State Mamba: Spatiotemporal EEG State-Space Model with Dynamic Brain Alignment for Cross-Subject Representation

Weining Weng[1, 2], **Yang Gu**[1, 2,*], **Yuan Ma**[1, 2], **Yuchen Liu**[1, 2], **Yingwei Zhang**[1, 2], **Yiqiang Chen**[1, 2]

[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China
{wengweining21b, guyang, mayuan20z, liuyuchen23s, zhangyingwei, yqchen}@ict.ac.cn

Abstract

Cross-subject EEG decoding remains a fundamental challenge due to substantial inter-subject variability in brain activity, which hinders the development of subject-independent EEG models. Despite progress in extracting cross-subject invariant features, existing studies neglect the shared neural responses that arise under similar cognitive or emotional states across individuals, limiting their ability to learn generalized and consistent EEG representations. To address the challenges, we propose **State Mamba**, a novel spatiotemporal EEG state-space model that explicitly models neural responses and their spatiotemporal state transitions to learn consistent and generalized representations across subjects. Innovatively, State Mamba theoretically formulates a multi-channel Mamba architecture that jointly models and aligns spatial and temporal brain state transitions, supporting principled analysis of neural responses. To enhance spatiotemporal feature fusion, we introduce the LGANN module, which adopts global-local attention to integrate long- and short-term brain activity into a compact EEG representation. Furthermore, we design two self-supervised pretext tasks to extract consistent neural patterns across subjects: (1) representation alignment to align EEG representation, and (2) pattern alignment to align their transition rules under identical conditions, jointly promoting subject-invariant EEG representations. Extensive experiments on three benchmark datasets, FACED, DEAP, and ISRUC, demonstrate the superior performance of State Mamba in cross-subject emotion and sleep recognition tasks, validating its robust generalization capability.

Introduction

Electroencephalogram (EEG) signals reflect dynamic neural activity across multiple brain regions and are intricately associated with a range of brain functions, including emotion (Li et al. 2022), cognition (Dahal et al. 2011), and sleep (Motamedi-Fakhr et al. 2014). Their high temporal resolution and rich spatial information make EEG widely used for decoding complex mental states (Song et al. 2020; Saha and Fels 2019; Haynes and Rees 2006). Early research focused on identifying task-relevant neural signatures by analyzing salient EEG waveforms evoked by specific stimuli (Kulkarni and Bairagi 2017). For instance, event-related potentials (ERPs) like the P300 (Polich 1997) and LPP (De Cesarei and Codispoti 2011) components exhibit increased amplitudes after emotional stimuli and serve as reliable markers of affective processing (Ding et al. 2017). With advances in artificial intelligence and deep learning, EEG analysis has shifted from handcrafted features to automated, data-driven modeling (Weng et al. 2024). Existing approaches leverage a wide range of computational methods—from classical machine learning algorithms to advanced deep learning models—to improve performance in tasks such as emotion recognition, cognitive assessment, and sleep stage classification.

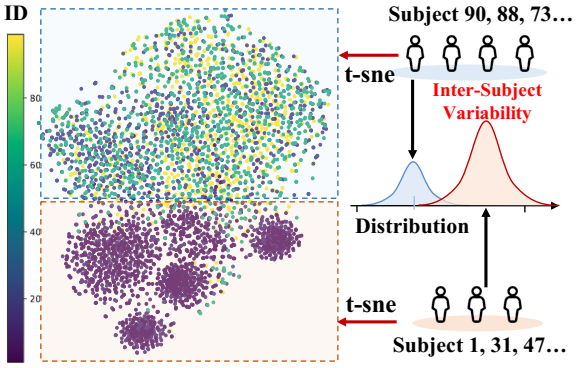

Figure 1: EEG signals exhibit substantial inter-subject variability and distributional discrepancies.

Cross-subject EEG decoding remains challenging due to substantial inter-subject variability in neural responses and data distributions. Variations in cognitive processing and neural expression across subjects significantly degrade EEG model performance when applied to new individuals (Hang et al. 2019). To address this, transfer learning (Li et al. 2019) and self-supervised learning (Li et al. 2024) have emerged as effective strategies for enhancing cross-subject generalization in various EEG tasks (Jiménez-Guarneros and Gómez-Gil 2020; Zhou et al. 2024). Transfer learning methods mitigate inter-subject distribution shifts by incorporating cross-subject distribution alignment losses (Zhao, Yan, and Lu 2021), while self-supervised methods employ contrastive or predictive pretext tasks to align EEG representations across subjects, which are then adapted to downstream tasks (Shen

*Corresponding Author

et al. 2022). Together, these methods advance cross-subject EEG decoding by addressing inter-subject variability, enabling more robust and generalizable models.

Despite significant progress in improving generalization, existing cross-subject EEG decoding methods predominantly rely on data-driven paradigms that extract invariant features only at the data level, neglecting the more fundamental neural invariance at the semantic level. Neuroscience studies reveal that different individuals exhibit similar neural activation patterns in core brain regions under identical stimuli (Engel and Wang 2011; Radua et al. 2014), suggesting the existence of shared neural response mechanisms. While deep models have advanced EEG representation learning by capturing spatiotemporal signal features (Li et al. 2023), they overlook those common neural responses across subjects. Current transfer and self-supervised methods focus mainly on aligning feature distributions, without explicitly modeling this cross-subject consistency in neural responses. As a result, they may discard meaningful and critical neural patterns, thereby limiting their ability to extract invariant neural features critical for explaining brain processes and enabling robust cross-subject generalization.

To solve the above challenges, we propose **State Mamba**, a spatiotemporal EEG State-Space Model for cross-subject EEG representation. Inspired by the State-Space Model (SSM) from control theory and the Mamba architecture (Gu and Dao 2023), State Mamba introduces a unified theoretical framework for modeling spatiotemporal state transition patterns occurring among multi-channel EEG signals. This novel framework incorporates both spatial and temporal state transition matrices to jointly capture temporal dynamics and inter-channel interactions of brain states, enabling the precise characterization of spatiotemporal neural response features. Besides, we design a local and global attention module (**LGANN**) to couple high-dimensional EEG features. LGANN integrates both short- and long-range dependencies across channels, yielding a compact representation that effectively preserves and enhances task-relevant brain activity patterns. To address cross-subject variability in neural responses, we propose two self-supervised pretext tasks: **representation alignment**, which enforces feature-level consistency, and **pattern alignment**, which synchronizes spatiotemporal transition matrices under identical brain states. Joint optimization of these pretext tasks explicitly enables State Mamba to learn consistent neural response patterns across subjects. Specifically, the contributions of this paper are summarized as follows:

(1) We propose State Mamba, a spatiotemporal EEG state-space model for cross-subject representation. By theoretically formulating temporal and spatial state transition matrices and developing a multi-channel Mamba architecture, State Mamba captures the dynamic evolution of brain states and extracts discriminative features from EEG signals.

