

NEURO-MOBRE: EXPLORING MULTI-SUBJECT MULTI-TASK INTRACRANIAL DECODING VIA EXPLICIT HETEROGENEITY RESOLVING

Anonymous authors

Paper under double-blind review

ABSTRACT

Neurophysiological decoding, fundamental to advancing brain-computer interface technologies, has significantly benefited from recent advances in deep learning. However, existing decoding approaches largely remain constrained to single-task scenarios and individual subjects, limiting their broader applicability and generalizability. Efforts towards creating large-scale neurophysiological foundation models have shown promise, but continue to struggle with significant challenges due to pervasive data heterogeneity across subjects and decoding tasks. Simply increasing model parameters and dataset size without explicitly addressing this heterogeneity fails to replicate the scaling successes seen in natural language processing. Here, we introduce the **Neural Mixture of Brain Regional Experts (Neuro-MoBRE)**, a general-purpose decoding framework explicitly designed to manage the ubiquitous data heterogeneity in neurophysiological modeling. Neuro-MoBRE incorporates a brain-regional-temporal embedding mechanism combined with a mixture-of-experts approach, assigning neural signals from distinct brain regions to specialized regional experts on a unified embedding basis, thus explicitly resolving both structural and functional heterogeneity. Additionally, our region-masked autoencoding pre-training strategy further enhances representational consistency among subjects, complemented by a task-disentangled information aggregation method tailored to effectively handle task-specific neural variations. Evaluations on intracranial recordings from 11 subjects across five diverse tasks, including complex language decoding and epilepsy diagnosis, demonstrate that Neuro-MoBRE surpasses prior art and exhibits robust generalization for **cross-subject** decoding on unseen subjects.

1 INTRODUCTION

Neurophysiological signals represent the direct electrical activity generated by the central and peripheral nervous systems, as characterized by the volume conduction theory (Nunez & Srinivasan, 2006). With advancements in biomedical engineering and continuous progress in neuroscience, neurophysiological decoding has increasingly become a critical focus of research, especially in the development of BCIs. Simultaneously, breakthroughs in deep learning, have significantly enhanced the decoding capabilities of BCIs, enabling sophisticated interpretations of human sensory, motor, and communicative functions through implantable technologies (Robinson et al., 2024; Bouton et al., 2016; Metzger et al., 2023; Willett et al., 2023; Hochberg et al., 2012; Benabid et al., 2019). Nonetheless, current methods remain predominantly confined to task-specific decoding scenarios and restricted datasets gathered from individual subjects, thereby excelling only in narrowly defined contexts. This limitation severely constrains their practical utility and generalizability across diverse neurophysiological contexts.

Recent studies (Yang et al., 2023; Zhang et al., 2024; Wang et al., 2023) have explored large-scale neurophysiological foundation models by employing various pre-training techniques with electroencephalogram (EEG) data aggregated from multiple tasks. These approaches aim to replicate the scalability benefits observed in large language models (LLMs) (Kaplan et al., 2020). Techniques such as contrastive learning, masked autoencoding, and vector quantization have been adopted to extract generalized neural representations from extensive datasets (Jiang et al., 2024; Wu et al., 2025). Nevertheless, such methods still require task-specific fine-tuning, restricting their general-

purpose decoding potential. Recently, NeuroLM (Jiang et al., 2025) introduced a multi-task EEG decoding model, framing diverse decoding tasks as a universal question-answering framework via pre-trained LLMs. Despite pioneering multi-task EEG decoding, NeuroLM exhibits substantially lower performance compared to specialized single-task models.

Despite these efforts, the development of large neurophysiological models (LNMs) capable of robust general-purpose decoding remains elusive due to several unresolved challenges:

- (1) **Ubiquitous data heterogeneity in neurophysiological modeling.** Data heterogeneity encapsulates the variability and inconsistency inherent in neurophysiological recordings, arising from biological differences among subjects as well as variations in experimental setups and conditions. For example, neurophysiological signals significantly differ across task paradigms due to distinct neural processes engaged in task-specific recruitment. Current approaches typically lack explicit strategies for addressing data heterogeneity, often limiting themselves to structural adjustments at the model architecture level, such as accommodating differences in channel numbers and signal duration, neglecting effective mechanisms to address neurophysiological variations across subjects and neural activity contexts. Consequently, large-scale modeling of heterogeneous data encounters significant consistency issues, resulting in suboptimal decoding performance.
- (2) **Limited spatiotemporal resolution in non-invasive recordings.** Non-invasive modalities such as EEG suffer from signal attenuation caused by the skull and dura mater, severely constraining their capacity to capture high-resolution neural dynamics. While existing LNMs can decode neural activity for relatively coarse tasks, such as abnormal event detection, they fall short for cognitively complex tasks, including fine-grained somatic motor control and language decoding, where access to high-frequency components (e.g., high-gamma bands) is critical. This limitation is compounded by the field’s reliance on heterogeneous, publicly available datasets collected with diverse protocols, equipment, and subject populations, further exacerbating variability. Together, these factors hinder the development of LNMs capable of consistent, multi-subject, multi-task decoding across domains of high cognitive and motor complexity.

