Paper ID: 1348

# MBrain: A Multi-channel Self-Supervised Learning Framework for Brain Signals

Donghong Cai[1*], Junru Chen[1*], Yang Yang[1†], Teng Liu[1], Yafeng Li[2]

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

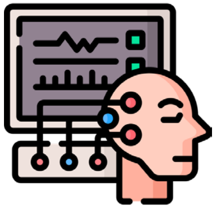
jrchen_cali@zju.edu.cn

yangya@zju.edu.cn

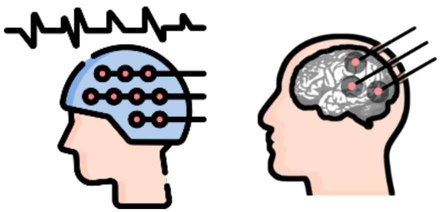
Group Homepage

# Background

Brain signals are **foundational** quantitative data for the study of human brain



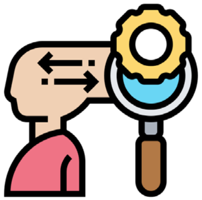
Brain signals can be measured by **various** methods



The patterns of brain signals help us to **understand** the brain functions

**Cognitive science**

The scientific investigation of the mind and intelligence



**Emotion recognition**

The process of identifying human emotion

**Neurological disorders**

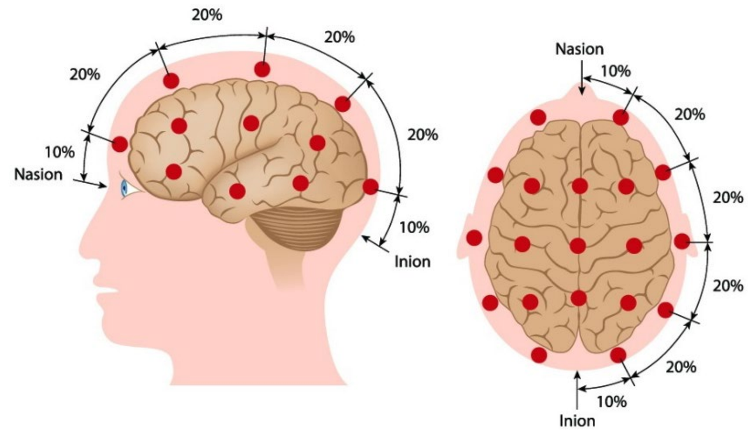
The diseases of the central and peripheral nervous system

# Background

## ❑ Non-invasive Methods (EEG)

– Easy to implement **without** any surgery.

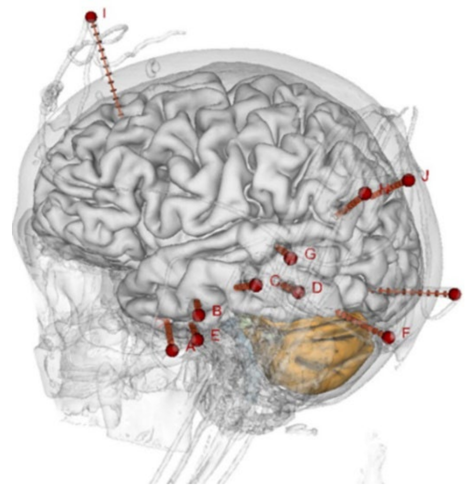
– **Cannot** simultaneously consider temporal and spatial resolution along with the deep brain information.



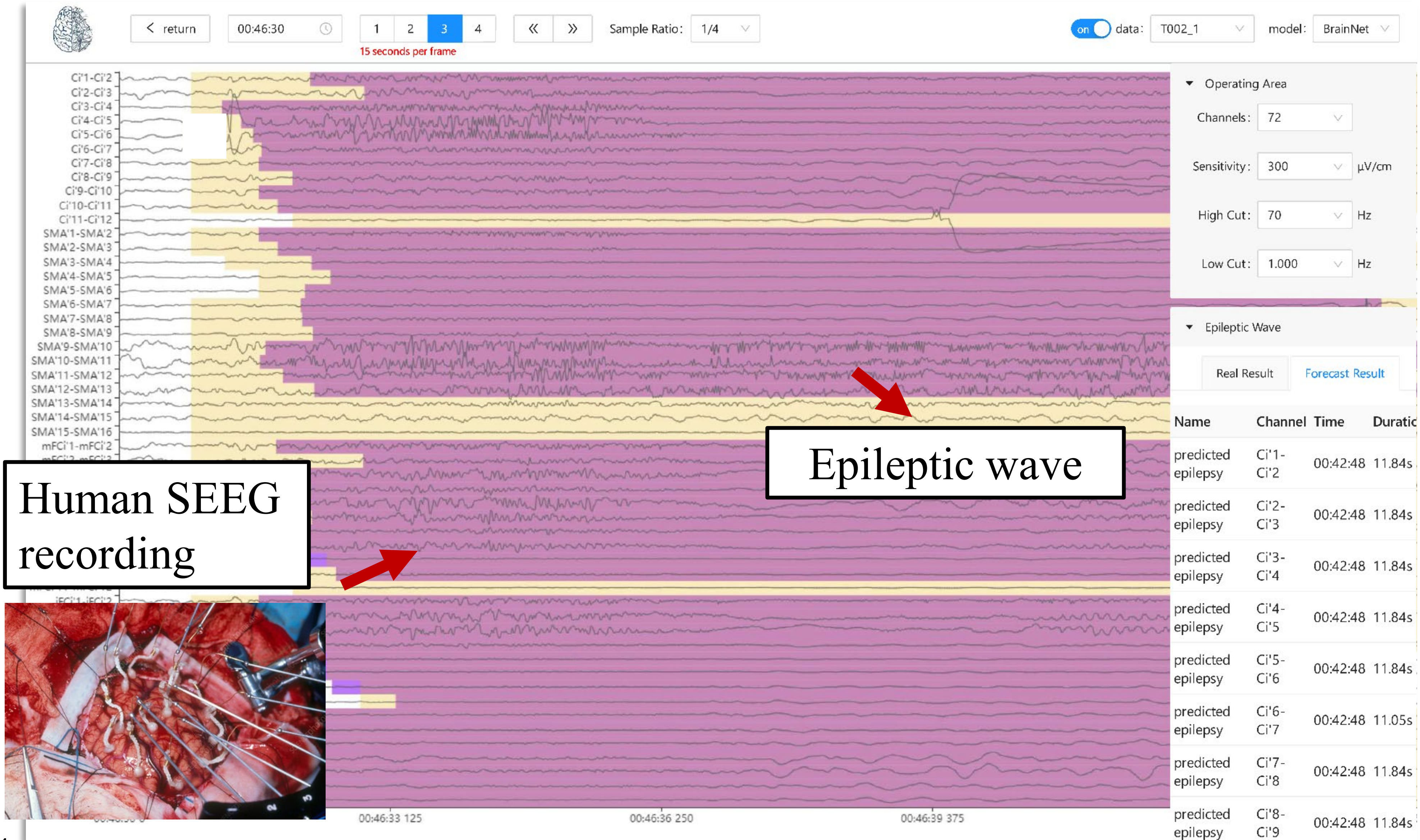

## ❑ Invasive Methods (SEEG)

– Require extra **surgeries** to insert the recording devices.

– Have access to **more** precise and **higher** signal-to-noise data.