(2) We introduce LGANN, a Local and Global Attention Neural Network that hierarchically integrates multi-channel EEG features by modeling both short-range local dependencies and long-range global interactions. This dual-attention mechanism enables compact and informative low-dimensional signal representation learning.

(3) We design two self-supervised pretext tasks: representation alignment and pattern alignment, to pretrain State Mamba. These tasks align EEG features and synchronize spatiotemporal transition patterns across subjects, enabling the model to learn subject-invariant and discriminative brain activity representations through unlabeled pre-training.

(4) We conduct comprehensive experiments on three benchmark datasets, demonstrating that State Mamba achieves state-of-the-art performance in cross-subject emotion and sleep stage recognition tasks, highlighting its strong generalization across diverse EEG decoding tasks.

Methodology

Preliminary

State-Space Models (SSMs) have undergone significant advancements in recent years. SSMs provide a mathematical framework for modeling the temporal evolution of dynamic systems (Aoki 2013), utilizing the discretized state transition matrix to control the dynamic transition of hidden states. The state space equation is formulated as follows:

$$h(t+1) = \overline{\boldsymbol{A}}h(t) + \overline{\boldsymbol{B}}x(t), y(t) = \boldsymbol{C}h(t), \quad (1)$$

where $h(t)$ and $h(t+1)$ denote the hidden states at time step t and $t+1$, respectively. $\overline{\boldsymbol{A}}$ is the state transition matrix governing the evolution of hidden states, while $\overline{\boldsymbol{B}}$ projects the input $x(t)$ into the hidden state space. $\boldsymbol{C}$ maps the hidden states to the output representation. $\overline{\boldsymbol{A}}$ and $\overline{\boldsymbol{B}}$ are both discretized by the time step Δ according to the zero-order hold principle (Lv et al. 2025), which can be shown as follows:

$$\overline{\boldsymbol{A}} = \exp(\Delta\boldsymbol{A}), \overline{\boldsymbol{B}} = (\Delta\boldsymbol{A})^{-1}(\exp(\Delta\boldsymbol{A}) - \boldsymbol{I})\Delta\boldsymbol{B}, \quad (2)$$

where $\boldsymbol{A} \in \mathbb{R}^{D\times N}$ is the trainable parameter, $\boldsymbol{B} \in \mathbb{R}^{|\mathcal{B}|\times P\times N}$ and $\boldsymbol{C} \in \mathbb{R}^{|\mathcal{B}|\times P\times N}$ are input-dependent matrices that implement the input selection mechanism within the Mamba module. $|\mathcal{B}|$ denotes the batch size, P represents the number of patches, and N is the dimensionality of the hidden states. Mamba leverages a state space recurrence mechanism with linear-time complexity and strong long-range modeling capacity, making it a scalable and efficient alternative to Transformers for time-series modeling tasks.

Multi-Channel EEG Patches

Similar to language models, temporal EEG signals are segmented and embedded into patches, which are shown in Figure 2. Given an EEG sample $s \in \mathbb{R}^{c\times l}$ with c channels l time points, the signal is divided into p non-overlapping segments, forming a structured EEG sequence $s \in \mathbb{R}^{c\times p\times l/p}$. Each segment is linearly projected into a shared feature space, followed by Z-score normalization to enforce a distribution with zero mean and unit variance across the features.

In language models, positional encoding typically captures the sequential order of tokens (Ke, He, and Liu 2021). In contrast, multi-channel EEG signals require modeling both temporal order and spatial electrode locations. To this end, we incorporate temporal and spatial position encoding for EEG patches. Temporal order is encoded using Sinusoidal Position Encoding (Liu et al. 2020), while spatial

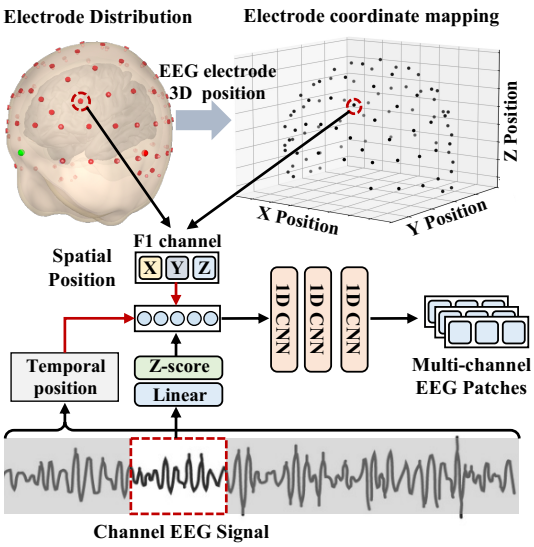

Figure 2: The process of embedding multichannel EEG signals into patches. Each EEG channel is embedded into patches through the multi-layer 1DCNN module, incorporating both temporal and 3D spatial positional encoding.

electrode location is captured by mapping the spatial distribution of EEG electrodes on the cerebral cortex to 3D coordinates based on the 10–20 or BioSemi64 system. These coordinates are concatenated with the corresponding EEG patches to provide spatial context.

Finally, 1D CNNs embed the continuous EEG into multi-channel patches $x \in \mathbb{R}^{c \times p \times d}$, where each patch encodes spatial-temporal structure and d is the feature dimension.

