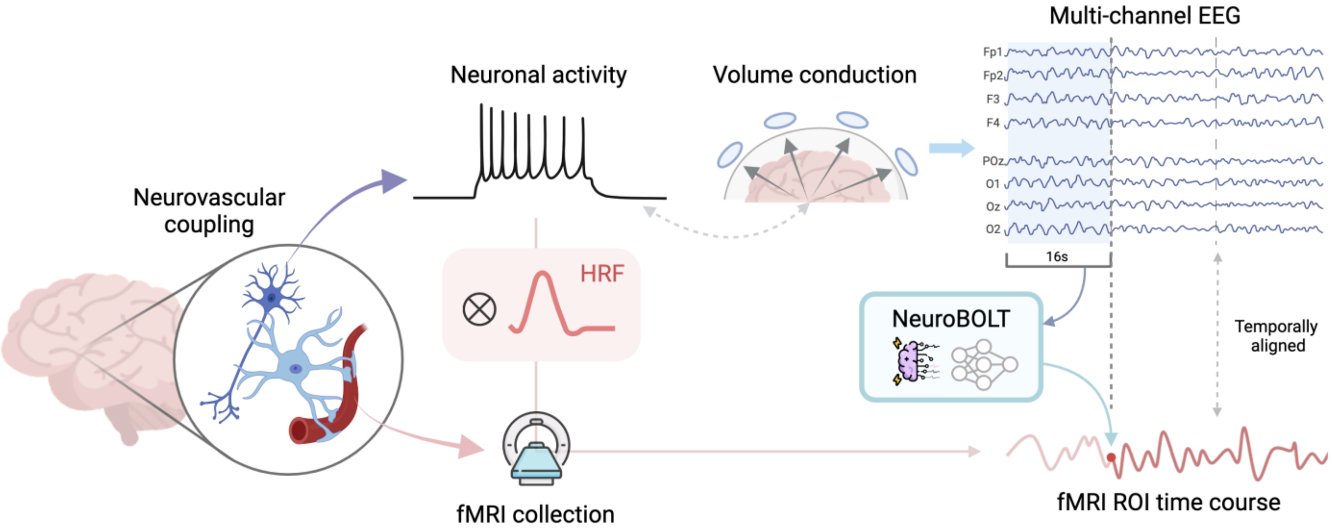
DOI: 10.5555/3737916.3738652
NIPS '24: Proceedings of the 38th International Conference on Neural Information Processing Systems
Article No.: 736, Pages 23378 - 23405
Published: 05 June 2025

Why?

- fMRI = high spatial resolution, slow, expensive, not portable.
- EEG = high temporal resolution, portable, cheap, but spatially fuzzy.
- Challenge: Can EEG approximate fMRI signals?
- Potential impact: accessible brain mapping for research & clinics.

How does it work?

- Input: 16s window of multi-channel EEG.
- Output: Predicted fMRI signal at a specific Region of Interest .
- Task: Sequence-to-one regression.
- Training: All ROI (64 ROIs in DiFuMo atlas).



Overall illustration of EEG-to-BOLD fMRI translation using NeuroBOLT

Existing approaches

- Correlation-based studies: link EEG rhythms (e.g., alpha power) with fMRI activity, but only partial correlations, not full synthesis
- Early machine learning: ridge regression on EEG temporal–spectral features to reconstruct fMRI signals, but limited to visual cortex or specific subcortical regions
- Deep learning (Seq-to-Seq, CNNs, etc.): used encoder–decoder or transformer frameworks to map EEG → fMRI
  - Some success in reconstructing fMRI time series in deep brain regions.
  - Focused mainly on task datasets (e.g., eye opening/closing, auditory cues).

Limitations

❌ Mostly task-based, resting-state EEG-fMRI largely unexplored (too noisy/random).
❌ Subject-specific models: trained & tested on same individual, poor generalization.
❌ Restricted to a few brain regions only.
❌ Often rely on fixed hemodynamic assumptions (delay between EEG and fMRI).

