NEURAL INFORMATION PROCESSING SYSTEMS

浙江大學 ZHEJIANG UNIVERSITY

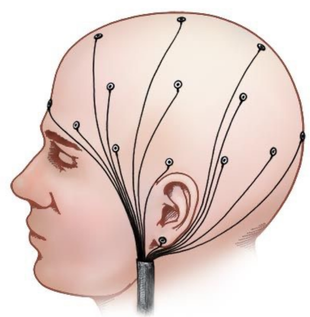
Brant: Foundation Model for Intracranial Neural Signal

Daoze Zhang[1*], Zhizhang Yuan[1*], Yang Yang[1†], Junru Chen[1], Jingjing Wang[1], Yafeng Li[2]

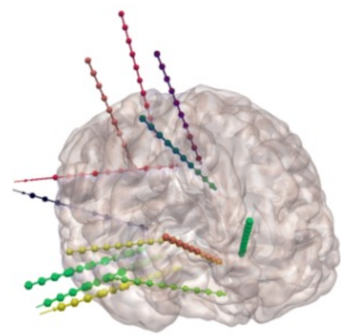
[1]Zhejiang University [2]Nuozhu Technology Co., Ltd

Background

- **Brain Signal**
  - Electrical impulses that are generated by brain neurons, providing important information about brain activity.
- **Two Monitoring Ways**
  - Scalp EEG: through electrodes placed on the scalp.
  - **Intracranial EEG**: through intracranial electrodes that implants into brain tissue directly.



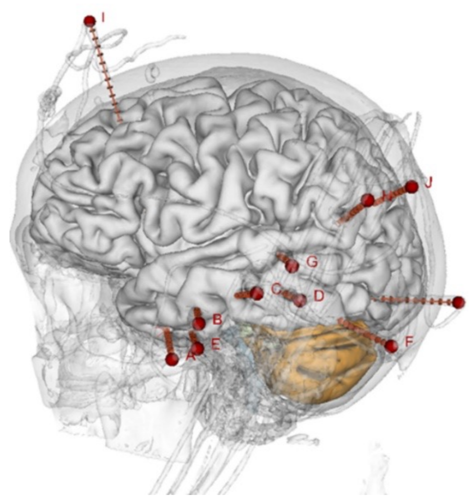
Scalp EEG

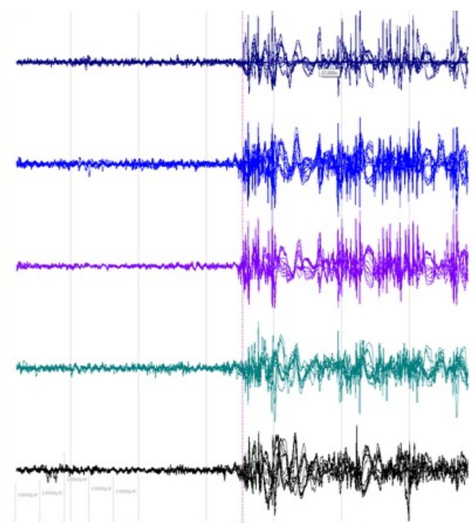


Intracranial EEG

Background

- **Intracranial Neural Signal (iEEG)**
  - Recorded by **deep** electrodes inside human brains.
  - Provide **stereotactic** information from deeper brain structures.
  - Furnish **more abundant and detailed** analysis about brain wave patterns.