# Background

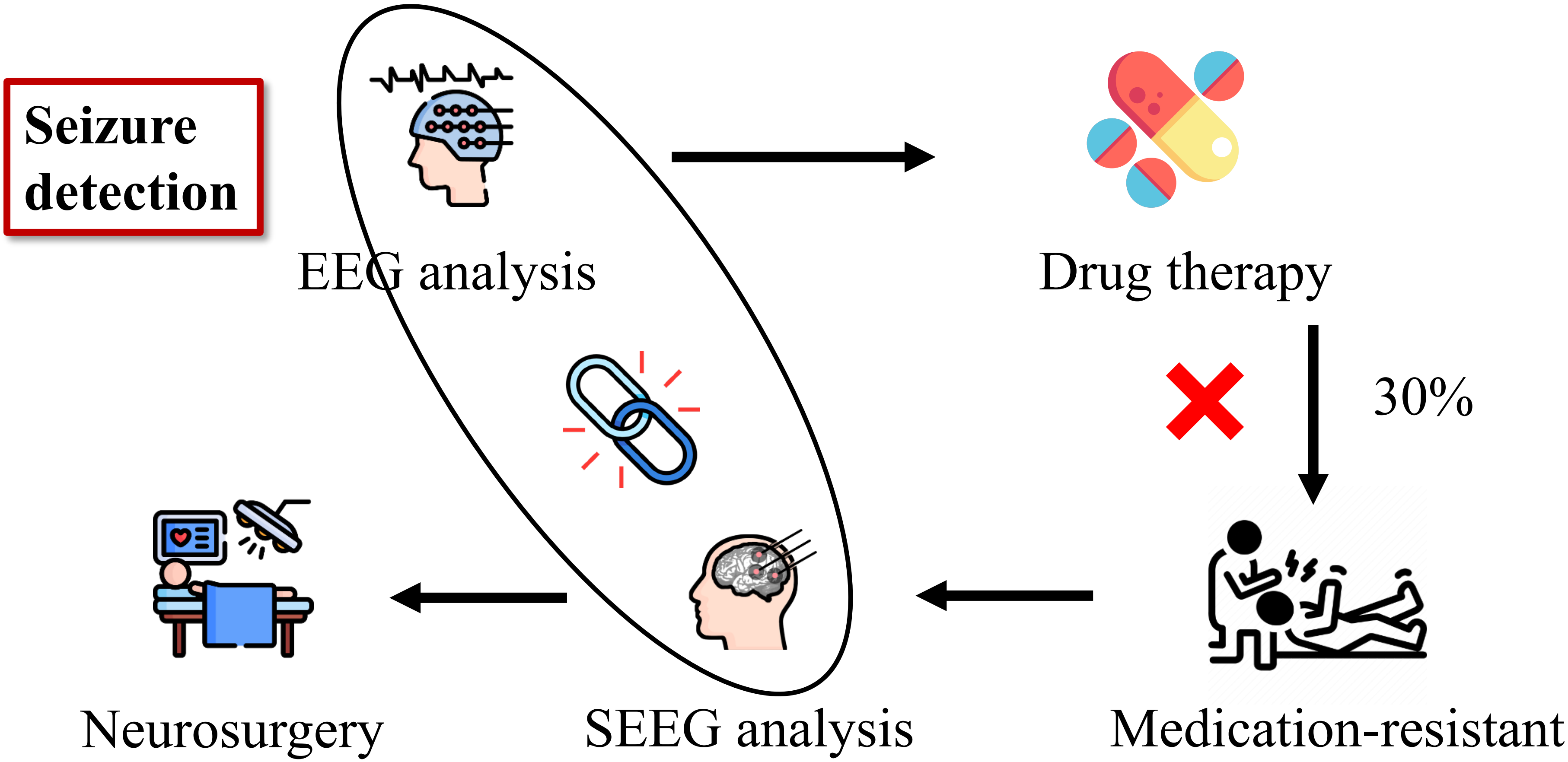


return
00:46:30
1
2
3
4
Sample Ratio: 1/4
15 seconds per frame
on
data: T002_1
model: BrainNet
Operating Area
Channels: 72
Sensitivity: 300 μV/cm
High Cut: 70 Hz
Low Cut: 1.000 Hz
Epileptic Wave
Real Result
Forecast Result
Name
Channel
Time
Duratic
predicted epilepsy Ci'1-Ci'2 00:42:48 11.84s
predicted epilepsy Ci'2-Ci'3 00:42:48 11.84s
predicted epilepsy Ci'3-Ci'4 00:42:48 11.84s
predicted epilepsy Ci'4-Ci'5 00:42:48 11.84s
predicted epilepsy Ci'5-Ci'6 00:42:48 11.84s
predicted epilepsy Ci'6-Ci'7 00:42:48 11.05s
predicted epilepsy Ci'7-Ci'8 00:42:48 11.84s
predicted epilepsy Ci'8-Ci'9 00:42:48 11.84s
00:46:33 125
00:46:36 250
00:46:39 375
Epileptic wave
Human SEEG recording

# Motivation

❑ **Why do we want to model EEG and SEEG signals uniformly?**

- Share similar physiological mechanisms
- Closely related in healthcare applications

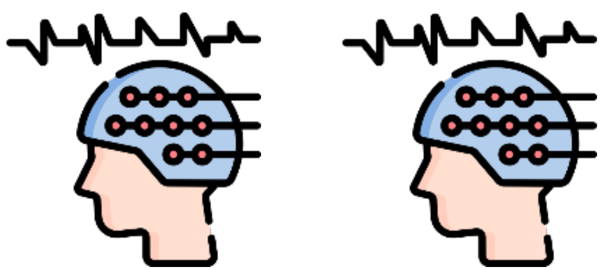


# Challenges

## ❑ Lack of a unified method for handling both signals.

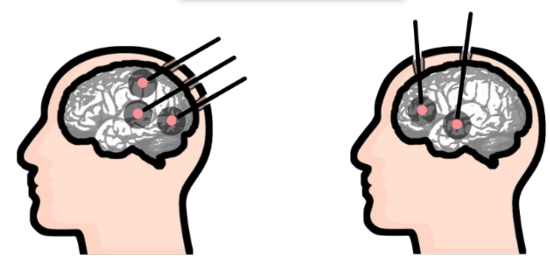
- Varying monitoring location for SEEG.
- Different signal patterns for EEG and SEEG signals.

**EEG**



- ➢ A gold-standard collection location
- ➢ Collect noisy and rough scalp signals