State Mamba: Spatiotemporal Brain Mamba

Innovatively, we propose State Mamba, a spatiotemporal brain state-space model that theoretically characterizes the spatiotemporal transition processes of EEG states, aiming to capture temporal and channel-wise neural response features evoked by specific brain stimuli. Based on the derivative definition $f'(x) = \lim_{\Delta \to 0} \frac{f(t+\Delta)-f(t)}{\Delta} \Rightarrow \Delta f'(x) = f(t+\Delta) - f(t)$, the continuous state transition for the i-th EEG channel can be approximated in discrete form as:

$$h_i(t+\Delta) = \Delta \cdot h_i'(t) + h(t), \tag{3}$$

where $h_i(t)$ denotes the hidden state of the i-th EEG channel at time t. In the Mamba architecture (Gu and Dao 2023), the state derivative $h'(t)$ is defined by a single-channel state transition differential equation: $h'(t) = \boldsymbol{A}h(t) + \boldsymbol{B}x(t)$. However, the spatial correlations among the states and inputs of other brain channels influence the state change of a given channel. Accordingly, we redefine the multi-channel state transition differential equation as follows:

$$h_i'(t) = \boldsymbol{A}h_i(t) + \boldsymbol{B}x_i(t) + \sum_{j \neq i}^{j=c} \boldsymbol{A_j^c} h_j(t) + \sum_{j \neq i}^{j=c} \alpha_j x_j(t), \tag{4}$$

where $\boldsymbol{A^c}$ is the spatial state transition matrix controlling the state interactions among the remaining EEG channels, and α is a weighting coefficient reflecting input correlations across EEG channels. Accordingly, Equation 3 can be expanded as:

$$\begin{aligned} h_i(t+\Delta) = & \underbrace{(\Delta \boldsymbol{A} + \boldsymbol{I}) h_i(t) + \Delta \boldsymbol{B} x_i(t)}_{\text{Current channel state transition}} \\ & \underbrace{+ \Delta \sum_{j \neq i}^{c} \boldsymbol{A_j^c} h_j(t) + \Delta \sum_{j \neq i}^{c} \alpha_j x_j(t),}_{\text{Spatial influence from other channels}} \end{aligned} \tag{5}$$

where Δ, $\boldsymbol{B}$, and $\boldsymbol{C}$ are input-dependent parameters that implement selection conditioned on the EEG sequence, while $\boldsymbol{A}$ and $\boldsymbol{A^c}$ govern temporal and spatial neural state transitions, and $Delta$ implements the discretization of $\boldsymbol{A}$ and $\boldsymbol{B}$.

To optimize the computational efficiency of state-space modeling for multi-channel EEG, we propose two simplification strategies that further streamline the spatiotemporal state transition equations in State Mamba.

Strategy 1: Simplification of inter-channel input dependencies. Given the strong temporal dependencies in EEG signals—where the current signal is highly correlated with the previous one—we simplify the state transition modeling by retaining only the influence of previous multi-channel states. The contribution of current inter-channel input correlations is omitted, as their effect is largely redundant and implicitly captured by the temporal dynamics. The resulting formulation is as follows:

$$\begin{aligned} h_i(t+\Delta) \overset{\text{simplify}}{=} & \underbrace{(\Delta \boldsymbol{A} + \boldsymbol{I}) h_i(t) + \Delta \boldsymbol{B} x_i(t)}_{\text{Current channel state transfer}} \\ & + \underbrace{\Delta \sum_{j=1, j \neq i}^{j=c} \boldsymbol{A_j^c} h_j(t)}_{\text{Spatial influence from other channels}}. \end{aligned} \tag{6}$$

Strategy 2: Shared spatial transition mapping. To further reduce the complexity of spatial modeling across EEG channels, we replace the set of channel-specific matrices A_j^c with a shared spatial transition matrix $\boldsymbol{A^c}$. This approximation assumes that the influence of other channels can be aggregated using a unified transformation, thereby simplifying the spatial interaction structure while retaining the core cross-channel dynamics, which can be shown as:

$$\begin{aligned} h_i(t+\Delta) \overset{\text{simplify}}{=} & \underbrace{(\Delta \boldsymbol{A} + \boldsymbol{I}) h_i(t) + \Delta \boldsymbol{B} x_i(t)}_{\text{Current channel state transfer}} \\ & + \underbrace{\Delta \boldsymbol{A^c} \sum_{j=1, j \neq i}^{j=c_n} h_j(t)}_{\text{Spatial influence from other channels}}. \end{aligned} \tag{7}$$

The proposed simplifications reduce the computational complexity of multi-channel State Mamba scanning from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log n)$, while maintaining parallel computation efficiency. Leveraging the revised state update mechanism, each State Mamba block models the brain state of individual patches across EEG channels. With residual connections and RMSNorm, the architecture supports deep stacking, enabling hierarchical modeling of complex spatiotemporal brain activity and neural response features.

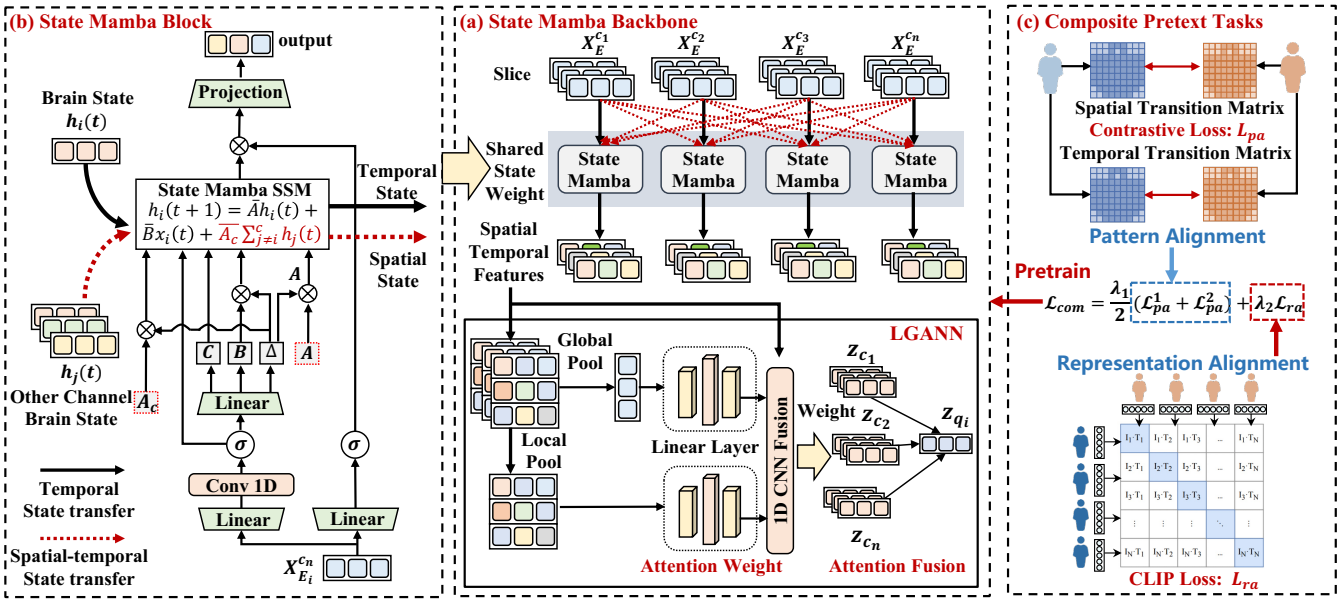

Figure 3: Detailed structure of State Mamba. State Mamba introduces a multi-channel state transition mechanism across EEG patches to extract spatiotemporal neural activity features. The LGANN module further integrates spatiotemporal state information into compact representations through the multi-scale attention. Composite pretext tasks are designed to align cross-subject neural response patterns under the identical brain states, guiding self-supervised learning of generalizable EEG representations.