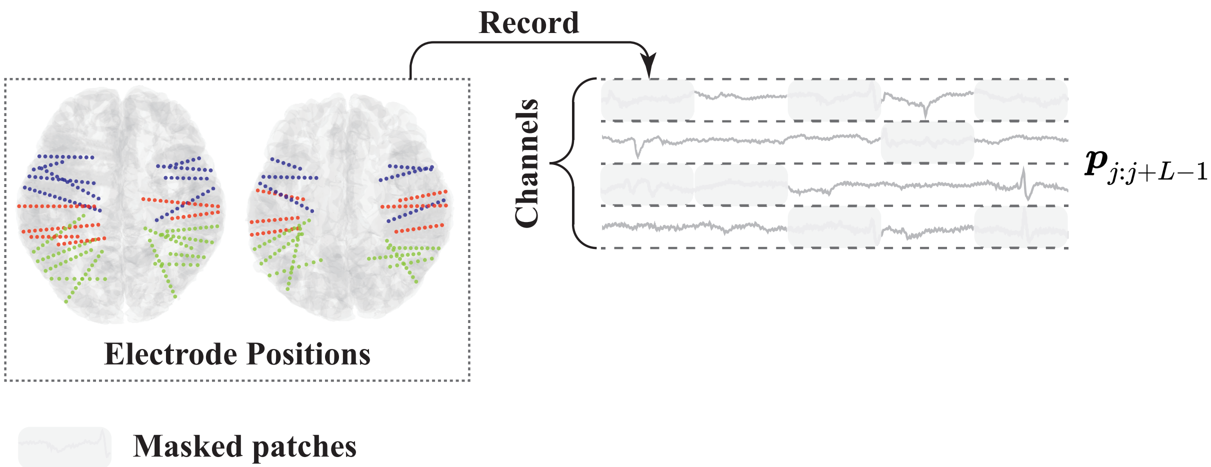


Modeling Intracranial Signal: Insights

- **Long-term Dependency**
  - Gradual changes in brain activity may only be captured by the long-period analysis.
- **Spatial Correlation**
  - Due to the fact that brain waves propagate through different brain regions, signals recorded from **different channels can be spatially correlated.**
- **Time and Frequency Domains**
  - Time domain: information about the amplitude and duration.
  - Frequency domain: underlying **oscillatory patterns and rhythms.**

Brant

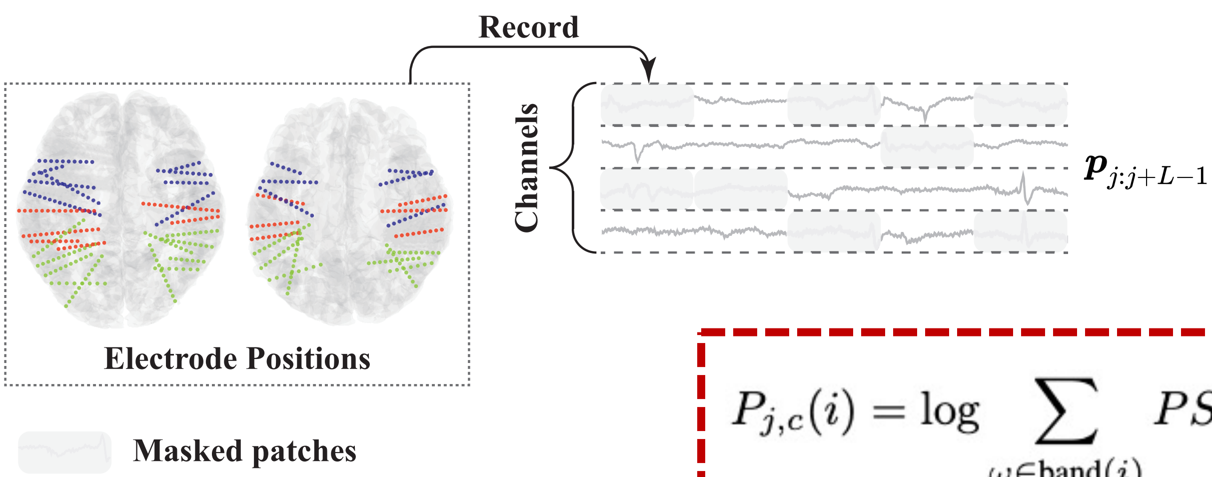
- ❑ **Patching**
- ❑ **Randomly Masking**
  - As the pre-training task is mask-reconstruction



Brant

❑ Frequency Encoding

- To add frequency domain information to the encoding
- The frequency encoding $\mathbf{F}_{j,c}$ of patch $\boldsymbol{p}_{j,c}$ is obtained as the **weighted sum of the learnable encodings** of each frequency band.



$$P_{j,c}(i) = \log \sum_{\omega \in \text{band}(i)} PSD_{p_{j,c}}(\omega), \quad i \in \{1, 2, \ldots, 8\},$$

$$\mathbf{F}_{j,c} = \sum_{i=1}^{8} \frac{\exp(P_{j,c}(i))}{\sum_{i'=1}^{8} \exp(P_{j,c}(i'))} \boldsymbol{f}_i.$$