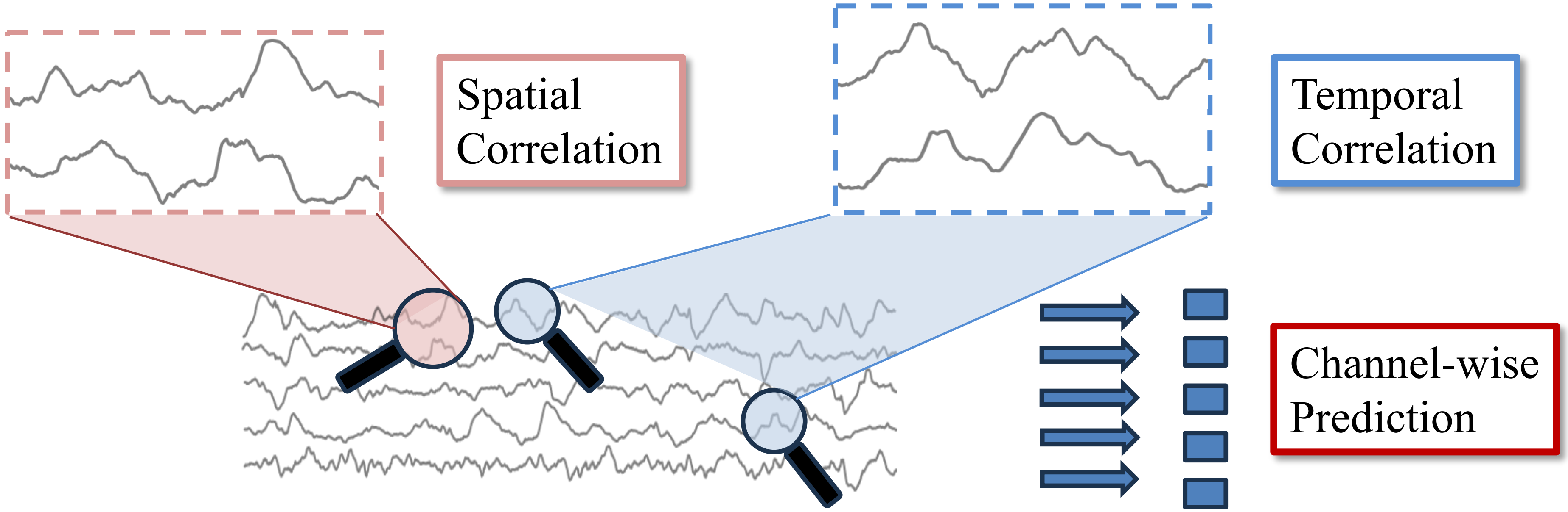
**SEEG**



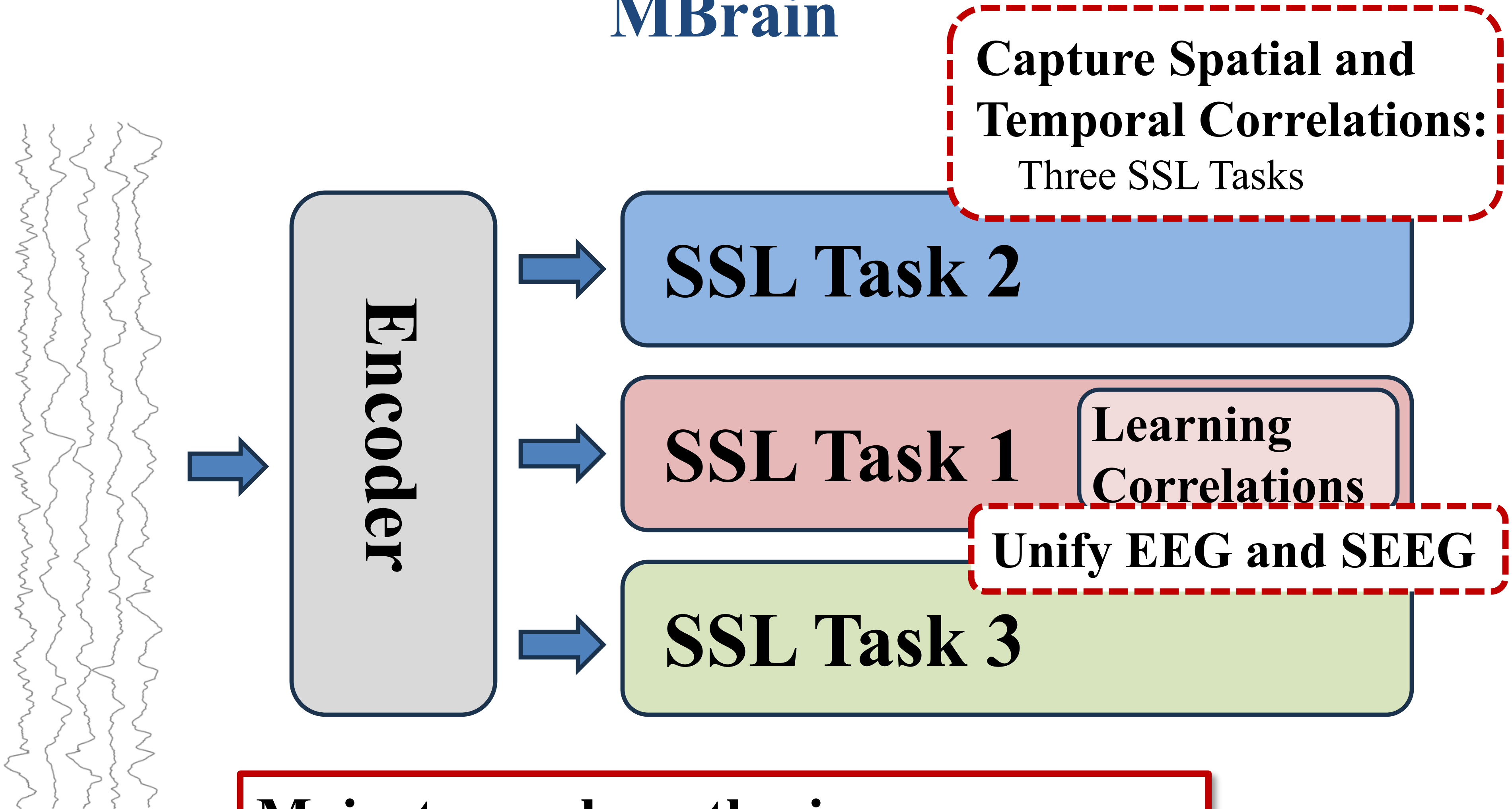
- ➢ Different number and position
- ➢ Collect more stereo and deeper signals

# Challenges

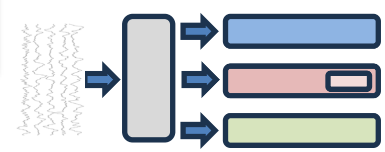
- **Lack of a unified method for handling both signals.**
  - Varying monitoring location for SEEG.
  - Different signal patterns for EEG and SEEG signals.
- **A gap between existing methods and applications.**
  - Capture the spatial and temporal correlations
  - Give channel-wise prediction