LGANN Module. We propose a local and global attention module to project high-dimensional EEG features into compact representations. **Global Attention** aggregates information across all patches and time steps to generate a global descriptor for each channel, followed by linear attention to compute weights reflecting holistic patterns of brain activity. **Local Attention** performs temporal pooling to retain patch-level granularity and applies linear attention to model fine-grained temporal dependencies within short windows. By integrating both attention mechanisms, LGANN produces a unified attention weight that balances global consistency and local specificity, which can be computed as:

$$\begin{aligned}\mathbf{a}^{(b)} = &\underbrace{\sigma(\mathbf{W}_2^{(g)} \cdot \mathrm{ReLU}(\mathbf{W}_1^{(g)} \cdot (\frac{1}{pl}\sum_{i=1}^{p}\sum_{j=1}^{l}\mathbf{X}_{b,c,i,j})}_{\text{Global attention path}} \\ &+ \underbrace{\sigma(\mathbf{W}_2^{(l)} \cdot \mathrm{ReLU}(\mathbf{W}_1^{(l)} \cdot (\frac{1}{l}\sum_{j=1}^{l}\mathbf{X}_{b,c,i,j})_{i=1}^{p})),}_{\text{Local attention path}}\end{aligned} \tag{8}$$

where $\boldsymbol{X}_{b,c} \in \mathbb{R}^{p\times l}$ denotes the channel-specific brain activity features output by the State Mamba Block, and σ is the sigmoid activation function. The combined attention weights $\boldsymbol{a}^{(b)}$ are applied to $\boldsymbol{X}$ to recalibrate the input features via element-wise multiplication $\boldsymbol{X}_a = \boldsymbol{a}^{(b)} \cdot \boldsymbol{X}$. Subsequently, a 1D-CNN layer is employed to reduce the dimensionality from $\mathbb{R}^{B\times c\times(l\cdot p)}$ to a compact representation of size $\mathbb{R}^{B\times L_r}$. By coupling inter-channel correlations and temporal dependencies, LGANN effectively guides the dimensionality reduction process, preserving critical brain dynamics while mitigating information loss.

Composite Pretext Tasks

We introduce two novel pretext tasks-**representation alignment** and **pattern alignment**-to jointly pre-train the model by aligning EEG features and synchronizing neural response patterns across subjects under the same brain state. The composite pretext task design is shown in Figure 3(c).

Representation Alignment. To extract cross-subject consistent neural features, the representation alignment module aligns representations from different subjects under the same brain state. We assume that subjects exposed to the same stimulus tend to exhibit similar brain states—for instance, EEG signals recorded at the same time point during an emotional movie or after equivalent sleep durations often show comparable patterns. We adopt a CLIP-inspired sampling strategy (Radford et al. 2021), where one-to-one positive pairs are constructed across subjects under matched conditions. In each training iteration, neural representations from different subjects are sampled under the same stimulus, and the model is trained to maximize similarity between positive pairs while minimizing similarity to mismatched ones. This encourages the learning of a shared latent space capturing semantically aligned, subject-invariant representations.

Pattern Alignment. We propose a pattern alignment pretext task to synchronize spatiotemporal brain state transition patterns across subjects. In State Mamba, neural response patterns are encoded through temporal and spatial state transition matrices, which capture the evolution of neural activity over time and across channels. Leveraging a CLIP-style

contrastive learning framework, this task aligns these transition patterns among subjects under similar brain states, enabling the extraction of shared neural regulatory patterns that reflect consistent cross-subject brain response mechanisms.

Experiments

We conduct comprehensive experiments on three benchmark multi-channel EEG datasets across diverse tasks to evaluate the cross-subject representation performance of State Mamba. The evaluation is designed to address the following critical research questions:

- **RQ1:** Does the proposed method outperform existing models in cross-subject EEG representation and generalize well across tasks? (**Comparison Experiments**)
- **RQ2:** How does model performance respond to variations in hyperparameters and architecture across different tasks? (**Sensitive Analysis**)
- **RQ3:** Are the model architecture and pretext tasks well-designed, and how do they contribute to overall performance? (**Ablation Study**)
- **RQ4:** Can the model capture subject-invariant spatiotemporal patterns, facilitating the discovery of neural mechanisms underlying brain states? (**Pattern Visualization**)

Experiment Setup

Dataset We conduct experiments on three benchmark datasets covering diverse downstream tasks. The specific details are as follows:

FACED (Chen et al. 2023) is a video-evoked emotion EEG dataset comprising recordings from 123 subjects, each exposed to 28 emotion-inducing clips labeled across nine emotional states (e.g., anger, joy, neutral). For each video, 30-second EEG data were collected using a 32-channel acquisition system at a sampling rate of 250 Hz.

DEAP (Koelstra et al. 2011) includes EEG data from 32 subjects watching 40 emotional video clips. Emotions were rated on valence and arousal dimensions and binarized for classification (high vs. low) (Tao et al. 2020). EEG was recorded with 30 channels at 512 Hz and downsampled to 128 Hz, yielding 60-second segments per clip.

ISRUC-S1 (Khalighi et al. 2016) is a sleep-stage dataset comprising EEG from 100 subjects with sleep disorders. EEG was recorded via 8 channels at 200 Hz. Each 30-second epoch was annotated into one of five sleep stages: Wake, N1, N2, N3, and REM.