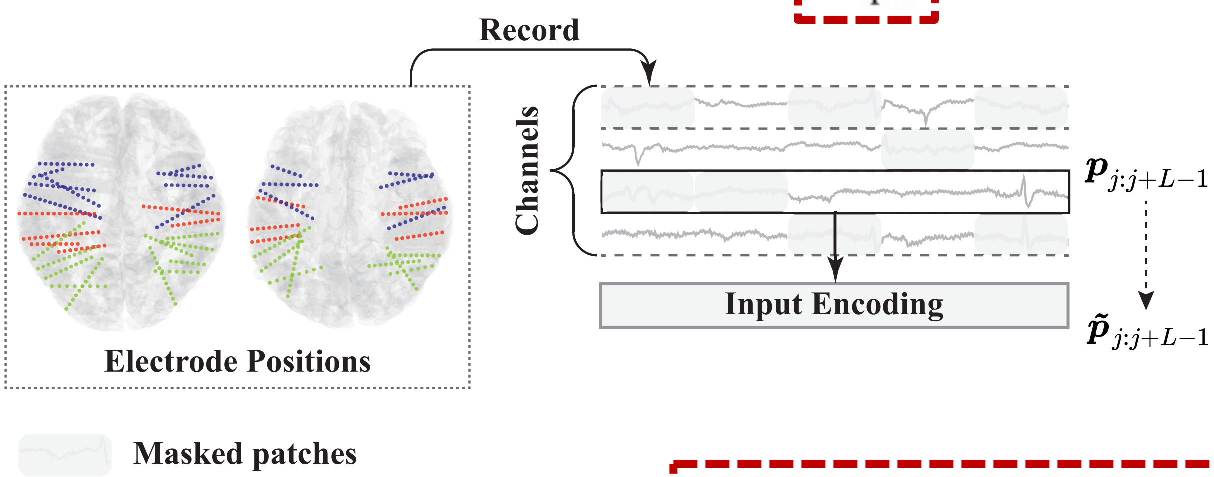
Brant

$\mathbf{W}_{\text{proj}}\boldsymbol{p}^{\mathrm{T}}_{j:j+L-1,c}$

❑ Input Encoding

– For each sequence of patches $\boldsymbol{p}_{j:j+L-1,c}$, we use a linear projection and add the **positional encoding** and the **frequency encoding**

$\mathbf{W}_{\text{pos}}$ $\mathbf{F}_{j:j+L-1,c}$

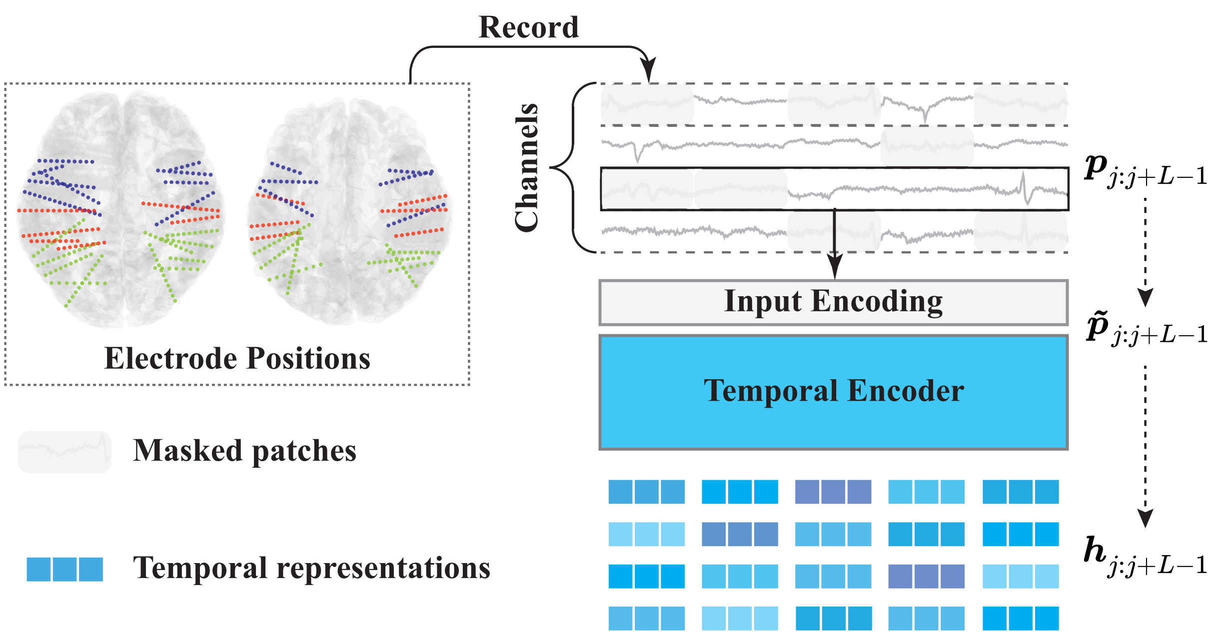


$$\tilde{\boldsymbol{p}}_{j:j+L-1,c} = (\mathbf{W}_{\text{proj}}\boldsymbol{p}^{\mathrm{T}}_{j:j+L-1,c})^{\mathrm{T}} + \mathbf{W}_{\text{pos}} + \mathbf{F}_{j:j+L-1,c},$$