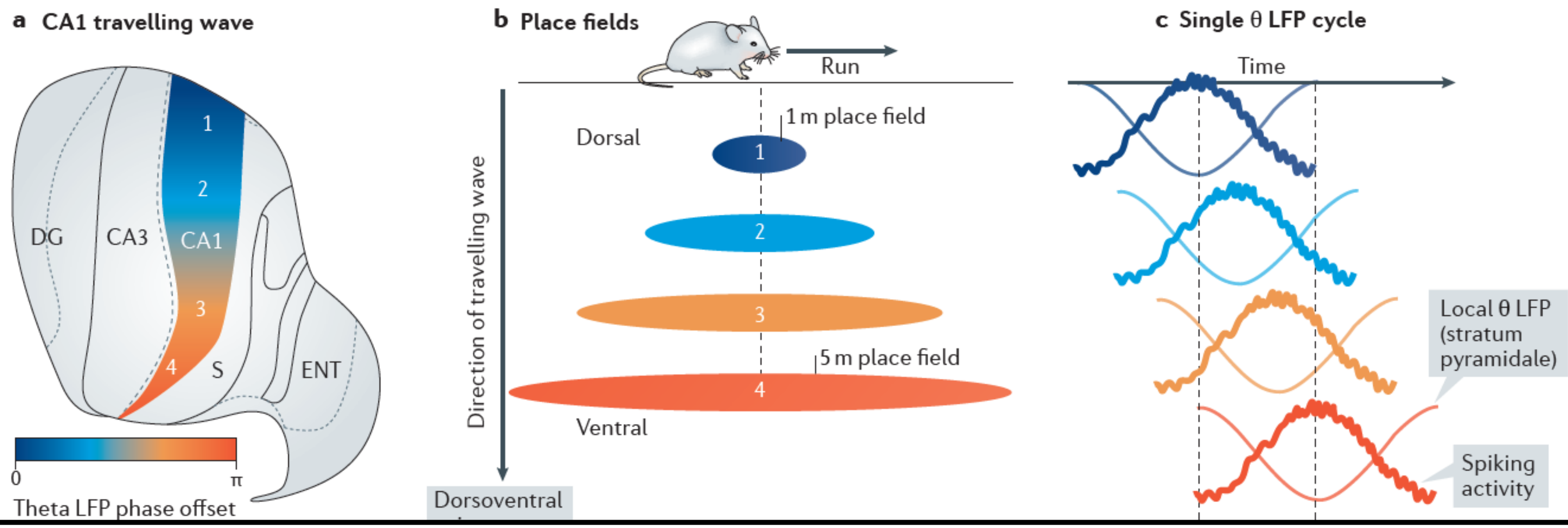
# MBrain



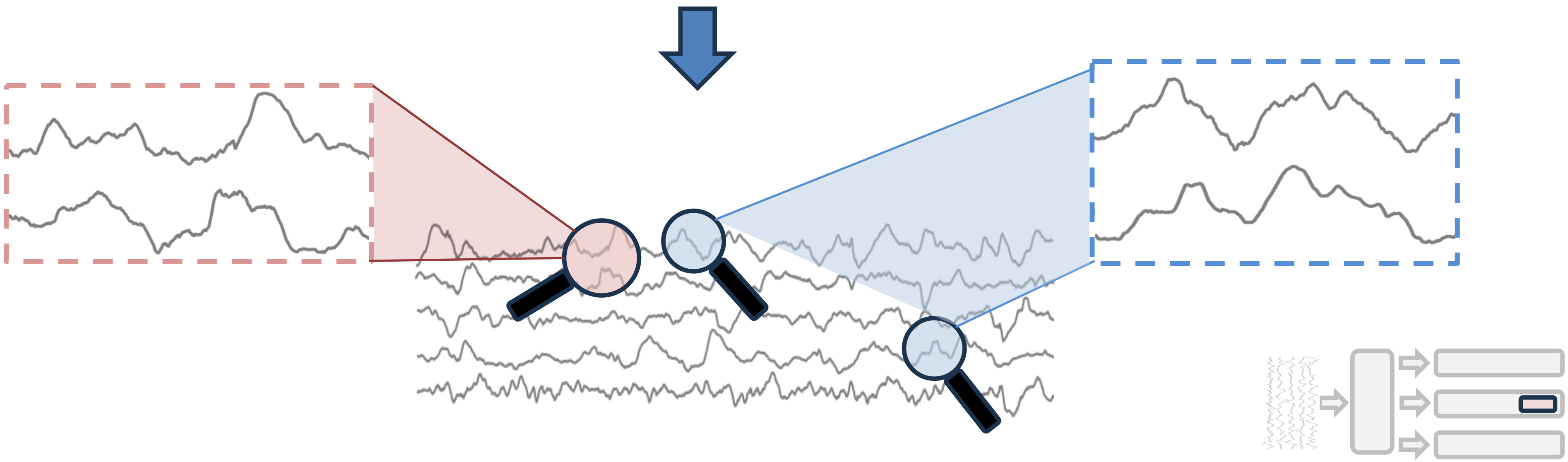

**Mainstream hypothesis:**
The synergistic effects between different brain regions reflect different brain functions.



# How to Model Brain Signals Uniformly?



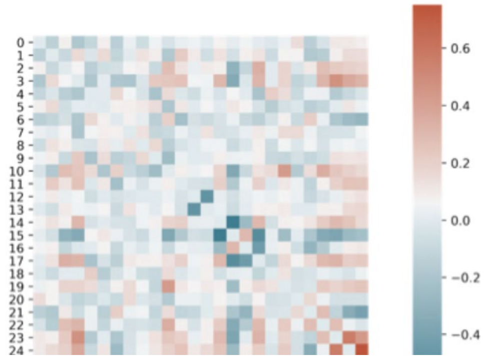

Muller, L., Chavane, F., Reynolds, J. *et al.* Cortical travelling waves: mechanisms and computational principles. *Nat Rev Neurosci* **19**, 255–268 (2018).

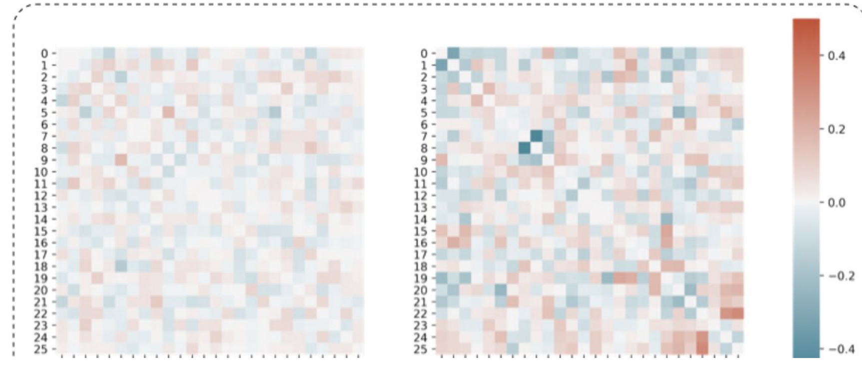


# How to Model Brain Signals Uniformly?

Normal correlation matrix

SEEG



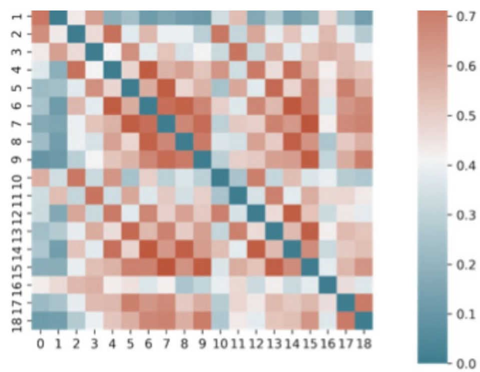


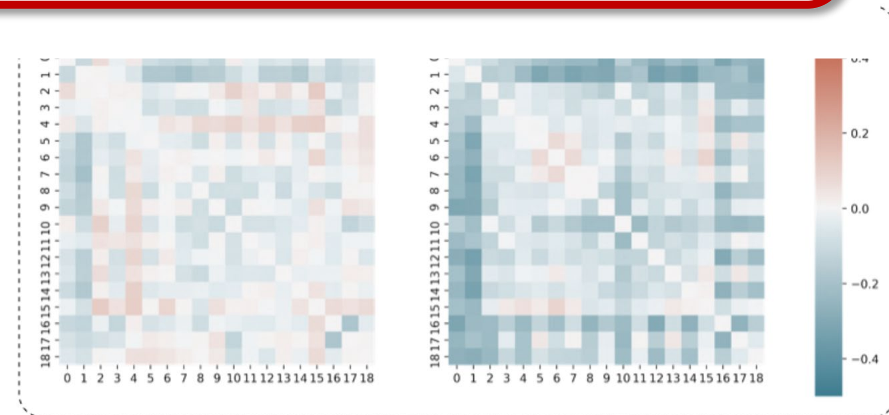
**Conclusion:**

The correlation patterns can help distinguish different brain states.

Seizure correlation matrix

EEG





# How to Model Brain Signals Uniformly?

$$\sigma_t(i,j) = \mathrm{SoftPlus}(\mathrm{MLP}(c^{\mathrm{self}}_{t,\tau,i}, c^{\mathrm{self}}_{t,\tau,j})),$$
$$n_t(i,j) \sim \mathcal{N}(0,1),$$
$$\mathrm{A}^{\mathrm{fine}}_t(i,j) = \mathrm{A}^{\mathrm{coarse}}(i,j) + \sigma_t(i,j) \times n_t(i,j).$$

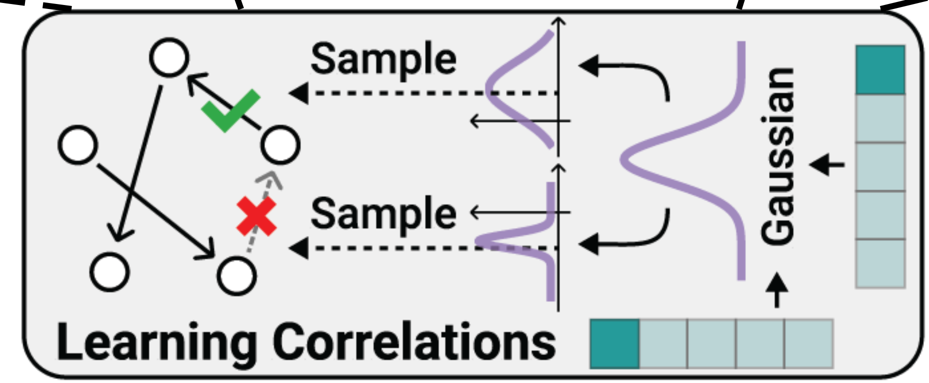
**Filter spurious correlations**

$$\mathrm{A}_t(i,j) = \begin{cases} \mathrm{A}^{\mathrm{fine}}_t(i,j), & \mathrm{A}^{\mathrm{fine}}_t(i,j) \geq \theta_1, \\ 0, & \mathrm{A}^{\mathrm{fine}}_t(i,j) < \theta_1. \end{cases}$$

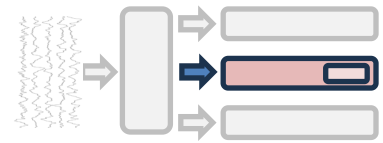
**Sample correlations**

**Coarse-grained correlation graph**

$$\mathrm{A}^{\mathrm{coarse}}(i,j) = \mathbb{E}_{s_t}[\mathrm{Cosine}(s_{t,i}, s_{t,j})],$$