Core idea: capture time, space, and frequency simultaneously.
Input: EEG windows (multi-channel, 16s).
Output: fMRI signal at ROI p and time t.
Model decomposes into:

- Spatiotemporal representation module.
- Multi-scale spectral representation module.
- Fusion + regression head.

EEG sequence split into patches (time slices × channels).
Each patch → projected to embedding vector $x \in \mathbb{R}^d$.

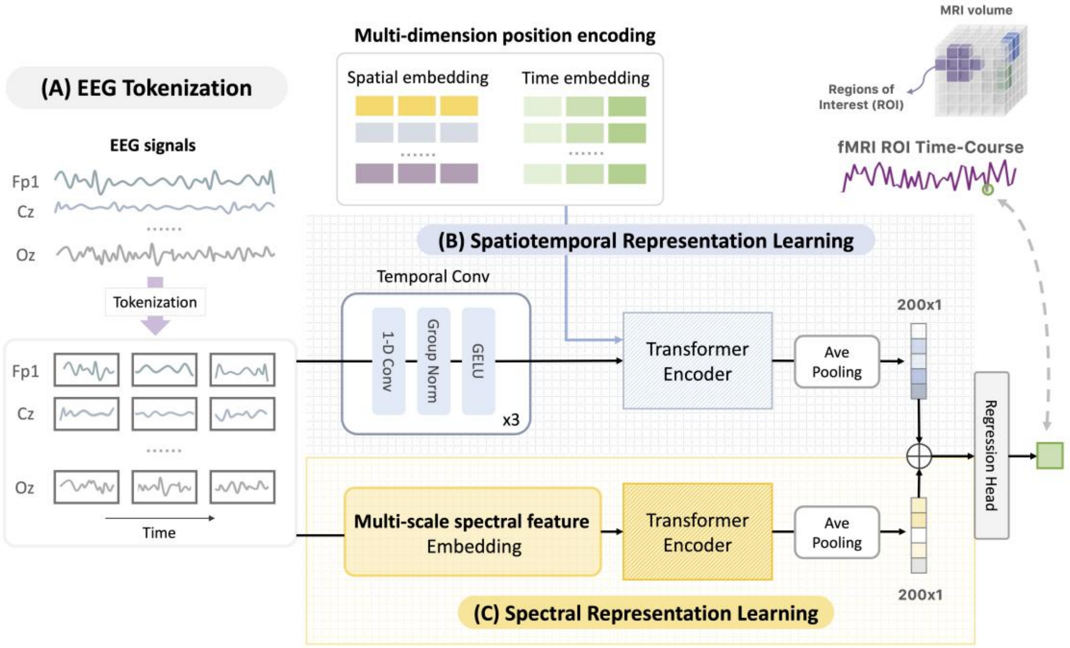
- Positional embeddings: encode temporal order.
- Channel embeddings: encode electrode identity (spatial context).