Brant

❑ Temporal Encoder: Long-term Dependency

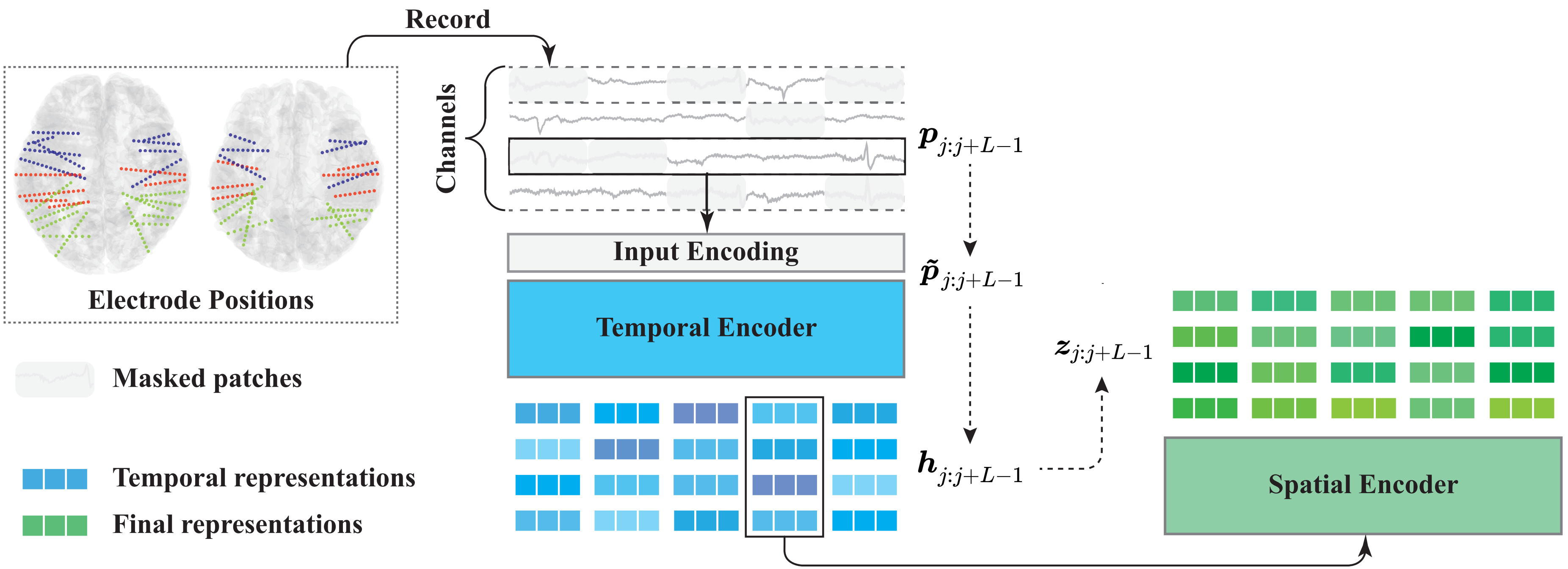
– The input encoding will be fed into the temporal encoder to obtain temporal hidden representations $\boldsymbol{h}_{j:j+L-1}$.



Brant

❑ Spatial Encoder: Channel Spatial Correlation

– The spatial encoder further captures the spatial correlation across channels, which outputs the final representations $\boldsymbol{z}_{j:j+L-1}$.