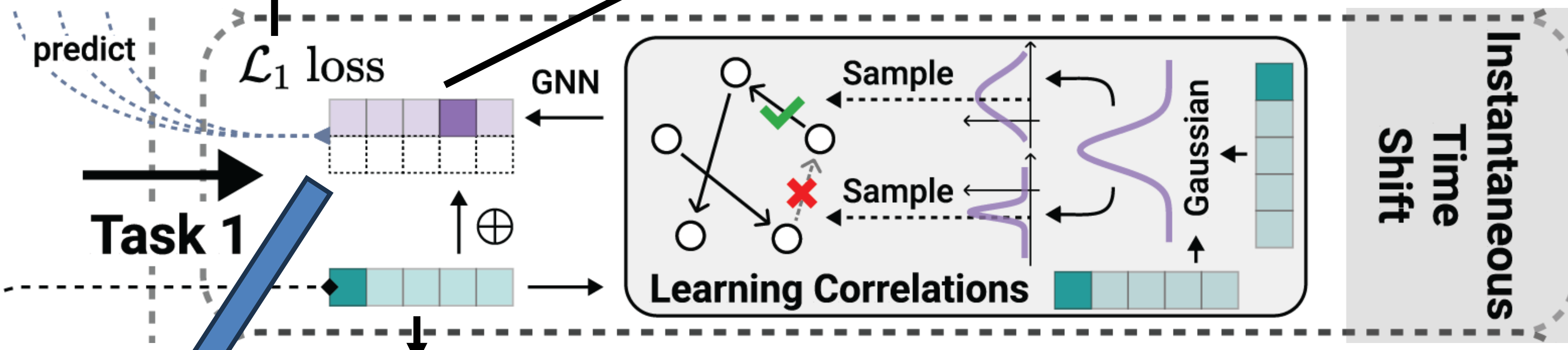


# How to capture spatial correlation patterns?



$$c_{t,\tau} = \mathrm{Concat}(c_{t,\tau}^{\mathrm{self}}, c_{t,\tau}^{\mathrm{other}}),$$

$$\mathcal{L}_1 = \mathcal{L}_N = -\mathbb{E}_{t,i,k_1}\left[\log \frac{c_{t,\tau,i}^{\top} W_{k_1} z_{t,\tau+k_1,i}}{\sum_{z_j \in X_t^i} c_{t,\tau,i}^{\top} W_{k_1} z_j}\right], \qquad c_{t,\tau,i}^{\mathrm{other}} = \mathrm{ReLU}\left(\frac{\sum_{j\neq i} \mathbf{A}_t(i,j)\cdot c_{t,\tau,j}^{\mathrm{self}}}{\sum_{j\neq i} \mathbf{A}_t(i,j)} \cdot \Theta\right),$$




$$c_{t,\tau}^{\mathrm{self}} = g_{\mathrm{ar}}(z_{t,1}, \cdots, z_{t,\tau}).$$

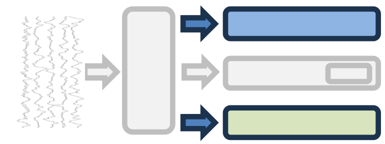
Proposition

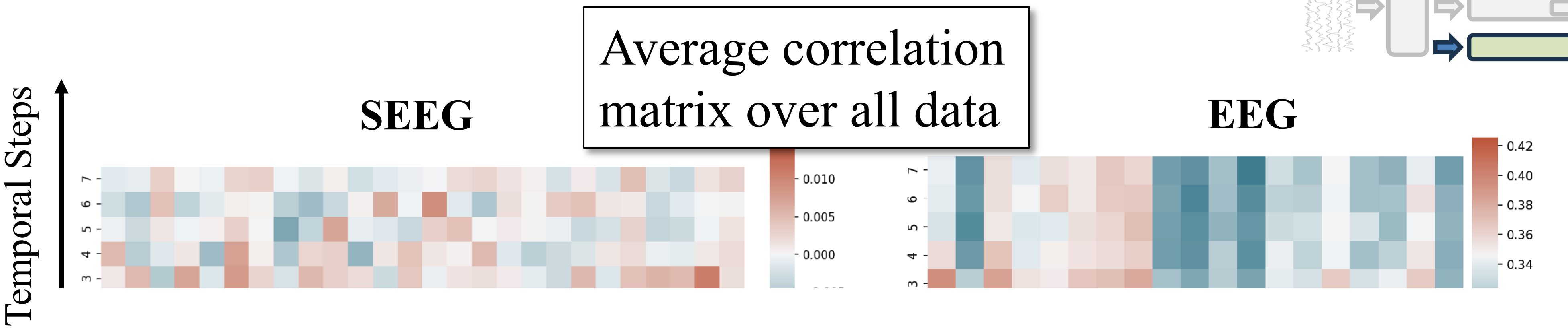
$$I(x_{t+k}^i; c_t^i, \Phi(\{c_t^j\}_{j\neq i})) \geq I(x_{t+k}^i; c_t^i), \qquad (1)$$

THEOREM 1. *Given a sample set for each channel* $X^i = \{x_1^i, \ldots, x_N^i\}$, $i = 1, \ldots, n$ *consisting of one positive sample from* $p(x_{t+k}^i|\Phi(c_t))$ *and* $N-1$ *negative samples from* $\sum_j p(x_{t+k}^j)/n$, *where* $n$ *is the number of channels. The optimal* $\mathcal{L}_N^{opt}$ *is the lower bound of* $\sum_i I(x_{t+k}^i; \Phi(c_t))$:

$$\mathcal{L}_N^{opt} \geq \sum_i \left[-I(x_{t+k}^i; \Phi(c_t)) + \log N\right]. \qquad (3)$$

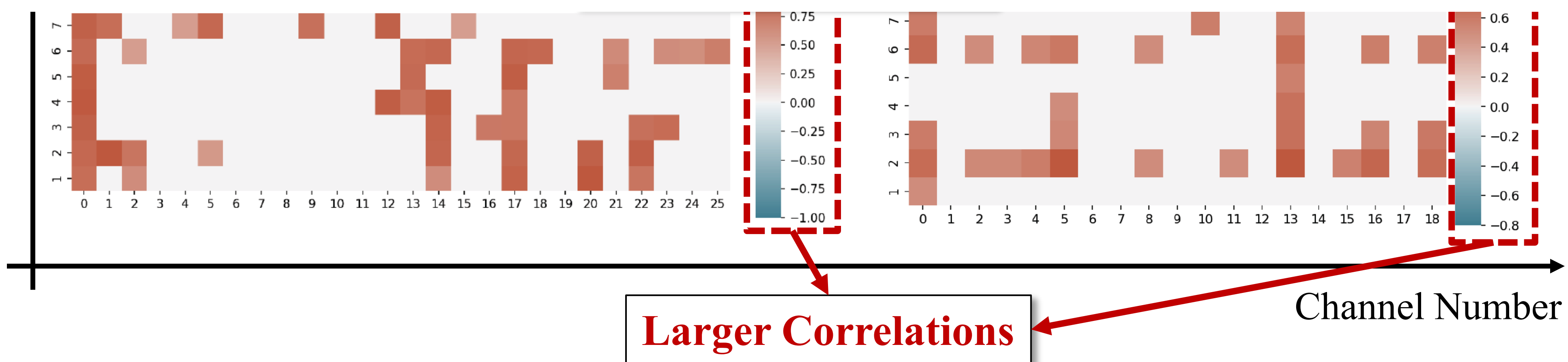
# How to capture temporal correlation patterns?





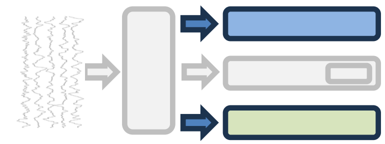

**Conclusion:**

It is necessary to model the correlations across time steps.