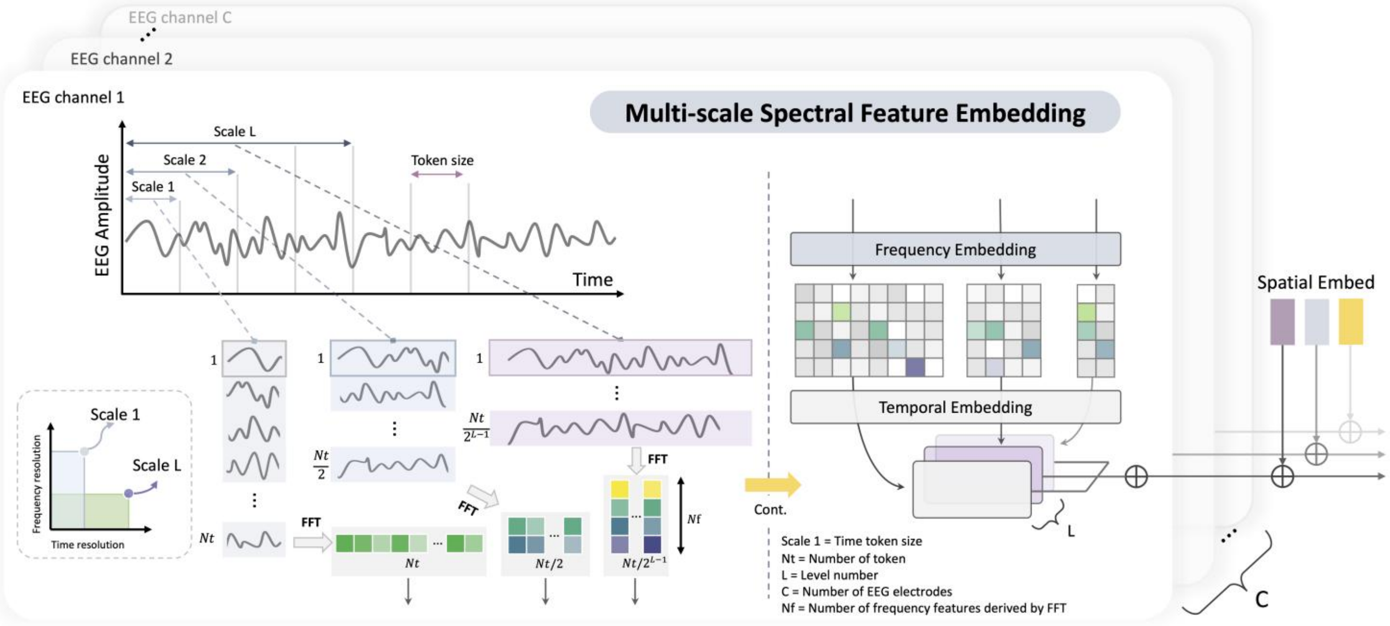
$$z_i = W \cdot \text{patch}_i + e_{time}(i) + e_{channel}(c)$$

patch embeddings enriched with temporal + spatial tokens.
Processed by multi-head self-attention transformer encoder:

$$h = \text{TransformerEncoder}(z_1, \ldots, z_n)$$

Captures dependencies across time (temporal dynamics) and channels (spatial distribution).
Output: spatiotemporal feature vector r_{st}.





EEG transformed into spectrograms via STFT.
Multi-scale windows:

- Small window → high temporal, low frequency resolution.
- Large window → low temporal, high frequency resolution.

Frequency embeddings added to encode spectral position.
Encoded by linear transformer (efficient attention scaling).
Output: spectral feature vector r_{sp}.

$$S(f,t) = \sum_{n} x[n] \cdot w(t-n) \cdot e^{-j2\pi fn/N}$$

Fusion head:

$$r = r_{st} + r_{sp}$$

Passed through regression head:

- GELU activation.
- Linear projection to scalar.

Final output: predicted fMRI at ROI p and time t.

$$\hat{Y}_{p,t} = W_2 \cdot \mathrm{GELU}(W_1 r)$$

Training Setup

- Dataset:
  - Resting-state dataset: 24 subjects, 20min each
  - Auditory task dataset (to evaluate the generalization performance): 10 subjects, 16 scans
- 26 EEG channels, 200 Hz sampling.
- ROIs: 64 (DiFuMo atlas).
- Optimizer: AdamW (lr = 3e-4, weight decay = 0.05).
- Batch size: 16 (intra-subject), 64 (inter-subject).