Evaluation Protocols We evaluate model performance under two settings: (1) 10-fold cross-validation, where all subjects are randomly divided into ten folds, with iterative training on nine and testing on the remaining one to assess sample-level generalization; and (2) Leave-One-Subject-Out, where one subject is held out for testing while the model is trained on the rest, providing a rigorous measure of subject-level generalization. Notably, all evaluations follow a domain generalization setup: subjects in downstream evaluation are entirely unseen during both pretraining and fine-tuning. This setting offers a stringent measure of the model's robustness to inter-subject variability.

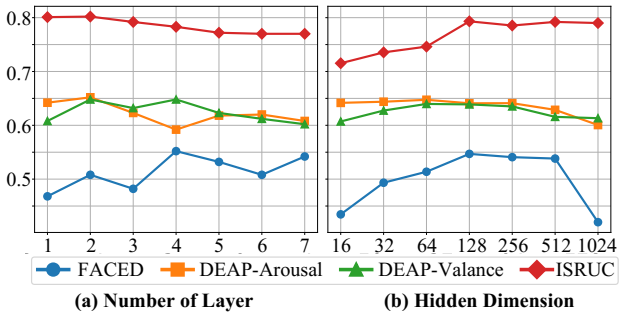

Figure 4: The results of the sensitive analysis.

Implementation For model training, we set the hidden dimension of State Mamba to 128 with 2 stacked layers and a batch size of 32. During pretraining, 50,000 positive pairs representing consistent brain states are randomly sampled from the training set in each epoch. To ensure balanced coverage, the sampling strategy is controlled to uniformly cover all consistent brain states across epochs. Optimization uses the Adam optimizer with a learning rate of 0.001 for pretraining and 0.0005 for downstream fine-tuning. All experiments are conducted on a server with 4×NVIDIA A100 GPUs and an Intel Xeon Platinum 8358P CPU (2.60GHz).

Baselines In our experiments, we compared the proposed State Mamba framework with several state-of-the-art EEG representation learning approaches (particularly self-supervised learning methods), including CBraMod (Wang et al. 2025), EEG2REP (Mohammadi Foumani et al. 2024), DMMR (Wang, Zhang, and Tang 2024), BIOT (Yang, Westover, and Sun 2023), TF-C (Zhang et al. 2022), CLISA (Shen et al. 2022), and mulEEG (Kumar et al. 2022)

RQ1: Comparison Experiments

As shown in Table 1, under the 10-fold cross-validation setting, State Mamba achieves over 2% accuracy improvement on the 9-class emotion recognition task (FACED) and approximately 3% gain on the 5-class sleep stage classification task (ISRUC), demonstrating strong generalization in extracting cross-subject features related to both emotional and sleep-related brain states. However, for the binary valence and arousal classification tasks, improvements over the advanced CBraMod (ICLR'25) baseline are marginal. This may be because valence and arousal are primarily associated with low-frequency, spatially distributed neural patterns, against which State Mamba's spatiotemporal modeling—designed to capture higher-frequency, localized dynamics—provides limited advantage given the tasks' low temporal and spatial resolution

As shown in Table 2, the results under the LOSO evaluation setting reveal similar trends to those observed in the 10-fold cross-validation, further validating the robustness of our model. Notably, LOSO exposes the model to a broader range of subjects during training, effectively expanding the source domains in the domain generalization setup. This increased subject diversity promotes better alignment of data distribu-

Model	FACED		DEAP-Arousal		DEAP-Valence		ISRUC-S1	
	Accuracy	Macro F1	Accuracy	Macro F1	Accuracy	Macro F1	Accuracy	Macro F1
CBraMod ICLR'25	52.39 ± 4.71	52.48 ± 2.37	64.09 ± 6.73	**64.28 ± 5.89**	62.95 ± 5.72	61.83 ± 6.76	76.83 ± 2.54	76.99 ± 4.02
EEG2REP KDD'24	41.35 ± 5.21	43.28 ± 4.19	57.83 ± 4.72	55.90 ± 5.62	59.01 ± 6.04	58.72 ± 5.36	73.26 ± 3.51	73.65 ± 3.74
DMMR AAAI'24	50.07 ± 3.71	44.32 ± 3.85	58.34 ± 5.16	55.30 ± 4.82	59.21 ± 6.33	56.70 ± 4.99	72.94 ± 4.25	74.63 ± 3.60
BIOT NeurIPS'23	50.84 ± 3.44	50.02 ± 2.67	63.09 ± 4.21	60.85 ± 6.79	61.04 ± 5.28	60.77 ± 4.07	73.58 ± 3.47	71.37 ± 3.45
TF-C NeurIPS'22	48.37 ± 5.03	49.02 ± 2.80	55.92 ± 2.86	52.86 ± 3.47	51.34 ± 4.80	54.06 ± 4.37	74.52 ± 2.71	73.93 ± 2.69
CLISA TAC'22	45.71 ± 5.59	44.98 ± 2.35	55.46 ± 4.70	53.82 ± 2.79	58.01 ± 3.95	59.46 ± 5.42	71.49 ± 2.32	73.50 ± 4.52
mulEEG MICCAI'22	47.82 ± 4.17	45.48 ± 3.95	59.46 ± 3.60	57.39 ± 2.86	60.45 ± 5.03	61.42 ± 5.37	70.83 ± 4.10	72.75 ± 3.08
State Mamba	**54.69 ± 5.34**	**53.89 ± 4.68**	**64.38 ± 7.21**	62.85 ± 5.29	**63.97 ± 5.57**	**64.43 ± 2.97**	**79.32 ± 3.26**	**78.51 ± 3.04**

Table 1: The experimental results on three datasets under the 10-fold cross-validation mode.