To effectively address the data heterogeneity challenge, we propose a general-purpose neurophysiological decoding framework termed **Neural Mixture of Brain Regional Experts (Neuro-MoBRE)**. Neuro-MoBRE simultaneously performs multi-task decoding of neurophysiological signals from intracranial recordings across multiple subjects. Our fundamental strategy is straightforward: **explicitly resolving consistency conflicts arising from inherent data heterogeneity to achieve unified and robust neural decoding.** Specifically, Neuro-MoBRE incorporates a brain-regional-temporal embedding mechanism within a decoder-only transformer architecture to effectively handle structural heterogeneity. Additionally, it utilizes a mixture-of-experts approach, assigning neural signals from distinct brain regions and individual subjects to specialized regional experts, thus explicitly addressing both task-induced and subject-induced heterogeneity. We further introduce a region-masked autoencoding pre-training scheme designed to unify parameters from independently trained subject-specific models, serving as an effective initialization step for subsequent multi-task and multi-subject decoding. This mechanism significantly enhances representational homogeneity among subjects. Recognizing that distinct tasks activate different brain regions and neural pathways, we also propose a task-disentangled information aggregation method to integrate neural features extracted by various regional experts, thereby further mitigating task-specific heterogeneity and boosting decoding performance.

To circumvent the limitations posed by limited spatiotemporal resolution in non-invasive neurophysiological recordings, we rigorously evaluate Neuro-MoBRE using intracranial data collected from 11 subjects across five distinct decoding tasks. Unlike prior studies focusing primarily on binary or simple multi-class classification tasks, our evaluations span complex decoding scenarios, including language decoding and epileptic seizure diagnosis. Extensive experimental results demonstrate that Neuro-MoBRE surpasses traditional single-task decoding methods and existing large-scale neurophysiological foundation models. Furthermore, Neuro-MoBRE exhibits robust generalization in **cross-subject** decoding for unseen subjects. Collectively, our contributions are as follows:

- (1) We introduce Neuro-MoBRE, a versatile general-purpose decoding framework explicitly designed to resolve inherent data heterogeneity in neurophysiological signals, enabling effective multi-subject and multi-task decoding;

- 108 (2) To facilitate the effective extraction of meaningful information required to manage data hetero-
 109 geneity, we conduct comprehensive evaluations using intracranial recordings from a diverse
 110 cohort of 11 subjects across complex tasks, including phoneme articulation and seizure diagnosis;
 111 (3) Our extensive experimental results validate the superior performance of Neuro-MoBRE, achieving
 112 an average top-1 decoding accuracy improvement of 17.98% compared to existing multi-task
 113 approaches. Additionally, Neuro-MoBRE demonstrates robust **cross-subject** generalization
 114 capabilities to previously unseen subjects.
 115

116 2 RELATED WORK

118 **Intracranial Neurophysiological Decoding.** Intracranial brain recordings, including electrocor-
 119 ticography (ECoG) and stereo-electroencephalography (sEEG), offer unparalleled spatial and tempo-
 120 ral resolution for investigating neural substrates of motor, sensory, and cognitive functions (Bouton
 121 et al., 2016; Benabid et al., 2019; Hochberg et al., 2012; Willett et al., 2023; Metzger et al., 2023).
 122 Recent work by Hochberg et al. (Hochberg et al., 2012) demonstrated that individuals with tetraple-
 123 gia could perform complex robotic arm tasks using signals from implanted microelectrode arrays.
 124 Likewise, advances in speech prosthesis show high-accuracy text or audio decoding from attempted
 125 speech through intracranial recordings (Willett et al., 2023; Metzger et al., 2023; Moses et al., 2021;
 126 Angrick et al., 2019), including systems achieving low word error rates using signals from ventral
 127 premotor and Broca’s areas (Willett et al., 2023). However, prior studies largely focus on single-task,
 128 single-subject scenarios, limiting generalizability. Our work addresses these constraints by integrating
 129 multi-task, multi-subject intracranial recordings to enable robust and generalizable decoding.