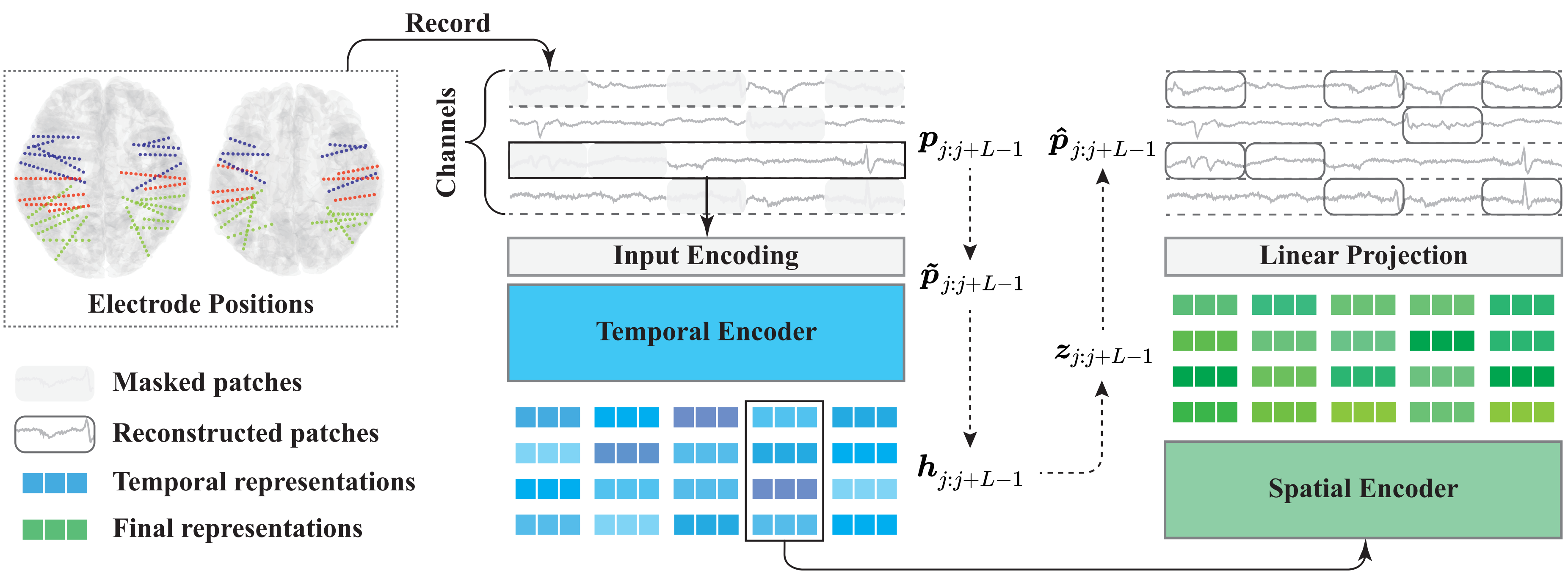


Brant

❑ SSL Task: Mask-Reconstruction

– The final representations will be fed into a flatten layer with linear head to get the reconstructed patches $\hat{\boldsymbol{p}}_{j:j+L-1}$.

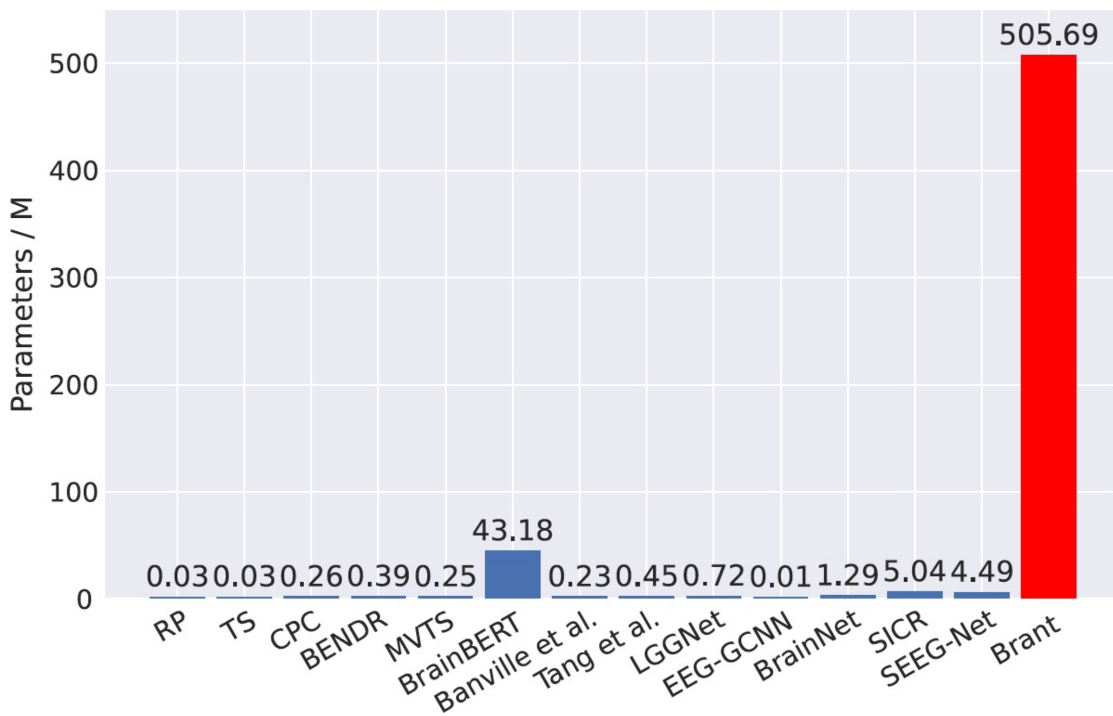
– The loss is calculated between $\boldsymbol{p}_{j:j+L-1}$ and $\hat{\boldsymbol{p}}_{j:j+L-1}$.