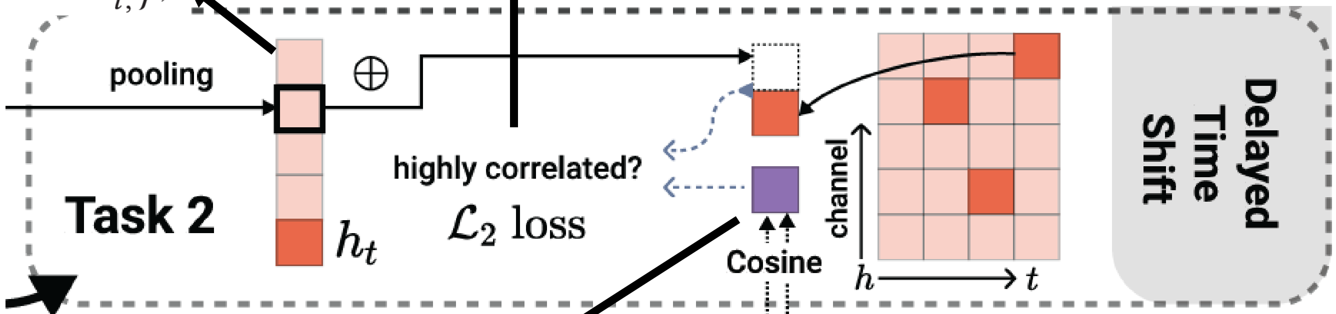


# How to capture temporal correlation patterns?

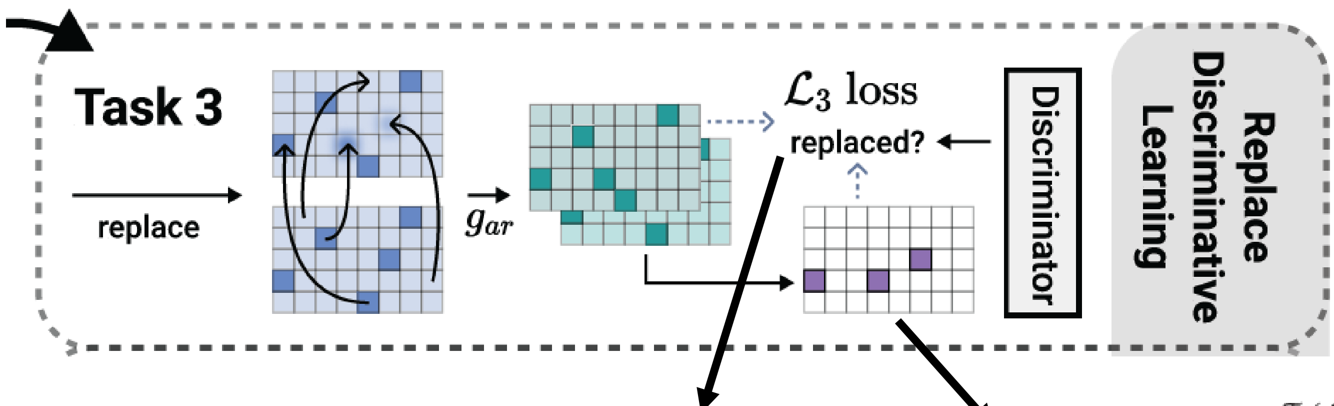
$$\mathcal{L}_2 = -\mathbb{E}_{t,i,k_2,j}\left[Y_t^i(k_2,j)\log\hat{p} + (1 - Y_t^i(k_2,j))\log(1-\hat{p})\right]$$



$$h_t = \text{Pooling}(c_{t,1}^{\text{self}}, \cdots, c_{t,\mathcal{T}}^{\text{self}}),$$



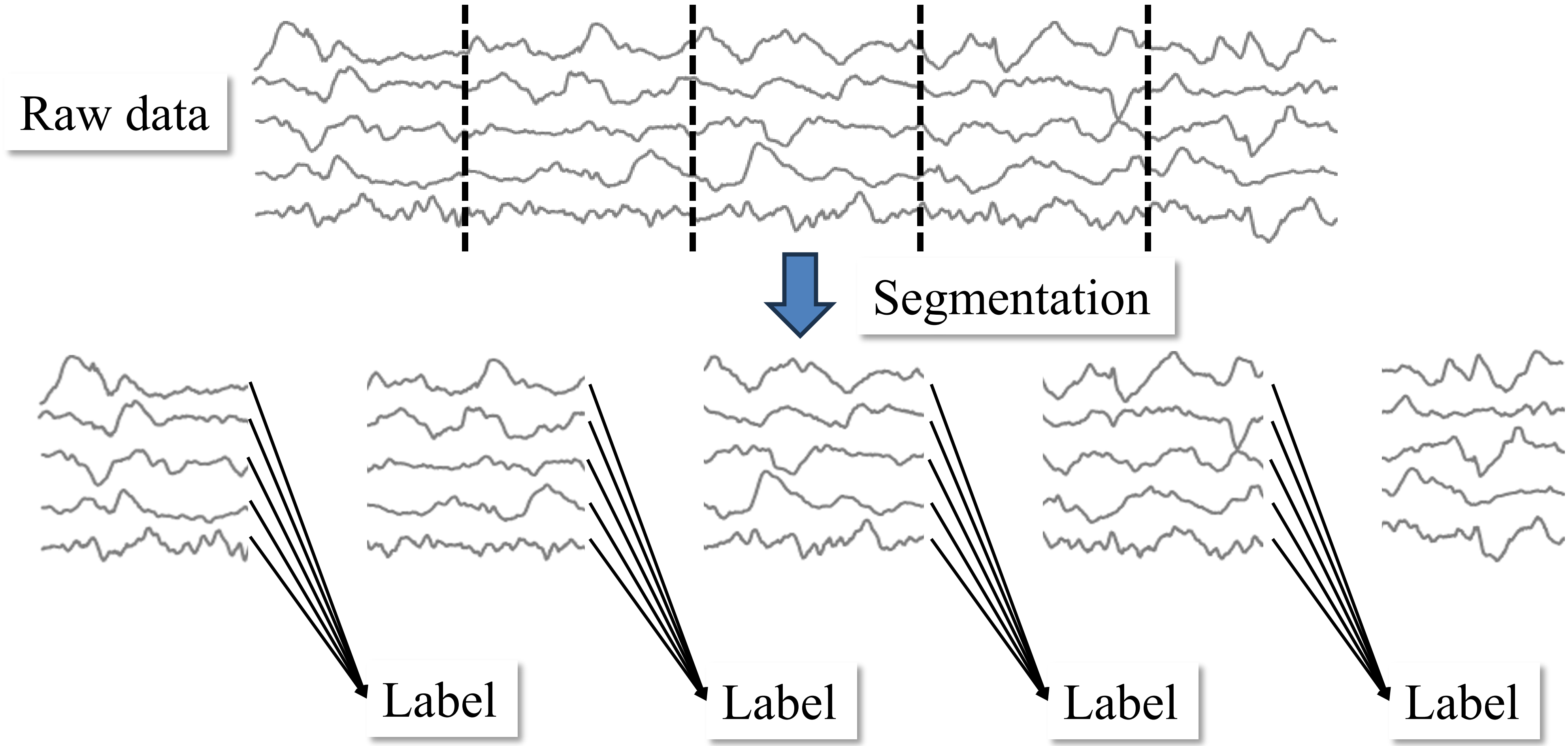

$$\mathrm{B}_t^i(k_2,j) = \text{Cosine}(s_{t,i}, s_{t+k_2,j}), \quad Y_t^i(k_2,j) = \begin{cases} 1, & \mathrm{B}_t^i(k_2,j) \geq \theta_2, \\ 0, & \mathrm{B}_t^i(k_2,j) < \theta_2. \end{cases}$$




$$\mathcal{L}_3 = -\mathbb{E}_{t,\tau,i}\left[Y_t(\tau,i)\log\hat{q} + (1 - Y_t(\tau,i))\log(1-\hat{q})\right],$$

$$Y_t(\tau,i) = \begin{cases} 1, & \mathcal{I}(\hat{z}_{t,\tau,i}) \neq i, \\ 0, & \mathcal{I}(\hat{z}_{t,\tau,i}) = i. \end{cases}$$