Evaluation Metrics

- Pearson correlation coefficient (R): the strength and direction of the linear relationship between prediction and ground truth.
- Mean squared error (MSE): the average squared differences between prediction and ground truth across samples

Results

	Model	Primary Sensory		High-level Cognitive		Subcortical		-	Avg. R↑
		Cuneus	Heschl's Gyrus	Middle Frontal	Precuneus Anterior	Putamen	Thalamus	Global Signal	
Intra-scan	BIOT [59]	0.531±0.223	0.518±0.207	0.490±0.162	0.459±0.110	0.410±0.205	0.411±0.231	0.493±0.133	0.473
	LaBraM [22]	0.540±0.176	0.519±0.197	0.493±0.153	0.490±0.176	0.411±0.179	0.449±0.177	0.487±0.167	0.484
	BEIRA [25]	0.357±0.241	0.396±0.240	0.294±0.228	0.320±0.220	0.234±0.194	0.328±0.197	0.456±0.240	0.341
	Li, et al. [31]	0.460±0.228	0.515±0.207	0.376±0.169	0.457±0.204	0.324±0.183	0.398±0.194	0.583±0.170	0.445
	NeuroBOLT (ours)	**0.588±0.166**	**0.566±0.183**	**0.502±0.168**	**0.559±0.141**	**0.437±0.184**	**0.480±0.213**	**0.587±0.162**	**0.531**
Inter-subject	FFCL [29]	0.326±0.094	0.412±0.039	0.327±0.078	0.437±0.091	0.243±0.125	0.373±0.082	0.512±0.048	0.376
	CNN Transformer [44]	0.218±0.204	0.412±0.114	0.298±0.097	0.316±0.153	0.232±0.086	0.180±0.106	0.282±0.185	0.273
	STT Transformer [48]	0.269±0.197	0.188±0.056	0.226±0.130	0.280±0.143	0.074±0.126	0.142±0.101	0.347±0.124	0.218
	BIOT [59]	0.457±0.123	0.512±0.039	0.393±0.128	0.445±0.084	0.299±0.063	0.413±0.073	0.529±0.110	0.435
	LaBraM [22]	0.177±0.116	0.211±0.105	0.153±0.132	0.170±0.152	0.047±0.111	0.147±0.122	0.150±0.152	0.151
	BEIRA [25]	0.421±0.112	0.482±0.063	0.384±0.147	0.452±0.149	0.241±0.135	0.410±0.097	0.492±0.106	0.412
	Li, et al. [31]	**0.505±0.063**	0.430±0.048	0.415±0.114	0.416±0.076	0.217±0.139	0.424±0.072	0.529±0.092	0.419
	NeuroBOLT (ours)	0.482±0.100	**0.561±0.046**	**0.423±0.115**	**0.496±0.136**	**0.335±0.144**	**0.453±0.106**	**0.564±0.115**	**0.473**

Table 1: Model performance (R) in intra- and inter-subject experiments. Bold: the best performance, the underlined: the second-best performance

Training	Testing	Primary Sensory		High-level Cognitive		Subcortical		-	Avg. R↑
		Cuneus	Heschl's Gyrus	Middle Frontal	Precuneus Anterior	Putamen	Thalamus	Global Signal	
RS	AT	0.387±0.087	0.431±0.026	0.419±0.099	0.451±0.050	0.240±0.202	0.361±0.164	0.372±0.087	0.380
AT	AT	0.428±0.141	0.479±0.084	0.407±0.058	0.460±0.071	0.187±0.253	0.362±0.166	0.287±0.120	0.373
RS-p+AT-f	AT	0.446±0.033	**0.547±0.060**	**0.437±0.089**	0.471±0.065	0.241±0.188	**0.401±0.177**	0.385±0.098	0.418
RS+AT	AT	**0.461±0.101**	0.516±0.044	0.434±0.106	**0.476±0.041**	**0.248±0.194**	0.401±0.220	**0.404±0.070**	**0.420**
RS	RS	**0.482±0.100**	**0.561±0.046**	0.423±0.115	**0.496±0.136**	**0.335±0.144**	**0.453±0.106**	**0.564±0.115**	**0.473**
RS+AT	RS	0.478±0.110	0.560±0.049	**0.437±0.086**	0.494±0.107	0.330±0.140	0.443±0.074	0.540±0.119	0.469