130 **Large-scale Neurophysiological Models.** Successes of LLMs (Brown et al., 2020) have inspired
 131 scalable brain decoding models that integrate signals from many subjects and tasks. MMM (Yi et al.,
 132 2023) standardizes EEG across sensor grids via topology-agnostic pretraining; BIOT (Yang et al.,
 133 2023) unifies biosignal representations using a handcrafted tokenizer. Brant (Zhang et al., 2024)
 134 and BrainBERT (Wang et al., 2023) both use masked autoencoding in the frequency domain for
 135 intracranial data, while LaBraM (Jiang et al., 2024) leverages discrete EEG tokenization. However,
 136 these methods require fine-tuning per task and lack true multi-task capability. NeuroLM (Jiang et al.,
 137 2025) reframes decoding as question answering for multi-task use, while H2DiLR (Wu et al., 2025)
 138 explicitly disentangles heterogeneity, though limited to single-task scenarios. Our framework extends
 139 this line with explicit modeling of both task- and subject-level heterogeneity for robust multi-task
 140 decoding; further related work details are provided in the Appendix A.

141 3 METHODOLOGY

142 Neuro-MoBRE consists of three key components designed to explicitly address the inherent data
 143 heterogeneity: a brain-regional-temporal tokenizer, a brain-regional mixture-of-experts module, and
 144 a task-disentangled information aggregation mechanism, each tailored to mitigate different aspects of
 145 heterogeneity (Fig. 1). We begin by introducing the overall decoding paradigm of Neuro-MoBRE in
 146 Sec. 3.1. Furthermore, in Sec. 3.2, we introduce a region-masked autoencoding pre-training (RMAE)
 147 strategy that leverages independently trained subject-specific models to initialize Neuro-MoBRE,
 148 effectively upcycling learned parameters for further heterogeneity resolving.
 149

150 3.1 NEURO-MOBRE OVERVIEW

151 **Problem Formulation.** Suppose a set of neurophysiological recordings \mathcal{X} are collected from
 152 m subjects $\{\mathcal{S}_i\}_{i=1}^m$, across a set of n decoding tasks $\{\mathcal{T}_j\}_{j=1}^n$. Each recording $\mathcal{X}_i \in \mathbb{R}^{N_i \times T_i \times C_i}$
 153 consists of N_i data samples, each with a signal length of T_i and C_i channels for the subject \mathcal{S}_i
 154 with substantial heterogeneity introduced by both neurophysiological variability across subjects and
 155 differences in experimental protocols across tasks. Neuro-MoBRE aims to achieve robust decoding
 156 in a multi-subject, multi-task setting by learning a mapping function $f(\cdot)$ that projects the input
 157 recordings to their corresponding task-specific labels, i.e., $f(\mathcal{X}) \rightarrow \mathcal{Y}_{\mathcal{T}}$ for all decoding tasks.
 158

159 **Brain-regional-temporal Tokenizer.** Since intracranial neurophysiological recordings from differ-
 160 ent subjects typically originate from diverse brain regions, each region potentially having a varying
 161

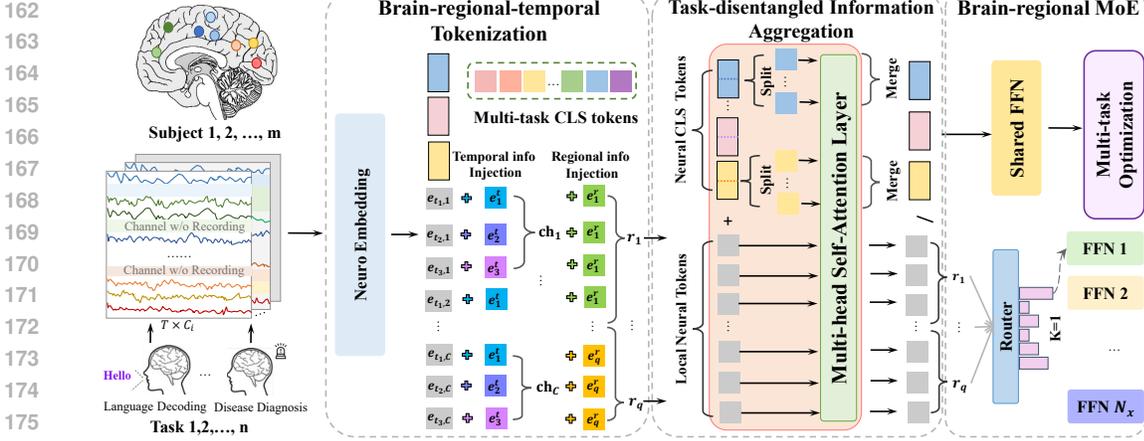


Figure 1: Overview of Neuro-MoBRE for multi-subject, multi-task neural decoding. The brain-regional-temporal tokenizer resolves structural heterogeneity by projecting neural signals collected from diverse recording configurations into a unified representation space, incorporating dedicated temporal and regional tokens (e^t , e^r). Channels and brain regions are denoted by ch and r , respectively. Disentangled “wider” neural CLS tokens aggregate local neural features and enhance global neural representation capacity, capturing diverse functional patterns across distinct cognitive tasks. To further mitigate functional heterogeneity among brain regions and tasks, all CLS tokens share a common FFN, while local tokens are dynamically routed to their associated regional expert networks.