Model	FACED		DEAP-Arousal		DEAP-Valence		ISRUC-S1	
	Accuracy	Macro F1	Accuracy	Macro F1	Accuracy	Macro F1	Accuracy	Macro F1
CBraMod ICLR'25	54.02 ± 5.44	53.73 ± 4.83	66.88 ± 7.29	**68.42 ± 6.72**	65.33 ± 7.35	65.89 ± 5.43	79.05 ± 2.43	75.42 ± 2.84
EEG2REP KDD'24	47.38 ± 6.41	49.10 ± 5.42	59.37 ± 3.50	60.12 ± 5.81	60.19 ± 4.72	59.82 ± 4.58	74.82 ± 2.71	75.88 ± 2.06
DMMR AAAI'24	52.38 ± 5.03	50.79 ± 4.70	60.33 ± 6.20	58.52 ± 4.11	61.59 ± 7.03	61.94 ± 6.97	75.89 ± 2.04	72.85 ± 2.73
BIOT NeurIPS'23	51.03 ± 4.53	52.58 ± 3.09	62.17 ± 6.05	63.55 ± 6.29	61.89 ± 5.75	62.04 ± 6.17	73.43 ± 3.40	72.09 ± 2.93
TF-C NeurIPS'22	48.74 ± 4.02	48.37 ± 2.95	55.83 ± 3.52	57.09 ± 4.02	57.95 ± 4.98	59.20 ± 4.79	76.80 ± 3.59	79.07 ± 3.34
CLISA TAC'22	48.69 ± 6.93	49.82 ± 5.71	57.35 ± 5.88	57.96 ± 6.01	59.89 ± 5.46	58.73 ± 6.22	72.45 ± 4.72	70.69 ± 4.80
mulEEG MICCAI'22	49.06 ± 3.57	46.98 ± 6.43	61.36 ± 4.82	58.04 ± 5.82	62.38 ± 6.77	61.05 ± 5.94	75.36 ± 2.23	76.08 ± 1.77
State Mamba	**58.33 ± 3.17**	**57.94 ± 4.25**	**68.04 ± 9.08**	67.51 ± 8.10	**67.81 ± 9.12**	**66.20 ± 8.43**	**81.43 ± 4.71**	**79.88 ± 5.27**

Table 2: The experimental results on three datasets under the Leave-One-Subject-Out (LOSO) mode.

tions and brain state representations, facilitating more effective cross-subject transfer.

RQ2: Sensitive Analysis

To investigate the sensitivity of the State Mamba model to key architectural hyperparameters, we conduct two sets of ablation studies on (1) hidden dimension and (2) stacked Mamba layers. As shown in Figure 4 (left), increasing the hidden dimension initially improves performance across all datasets, with most tasks reaching optimal accuracy at intermediate dimensions (e.g., 128 or 256). Beyond this point, performance tends to plateau or decline, suggesting overparameterization. In Figure 4 (right), we vary the number of stacked State Mamba layers from 1 to 7. While ISRUC and DEAP-Valence maintain stable performance, FACED and DEAP-Arousal exhibit mild fluctuations, with no clear monotonic trend. These results suggest that State Mamba is robust to moderate changes in depth and width, and performance is more sensitive to hidden size than layer count.

RQ3: Ablation Study

Furthermore, we conduct ablation studies to evaluate the contribution of key architectural components and pretext tasks on State Mamba's performance across various EEG downstream tasks. All experiments are evaluated under 10-fold cross-validation, with results summarized in Table 3.

Firstly, the ablation study validates the effectiveness of the composite pretext tasks. Removing the entire pretraining phase (i.e., supervised learning on the source domain) causes a significant performance drop across all datasets (e.g., FACED: 54.69 $\rightarrow$ 49.83; DEAP-Arousal: 64.38 $\rightarrow$ 60.62), and removing either alignment strategy also leads to noticeable degradation. These findings highlight that the pretext tasks not only provide a strong initialization but also enable the learning of semantically robust and subject-invariant representations, which are crucial for generalization in label-scarce and cross-subject EEG settings.

We further validated the architectural design of State Mamba by comparing it with a traditional single-channel Mamba model, in which the channel dimension is flattened into the batch dimension, modeling only temporal dependencies. Removing the spatial channel-wise state pathway caused a significant performance drop, confirming that inter-channel state interactions play a crucial role in enhancing discriminative representations and improving downstream task performance. Besides, we assessed the effectiveness of LGANN by comparing it with two alternatives: direct concatenation and average pooling of multi-channel EEG representations. While direct concatenation preserves full information, it yields high-dimensional features that increase model complexity and overfitting risk. In contrast, average pooling reduces dimensionality but discards informative features, degrading performance. LGANN achieves a favorable balance by compressing channel-wise features while retaining essential discriminative information. On the FACED and DEAP datasets, its performance closely matches that of direct concatenation, demonstrating its capacity to preserve task-relevant features. On the ISRUC dataset, where feature dimensionality is inherently lower due to fewer channels, the benefit of LGANN is diminished, as concatenation remains effective without excessive dimensionality.

RQ4: Neural Mechanism Visualization

To investigate the subject-independent neural response and brain activation pattern, we visualized the spatial and temporal state transition matrices during emotion and sleep modeling. Since these high-dimensional matrices encode complex neural dynamics, simple statistics such as element-wise

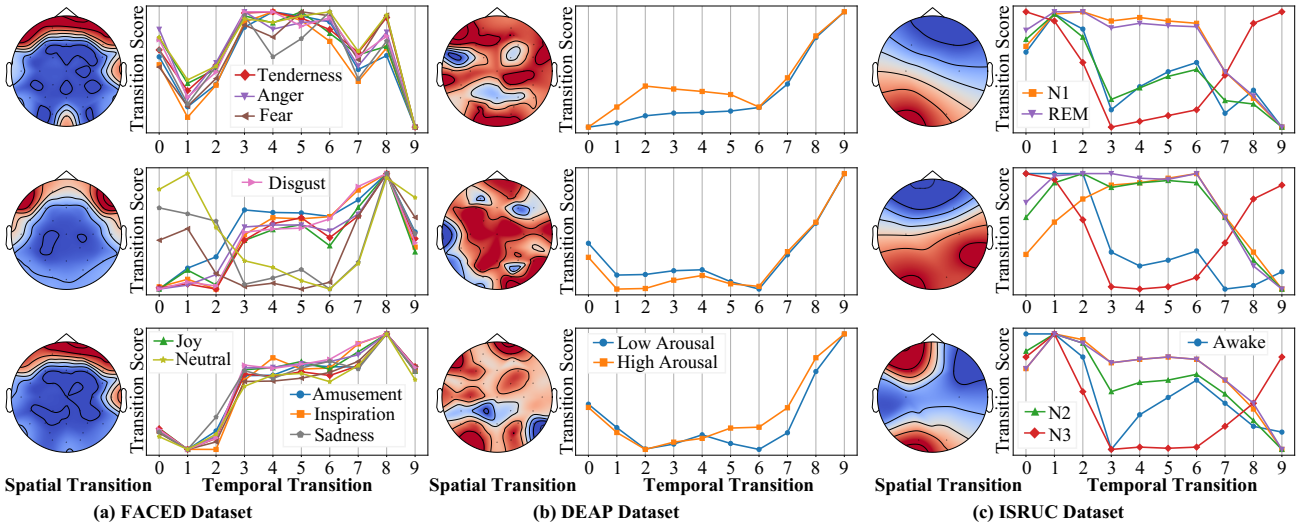

Figure 5: Visualization of temporal and spatial state transition matrices for three representative subjects from each dataset. Spatial transitions are shown as topographic graphs illustrating inter-regional connectivity, while temporal transitions are presented as line plots reflecting the evolution of state transition scores over time.