Brant

❑ The Largest Model on Brain Signals

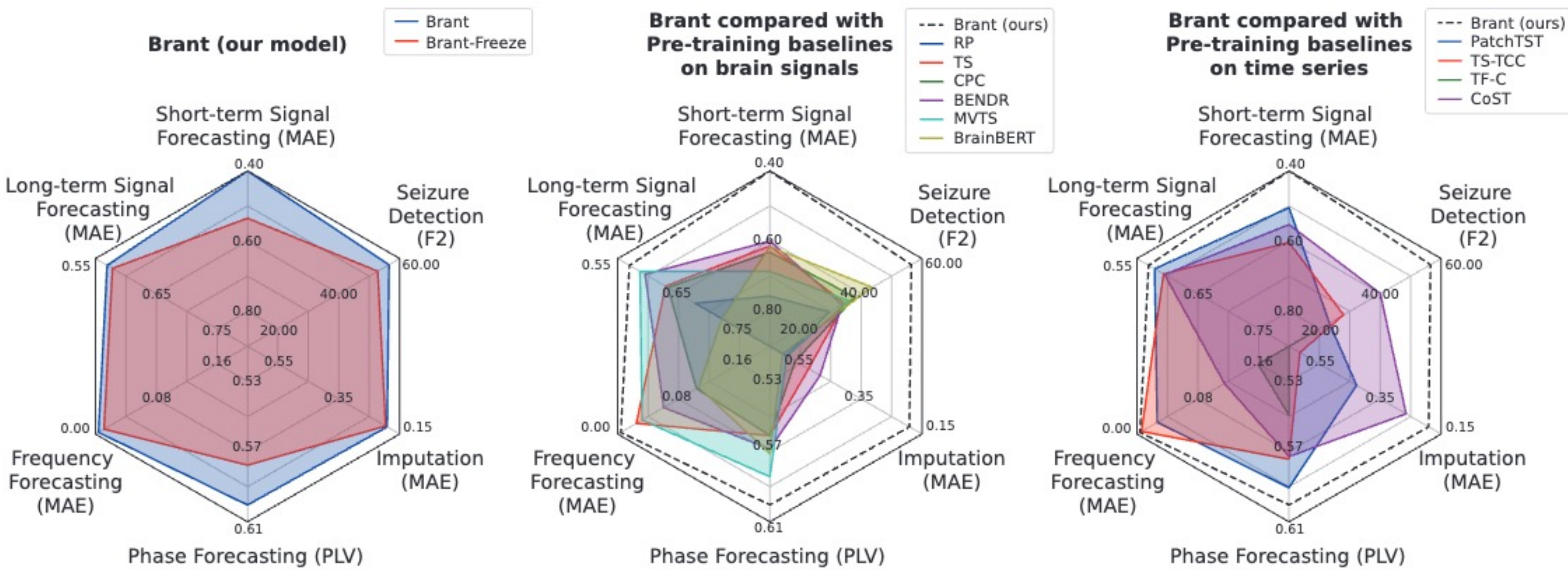
– Using the method above, Brant is pre-trained on a large intracranial dataset with 1.01 TB data, containing more than 500M parameters.



Experiments

❑ Overall Performance

– Brant achieves **consistent SOTA performance on a variety of tasks** compared with other baselines.



Please refer to the paper for detailed results.