# Experimental Setup




**Channel-wise Prediction**

# Experimental Setup

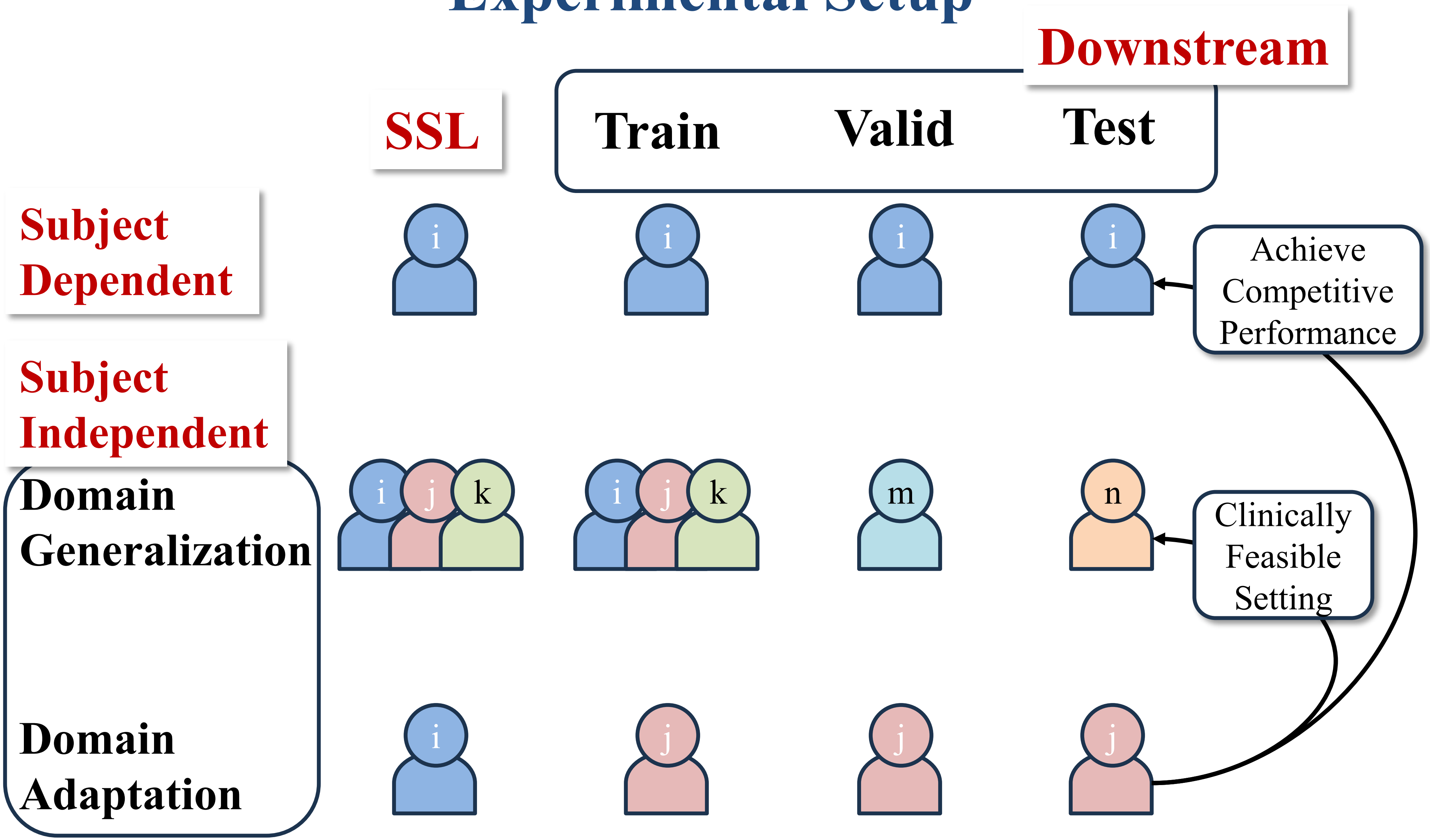
| Models |
| --- |
| MiniRocket |
| CPC |
| SimCLR |
| T-Loss |
| TST |
| GTS |
| TS-TCC |
| TS2Vec |
| *MBrain* |

## Baseline methods

- Supervised model MiniRocket [1]
- Graph-based model GTS [6]
- Reconstruction-based model TST [5]
- Contrastive-based models CPC [2], SimCLR [3], T-Loss [4], TS-TCC [7], TS2Vec [8]

[1] Angus Dempster, Daniel F Schmidt, and Geoffrey I Webb. 2021. Minirocket: A very fast (almost) deterministic transform for time series classification. In KDD. 248–257.
[2] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018).
[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In ICML. 1597–1607.
[4] Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. 2019. Unsupervised scalable representation learning for multivariate time series. In NeurIPS.
[5] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. 2021. A Transformer-based framework for multivariate time series representation learning. In KDD. 2114–2124.
[6] Chao Shang, Jie Chen, and Jinbo Bi. 2021. Discrete graph structure learning for forecasting multiple time series. In ICLR.
[7] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. 2021. Time-series representation learning via temporal and contextual contrasting. In IJCAI. 2352–2359.
[8] Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. 2021. TS2Vec: Towards universal representation of time series. arXiv preprint arXiv:2106.10466 (2021).

# Experimental Setup



# Subject Dependent Experiment

**SEEG dataset**

| | | | 28.92% ↑ | 26.85% ↑ |
|---|---|---|---|---|
| **Models** | Pre. | Rec. | $F_1$ | $F_2$ |
| MiniRocket | 22.98±0.15 | **66.24±0.26** | 31.79±0.19 | 43.58±0.22 |
| CPC | 27.65±4.49 | 55.07±3.52 | 34.20±3.40 | 42.73±2.57 |
| SimCLR | 11.06±3.95 | 51.54±5.87 | 16.60±4.68 | 25.41±4.95 |
| T-Loss | 29.29±2.65 | 51.55±2.53 | 36.00±1.97 | 43.13±1.57 |
| TST | 13.60±3.48 | 44.65±4.21 | 19.80±3.73 | 28.41±3.29 |
| GTS | 24.29±4.26 | 40.39±5.80 | 29.16±2.97 | 34.17±2.36 |
| TS-TCC | 22.10±7.65 | 49.94±5.41 | 25.32±8.02 | 32.74±7.95 |
| TS2Vec | 30.56±2.17 | 52.83±2.89 | 36.03±1.72 | 43.35±1.59 |
| *MBrain* | **37.97**±2.75 | 65.07±2.68 | **46.45**±2.25 | **55.28**±1.77 |

# Domain Generalization Experiment

**SEEG dataset**

**EEG dataset**

| Models | SEEG | | | | EEG | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Pre. | Rec. | $F_1$ | $F_2$ | Pre. | Rec. | $F_1$ | $F_2$ | AUROC |
| MiniRocket | 5.85±0.20 | **39.18±0.59** | 9.93±0.29 | 17.24±0.37 | **22.86±0.84** | 63.08±1.47 | 33.56±1.11 | 46.66±1.33 | 75.30±0.77 |
| CPC | 22.88±5.06 | 23.92±3.90 | 20.11±3.27 | 21.23±2.49 | 22.81±2.04 | 58.31±7.55 | 32.50±1.24 | 44.02±2.43 | 74.53±1.00 |
| SimCLR | 14.02±3.71 | 26.36±4.99 | 11.07±3.49 | 13.47±4.01 | 12.63±1.62 | 74.88±16.77 | 21.33±1.95 | 36.78±2.61 | 55.86±5.36 |
| T-Loss | 21.38±4.25 | 28.50±4.07 | 23.48±3.30 | 25.90±3.06 | 20.72±1.26 | 69.25±3.99 | 31.82±1.08 | 47.00±0.50 | 75.88±0.49 |
| TST | 8.37±3.96 | 32.48±8.25 | 11.80±3.91 | 15.67±3.69 | 15.65±1.54 | 28.59±12.93 | 19.65±4.36 | 23.87±8.09 | 58.20±4.27 |
| GTS | 24.16±5.91 | 27.99±4.98 | 22.77±2.69 | 24.15±2.79 | 18.86±1.09 | 62.51±5.04 | 28.88±0.88 | 42.54±1.48 | 71.69±1.88 |
| TS-TCC | 24.24±4.51 | 26.61±5.96 | 19.89±5.23 | 22.11±5.08 | 15.55±0.88 | 39.76±11.08 | 21.89±1.20 | 29.60±4.64 | 58.63±1.62 |
| TS2Vec | 27.93±5.23 | 29.49±3.97 | 26.78±3.29 | 27.88±3.52 | 21.40±0.63 | 58.31±6.14 | 31.24±1.18 | 43.24±2.78 | 73.35±1.02 |
| *MBrain* | **30.69±5.92** | 38.94±4.34 | **32.61±3.60** | **35.64±3.04** | 22.13±1.03 | **76.99±4.49** | **34.32±0.90** | **51.34±0.97** | **77.96±0.97** |