Performance of NeuroBOLT in inter-subject prediction in resting-state and auditory task fMRI. Mean R values between prediction and g.t. are shown. RS: Resting-State, AT: Auditory Task, RS-p+AT-f: Pretraining on RS and finetuning on AT, RS+AT: joint training of RS and AT

Results

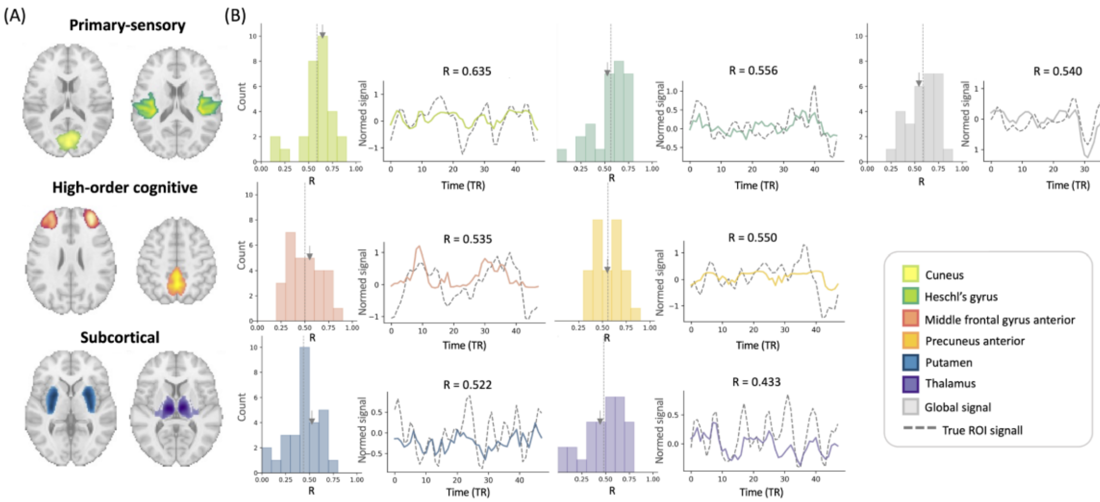


Figure 4: Intra-subject prediction results.

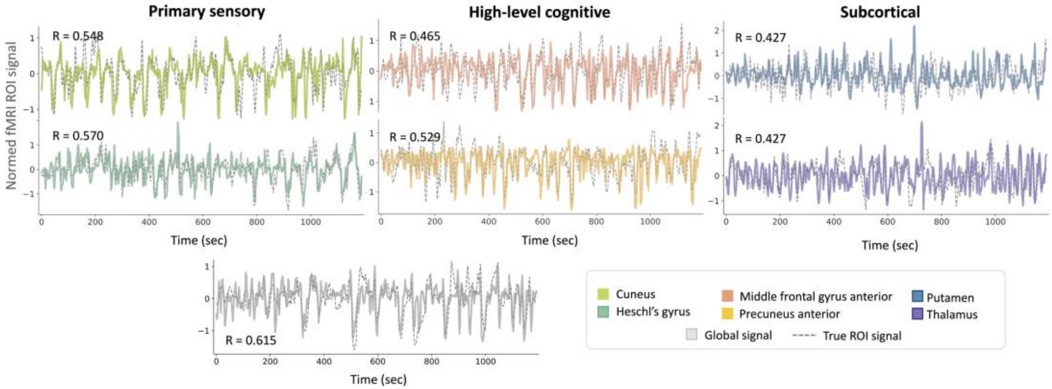


Figure 5: Examples of reconstruction of the unseen scans.

Contributions

- Outperforms baselines on both sensory and cognitive regions.
- Predicts fMRI time series from held-out subjects across entire 20-minute scans, which is a first in this field.
- Handles deep structures (thalamus, putamen).
- Works across subjects → good for clinical use.

Limitations

❌ Model is trained separately for each ROI (not whole brain jointly).
❌ Needs validation on more diverse datasets (different tasks, populations)..

Thanks

Any questions?