Model Analysis

❑ Low-resource Labeled Data Evaluation

– In medical scenarios, collecting labeled data is a huge investment…

Model	200 minutes	60 minutes		20 minutes	
	F2	F2	Decrease	F2	Decrease
SEEG-Net [9]	*42.28±1.10	35.54±1.90	15.94%	12.76±2.13	69.82%
RP [17]	29.59±1.97	27.62±2.03	*6.66%	25.05±1.98	15.34%
TS [17]	34.57±1.66	30.15±3.05	12.79%	29.61±3.34	*14.35%
CPC [17]	37.96±1.42	30.55±3.01	19.52%	29.57±3.74	22.10%
BENDR [18]	33.77±1.81	25.37±3.12	24.87%	22.18±4.09	34.32%
MVTS [19]	35.90±1.94	26.62±3.11	25.85%	24.39±4.01	32.06%
BrainBERT [20]	43.60±0.98	41.93±2.09	3.84%	36.35±3.23	16.63%
PatchTST [27]	23.27±1.26	18.02±2.23	22.55%	17.07±2.11	26.64%
TS-TCC [39]	27.91±1.19	25.35±2.07	9.17%	20.36±1.90	27.05%
TF-C [38]	19.02±1.24	15.97±1.23	16.04%	13.66±2.10	28.18%
CoST [37]	40.03±1.88	*39.18±3.02	**2.12%**	36.10±4.12	9.82%
Brant	**56.50**±1.08	**52.30**±2.04	7.43%	**51.03**±2.74	**9.68%**

Model Analysis

❑ Low-resource Labeled Data Evaluation

– In medical scenarios, collecting labeled data is a huge investment…

Model	200 minutes	60 minutes		20 minutes	
	F2	F2	Decrease	F2	Decrease
SEEG-Net [9]	*42.28±1.10	35.54±1.90	15.94%	12.76±2.13	69.82%
RP [17]	29.59±1.97	27.62±[illegible]	[illegible]	[illegible]	[illegible]
TS [17]	34.57±1.66	30.15±[illegible]	[illegible]	[illegible]	[illegible]
CPC [17]	37.96±1.42	30.55±3.01	19.52%	29.57±3.74	22.10%
BENDR [18]	33.77±1.81	25.37±3.12	24.87%	22.18±4.09	34.32%
MVTS [19]	35.90±1.94	26.62±3.11	25.85%	24.39±4.01	32.06%
BrainBERT [20]	43.60±0.98	41.93±2.09	3.84%	36.35±3.23	16.63%
PatchTST [27]	23.27±1.26	18.02±2.23	22.55%	17.07±2.11	26.64%
TS-TCC [39]	27.91±1.19	25.35±2.07	9.17%	20.36±1.90	27.05%
TF-C [38]	19.02±1.24	15.97±1.23	16.04%	13.66±2.10	28.18%
CoST [37]	40.03±1.88	*39.18±3.02	**2.12%**	36.10±4.12	9.82%
Brant	**56.50**±1.08	**52.30**±2.04	7.43%	**51.03**±2.74	**9.68%**

Brant maintains the most stable performance on 20-min labeled data.

Model Analysis

❑ Low-resource Labeled Data Evaluation

– In medical scenarios, collecting labeled data is a huge investment…

Model	200 minutes	60 minutes		20 minutes	
	F2	F2	Decrease	F2	Decrease
SEEG-Net [9]	*42.28				
RP [17]	29.59				
TS [17]	34.57				
CPC [17]	37.96±1.42	30.55±3.01	19.52%	29.57±3.74	22.10%
BENDR [18]	33.77±1.81	25.37±3.12	24.87%	22.18±4.09	34.32%
MVTS [19]	35.90±1.94	26.62±3.11	25.85%	24.39±4.01	32.06%
BrainBERT [20]	43.60±0.98	41.93±2.09	3.84%	36.35±3.23	16.63%
PatchTST [27]	23.27±1.26	18.02±2.23	22.55%	17.07±2.11	26.64%
TS-TCC [39]	27.91±1.19	25.35±2.07	9.17%	20.36±1.90	27.05%
TF-C [38]	19.02±1.24	15.97±1.23	16.04%	13.66±2.10	28.18%
CoST [37]	40.03±1.88	*39.18±3.02	**2.12%**	36.10±4.12	9.82%
Brant	**56.50**±1.08	**52.30**±2.04	7.43%	**51.03**±2.74	**9.68%**