Methods	FACED	DEAP-A	DEAP-V	ISRUC
w/o Pattern Alignment	53.82	63.25	62.74	77.26
w/o Representation Alignment	51.37	62.17	62.89	77.51
w/o Pretrain	49.83	60.62	60.43	75.10
w/o LGANN (concat)	54.61	64.35	63.88	**80.03**
w/o LGANN (average)	52.91	64.29	63.72	78.06
Mamba (w/o spatial transition)	52.84	62.77	61.02	78.13
State Mamba	**54.69**	**64.38**	**63.97**	79.32

Table 3: The results of the Ablation Study.

means are insufficient to summarize the underlying spatiotemporal neural transition patterns. To better preserve and interpret the information embedded in the transition matrix, we adopt two metrics: average singular value decomposition (SVD) and the Frobenius norm. The average SVD captures dominant modes of variation, highlighting coordinated and low-rank neural response patterns that reflect task-relevant features. The Frobenius norm quantifies the overall energy or intensity of state transition and activation, offering a robust scalar measure of global neural engagement.

Figure 5 presents the visualization results for representative subjects. In FACED, the frontal regions—particularly the prefrontal cortex—exhibit marked activation and strong state transitions across different emotional states, aligning with prior evidence of frontal-limbic involvement in emotion processing (Chayer and Freedman 2001). Temporally, emotion-related brain activities follow consistent trajectories, with peak state transitions occurring at similar time points, suggesting synchronized neural responses to emotional stimuli. In DEAP, arousal-related transitions are concentrated in prefrontal and occipital regions, while midbrain and parietal areas show higher inter-subject variability, possibly due to the activation of visual-processing neurons in the occipital cortex elicited by emotional visual stimuli (Lang et al. 1998). In the ISRUC dataset, spatial transitions vary across subjects but consistently involve frontal and occipital regions, aligning with prior findings that highlight the role of occipital channels in sleep staging (Wang et al. 2022). Furthermore, temporal state transitions can distinguish sleep stages: N1 and N2 show smooth, gradually declining transition curves, reflecting reduced brain state variability, while wakefulness and N3 exhibit greater temporal fluctuations. Despite significant inter-subject variability in brain responses to both emotional and sleep-related tasks, State Mamba is capable of uncovering highly consistent and generalizable spatiotemporal patterns across individuals.

Conclusion

In this work, we addressed the critical challenge of cross-subject variability in EEG-based neural decoding by proposing State Mamba, a novel spatiotemporal EEG State-Space Model. State Mamba models latent brain states and their spatiotemporal transitions to uncover consistent neural response patterns shared across individuals. By theoretically formulating multi-channel Mamba architecture and incorporating the LGANN module for multi-scale attention, our model captures both global and local EEG neural activity features. Composite self-supervised pretext tasks further enable cross-subject alignment of brain states and transition patterns without reliance on labeled data. State Mamba's inherent support for multi-channel processing enables comprehensive modeling of spatial-temporal brain activity, with the integration of pretext tasks further promoting the learning of consistent neural response paradigms across subjects. Moreover, the modular design and flexibility of State Mamba make it readily extensible to other neural decoding tasks and modalities, laying the groundwork for future large-scale, general-purpose EEG foundation models.

Acknowledgments

This work was supported by National Key Research and Development Plan of China (No.2023YFC3604805), Beijing Municipal Science & Technology Commission (No. Z221100002722009), and Natural Science Foundation of China (No. 62302487).

References

Aoki, M. 2013. *State space modeling of time series*. Springer Science & Business Media.

Chayer, C.; and Freedman, M. 2001. Frontal lobe functions. *Current neurology and neuroscience reports*, 1(6): 547–552.

Chen, J.; Wang, X.; Huang, C.; Hu, X.; Shen, X.; and Zhang, D. 2023. A large finer-grained affective computing EEG dataset. *Scientific Data*, 10(1): 740.

Dahal, N.; Nandagopal, N.; Nafalski, A.; and Nedic, Z. 2011. Modeling of cognition using EEG: a review and a new approach. In *TENCON 2011-2011 IEEE Region 10 Conference*, 1045–1049. IEEE.

De Cesarei, A.; and Codispoti, M. 2011. Affective modulation of the LPP and α-ERD during picture viewing. *Psychophysiology*, 48(10): 1397–1404.

Ding, R.; Li, P.; Wang, W.; and Luo, W. 2017. Emotion processing by ERP combined with development and plasticity. *Neural plasticity*, 2017(1): 5282670.

Engel, T. A.; and Wang, X.-J. 2011. Same or different? A neural circuit mechanism of similarity-based pattern match decision making. *Journal of Neuroscience*, 31(19): 6982–6996.

Gu, A.; and Dao, T. 2023. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.

Hang, W.; Feng, W.; Du, R.; Liang, S.; Chen, Y.; Wang, Q.; and Liu, X. 2019. Cross-subject EEG signal recognition using deep domain adaptation network. *IEEE Access*, 7: 128273–128282.

Haynes, J.-D.; and Rees, G. 2006. Decoding mental states from brain activity in humans. *Nature reviews neuroscience*, 7(7): 523–534.

Jiménez-Guarneros, M.; and Gómez-Gil, P. 2020. Custom Domain Adaptation: A new method for cross-subject, EEG-based cognitive load recognition. *IEEE Signal Processing Letters*, 27: 750–754.

Ke, G.; He, D.; and Liu, T. 2021. Rethinking Positional Encoding in Language Pre-training. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*.

Khalighi, S.; Sousa, T.; Santos, J. M.; and Nunes, U. 2016. ISRUC-Sleep: A comprehensive public dataset for sleep researchers. *Computer methods and programs in biomedicine*, 124: 180–192.

Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; and Patras, I. 2011. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing (TAFFC)*, 3(1): 18–31.