21.77% ↑

27.83% ↑

4.74% ↑
on average

# Domain Adaptation Experiment (DA)

**Domain Adaptation Experiment**

| Setting | Group *A* | | | Group *B* | | |
|---|---|---|---|---|---|---|
| | *B*→*A* | *C*→*A* | *D*→*A* | *A*→*B* | *C*→*B* | *D*→*B* |
| DA | 68.55±4.27 | 69.14±6.54 | 68.78±4.12 | 41.08±2.59 | 46.06±3.05 | 46.12±2.04 |
| Max-base | | 62.49±2.30 | | | 39.78±2.04 | |
| Non-DA | | **70.63±1.41** | | | **46.62±2.42** | |

Less than 15% ↓

| Setting | Group *C* | | | Group *D* | | |
|---|---|---|---|---|---|---|
| | *A*→*C* | *B*→*C* | *D*→*C* | *A*→*D* | *B*→*D* | *C*→*D* |
| DA | 40.04±3.98 | 39.34±2.11 | **48.64±5.48** | 80.82±0.65 | 79.90±1.11 | 80.72±1.31 |
| Max-base | | 33.59±2.23 | | | 75.35±0.79 | |
| Non-DA | | 46.09±2.35 | | | **83.27±0.95** | |

**Beat**

**Domain Generalization Experiment**

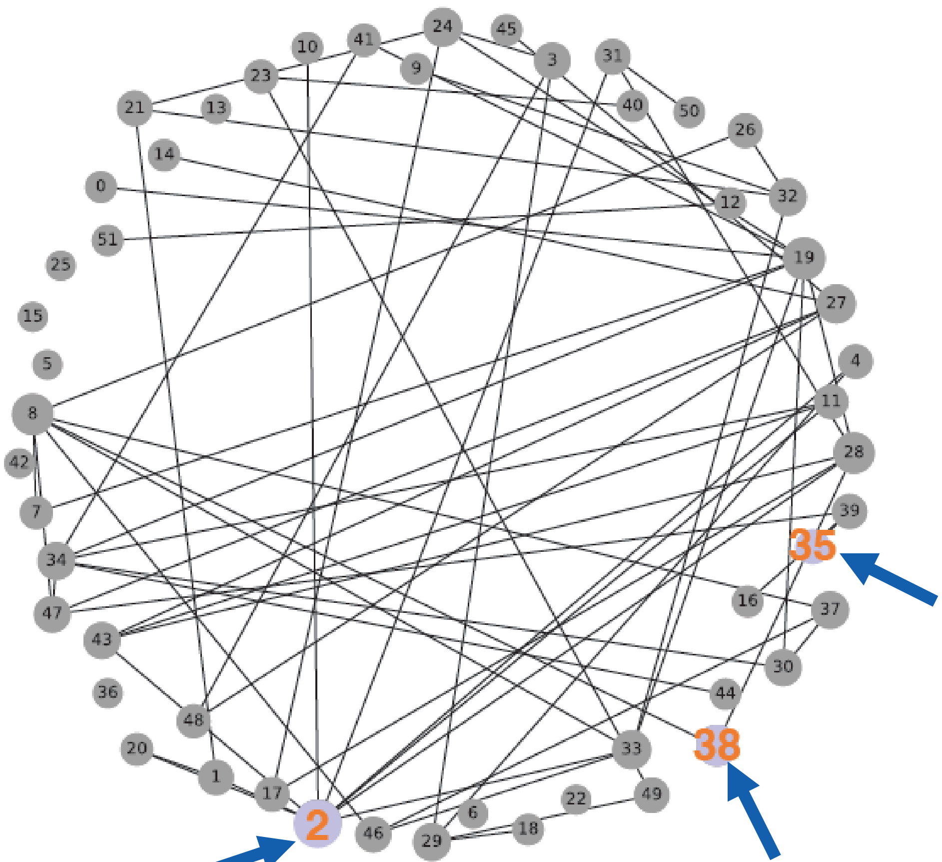
| | | | | |
|---|---|---|---|---|
| *MBrain* | **30.69±5.92** | 38.94±4.34 | **32.61±3.60** | **35.64±3.04** |

41.73% ↓

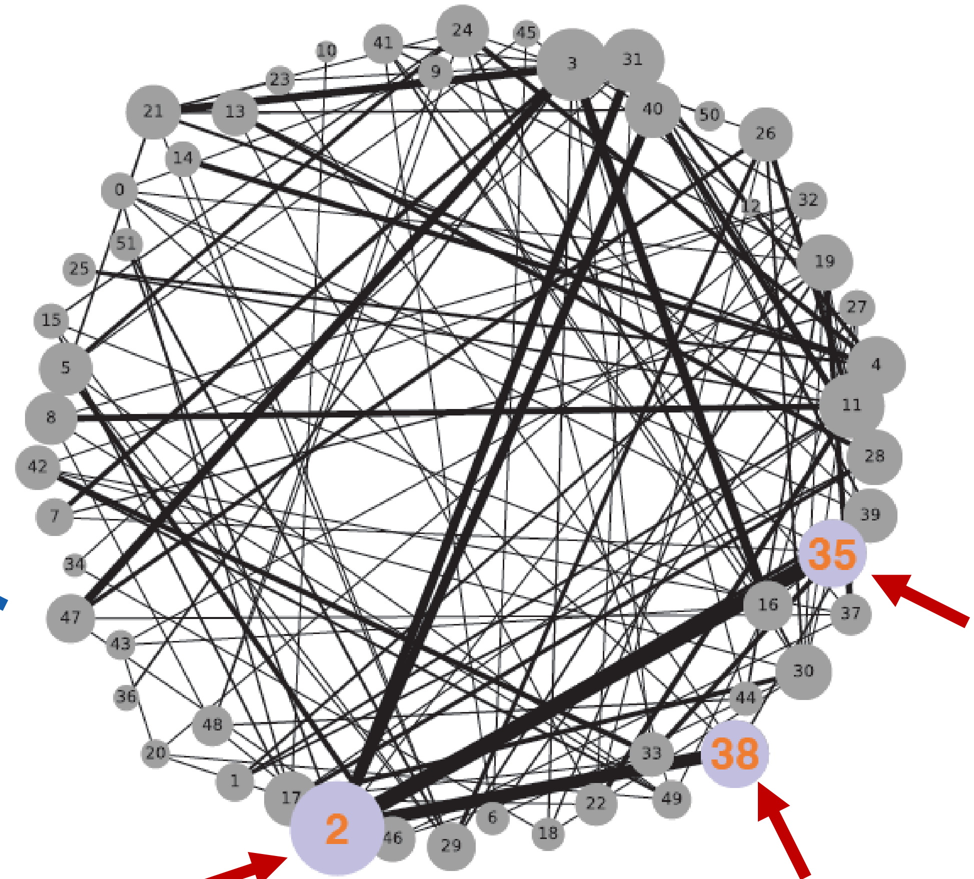
**Subject Dependent Experiment**

| | | | | |
|---|---|---|---|---|
| *MBrain* | **37.97±2.75** | 65.07±2.68 | **46.45±2.25** | **55.28±1.77** |

# Case Study




(a) Normal correlation graph.




(b) Seizure correlation graph.

**Conclusion:**

MBrain can learn the correlation patterns of brain signals.

# Conclusions

- We are the **first** to design a generalized self-supervised learning framework MBrain to pre-train both EEG and SEEG signals.
- MBrain explicitly capture the spatial and temporal correlations of brain signals while giving channel-wise predictions.
- We validate the effectiveness and clinical value of MBrain through extensive experiments on real-world EEG and SEEG datasets.

**THANKS | Q&A**

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

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

Group Homepage