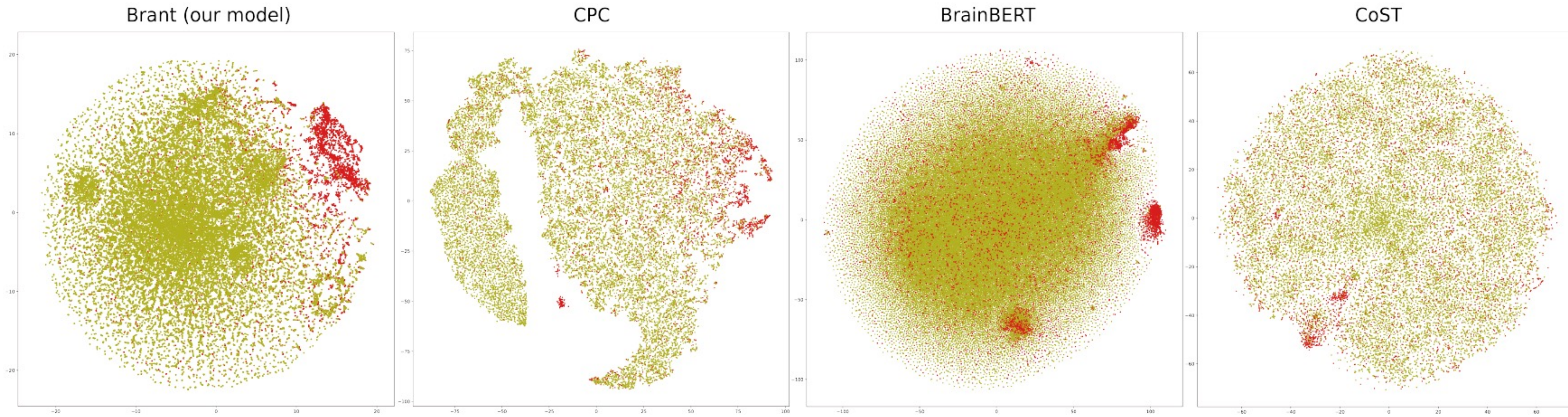
F2 score of our model on 20-min labeled data is even higher than that of the best baseline on 200-min labeled data.

Model Analysis

❑ Representation Analysis

– We visualize the pre-trained representations of Brant and three most representative methods using t-SNE.

Compared to other methods, the representations of seizure and normal signals learned from Brant are separated more clearly.

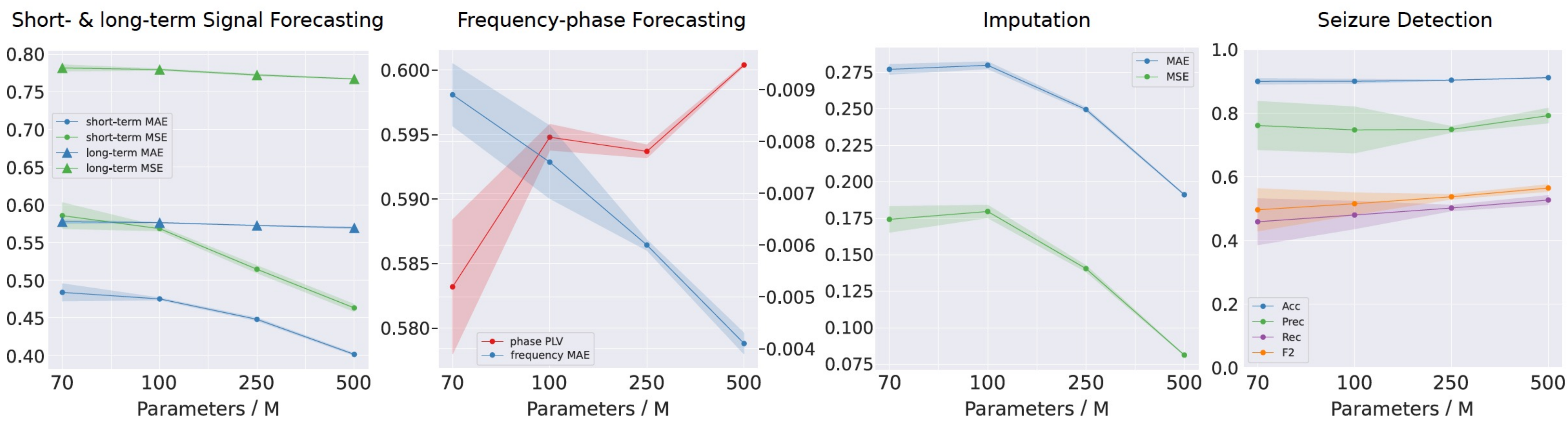


Model Analysis

❑ Model Scale Analysis

– As the model size increases,

- the performances show an overall upward trend, indicating that **a larger model with a higher capacity results in better ability.**
- the decrease in the standard deviation indicates **more stable performance for larger models.**



Conclusions

- ❑ We propose a task-agnostic foundation model, Brant, which is **the largest** pre-training model on brain signals.
- ❑ Experimentally, Brant achieves **consistent SOTA** performance on various downstream tasks w.r.t. medical scenarios.
- ❑ Brant is an **off-the-shelf** model with its code and weights, which can directly participate in other medical research and treatment.

THANKS

More relevant research of our group: ***http://yangy.org***

Contact: ***zhangdz@zju.edu.cn; yangya@zju.edu.cn***