Kulkarni, N.; and Bairagi, V. 2017. Extracting salient features for EEG-based diagnosis of Alzheimer's disease using support vector machine classifier. *IETE Journal of Research*, 63(1): 11–22.

Kumar, V.; Reddy, L.; Kumar Sharma, S.; Dadi, K.; Yarra, C.; Bapi, R. S.; and Rajendran, S. 2022. mulEEG: a multi-view representation learning on EEG signals. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 398–407. Springer.

Lang, P. J.; Bradley, M. M.; Fitzsimmons, J. R.; Cuthbert, B. N.; Scott, J. D.; Moulder, B.; and Nangia, V. 1998. Emotional arousal and activation of the visual cortex: an fMRI analysis. *Psychophysiology*, 35(2): 199–210.

Li, J.; Qiu, S.; Shen, Y.-Y.; Liu, C.-L.; and He, H. 2019. Multisource transfer learning for cross-subject EEG emotion recognition. *IEEE transactions on cybernetics*, 50(7): 3281–3293.

Li, W.; Hou, B.; Shao, S.; Huan, W.; and Tian, Y. 2023. Spatial-temporal constraint learning for cross-subject EEG-based emotion recognition. In *2023 International joint conference on neural networks (IJCNN)*, 1–8. IEEE.

Li, W.; Li, H.; Sun, X.; Kang, H.; An, S.; Wang, G.; and Gao, Z. 2024. Self-supervised contrastive learning for EEG-based cross-subject motor imagery recognition. *Journal of Neural Engineering*, 21(2): 026038.

Li, X.; Zhang, Y.; Tiwari, P.; Song, D.; Hu, B.; Yang, M.; Zhao, Z.; Kumar, N.; and Marttinen, P. 2022. EEG based emotion recognition: A tutorial and review. *ACM Computing Surveys*, 55(4): 1–57.

Liu, X.; Yu, H.-F.; Dhillon, I.; and Hsieh, C.-J. 2020. Learning to encode position for transformer with continuous dynamical model. In *International conference on machine learning (ICML)*, 6327–6335. PMLR.

Lv, X.; Sun, Y.; Zhang, K.; Qu, S.; Zhu, X.; Fan, Y.; Wu, Y.; Hua, E.; Long, X.; Ding, N.; et al. 2025. Technologies on Effectiveness and Efficiency: A Survey of State Spaces Models. *arXiv preprint arXiv:2503.11224*.

Mohammadi Foumani, N.; Mackellar, G.; Ghane, S.; Irtza, S.; Nguyen, N.; and Salehi, M. 2024. Eeg2rep: enhancing self-supervised eeg representation through informative masked inputs. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5544–5555.

Motamedi-Fakhr, S.; Moshrefi-Torbati, M.; Hill, M.; Hill, C. M.; and White, P. R. 2014. Signal processing techniques applied to human sleep EEG signals—A review. *Biomedical Signal Processing and Control*, 10: 21–33.

Polich, J. 1997. On the relationship between EEG and P300: individual differences, aging, and ultradian rhythms. *International journal of psychophysiology*, 26(1-3): 299–317.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning (ICML)*, 8748–8763. PmLR.

Radua, J.; Sarró, S.; Vigo, T.; Alonso-Lana, S.; Bonnín, C. M.; Ortiz-Gil, J.; Canales-Rodríguez, E. J.; Maristany, T.; Vieta, E.; Mckenna, P. J.; et al. 2014. Common and specific brain responses to scenic emotional stimuli. *Brain Structure and Function*, 219(4): 1463–1472.

Saha, P.; and Fels, S. 2019. Hierarchical deep feature learning for decoding imagined speech from EEG. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 10019–10020.

Shen, X.; Liu, X.; Hu, X.; Zhang, D.; and Song, S. 2022. Contrastive learning of subject-invariant EEG representations for cross-subject emotion recognition. *IEEE Transactions on Affective Computing (TAFFC)*, 14(3): 2496–2511.

Song, T.; Liu, S.; Zheng, W.; Zong, Y.; and Cui, Z. 2020. Instance-adaptive graph for EEG emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2701–2708.

Tao, W.; Li, C.; Song, R.; Cheng, J.; Liu, Y.; Wan, F.; and Chen, X. 2020. EEG-based emotion recognition via channel-wise attention and self attention. *IEEE Transactions on Affective Computing (TAFFC)*, 14(1): 382–393.

Wang, J.; Zhao, S.; Luo, Z.; Zhou, Y.; Jiang, H.; Li, S.; Li, T.; and Pan, G. 2025. CBraMod: A Criss-Cross Brain Foundation Model for EEG Decoding. In *13th International Conference on Learning Representations, ICLR 2025*.

Wang, Y.; Zhang, B.; and Tang, Y. 2024. DMMR: Cross-subject domain generalization for EEG-based emotion recognition via denoising mixed mutual reconstruction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 628–636.

Wang, Z.; Fei, X.; Liu, X.; Wang, Y.; Hu, Y.; Peng, W.; Wang, Y.-w.; Zhang, S.; and Xu, M. 2022. REM sleep is associated with distinct global cortical dynamics and controlled by occipital cortex. *Nature communications*, 13(1): 6896.

Weng, W.; Gu, Y.; Guo, S.; Ma, Y.; Yang, Z.; Liu, Y.; and Chen, Y. 2024. Self-supervised learning for electroencephalogram: A systematic survey. *ACM Computing Surveys*.

Yang, C.; Westover, M.; and Sun, J. 2023. Biot: Biosignal transformer for cross-data learning in the wild. *Advances in Neural Information Processing Systems (NeurIPS)*, 36: 78240–78260.

Zhang, X.; Zhao, Z.; Tsiligkaridis, T.; and Zitnik, M. 2022. Self-supervised contrastive pre-training for time series via time-frequency consistency. *Advances in neural information processing systems (NeurIPS)*, 35: 3988–4003.

Zhao, L.-M.; Yan, X.; and Lu, B.-L. 2021. Plug-and-play domain adaptation for cross-subject EEG-based emotion recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 863–870.

Zhou, R.; Ye, W.; Zhang, Z.; Luo, Y.; Zhang, L.; Li, L.; Huang, G.; Dong, Y.; Zhang, Y.-T.; and Liang, Z. 2024. Eegmatch: Learning with incomplete labels for semisupervised eeg-based cross-subject emotion recognition. *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*.