channel count, structural heterogeneity naturally arises. To resolve this, we propose embedding the recordings in a regional-temporal manner. Specifically, for each one of the N_i neuro recordings $X \in \mathbb{R}^{T \times C}$ from recording \mathcal{X}_i , we employ a cascade of 1-D convolutional operations with receptive field size k over the temporal dimension, transforming the input into a sequence of embeddings $E = [\mathbf{e}_{p,c} \in \mathbb{R}^d \mid p = 1, \dots, \lfloor \frac{T}{k} \rfloor; c = 1, \dots, C]$, where each $\mathbf{e}_{p,c}$ represents the p^{th} temporal token corresponding to channel c . To explicitly capture temporal information, we augment these embeddings with learnable temporal positional embeddings $[\mathbf{e}_p^t \mid p = 1, \dots, \lfloor \frac{T}{k} \rfloor]$ for each channel, resulting in the temporally-aware embeddings $[\mathbf{e}_{p,c} + \mathbf{e}_p^t \mid p = 1, \dots, \lfloor \frac{T}{k} \rfloor; c = 1, \dots, C]$. Furthermore, as the intracranial channel configuration does not adhere to a standardized arrangement (such as the EEG 10-20 system), we explicitly inject regional information into each embedding to encode brain region specificity. A set of brain regions \mathcal{R}_j is considered relevant to a decoding task \mathcal{T}_j . We define a set of learnable region-specific embeddings $[\mathbf{e}_q^r \mid q = 1, \dots, |\mathcal{R}_j|]$. For each token $\mathbf{e}_{p,c}$, we integrate regional information by adding the embedding \mathbf{e}_q^r corresponding to the region q from which channel c was recorded:

$$\mathbf{e}_{p,c}^{\text{final}} = \mathbf{e}_{p,c} + \mathbf{e}_p^t + \mathbf{e}_q^r. \quad (1)$$

Brain-regional MoE Block. The architecture of the proposed Neuro-MoBRE follows a typical decoder-only Transformer with layers, each containing a multi-head self-attention (MSA) structure with pre-normalization and a feed-forward network (FFN) with residual connection. To resolve data heterogeneity caused by brain region inconsistency due to brain region involvement and neuro-physiological differences, we replace the FFNs of each layer with a Brain-regional MoE module, which comprises a router and several experts, $\{\text{FFN}_x\}$. In this work, we specifically set the number of experts $N_x = |\{\text{FFN}_x\}|$ equal to the number of brain regions of interest. The rationale behind aligning the number of experts with brain regions is based on the assumption that neural activities across different brain regions, and particularly across different decoding tasks, vary significantly. Thus, tokens corresponding to different brain regions should ideally be processed by specialized experts. The router dynamically determines the assignment of each token to the most suitable expert, allowing each expert network to specialize in capturing distinct neural features associated with specific brain regions and their corresponding functional characteristics. Formally, given an input

neural token $\mathbf{e}_{t,c}^l$ at time stamp t and channel c of the layer l , the computation proceeds as follows:

$$h_{t,c}^l = \text{MSA}(\text{Norm}(\mathbf{e}_{t,c}^l)) + \mathbf{e}_{t,c}^l, \quad (2)$$

$$\bar{h}_{t,c}^l = \text{Norm}(h_{t,c}^l), \quad (3)$$

$$\mathbf{e}_{t,c}^{l+1} = \left(\sum_{x=1}^{N_x} \Psi(\bar{h}_{t,c}^l) \cdot \text{FFN}_x(\bar{h}_{t,c}^l) \right) + h_{t,c}^l, \quad (4)$$

where MSA denotes the multi-head self-attention operation, Norm is a normalization layer, and $\Psi(\cdot)$ represents the routing function indicating the contribution of each expert FFN_x .

To ensure that neural tokens originating from the same channel are consistently processed by a single expert, the routing function $\Psi(\cdot)$ operates in a channel-wise manner:

$$\hat{h}_{t,c}^l = \sum_{t=1}^{\frac{T}{k}} \bar{h}_{t,c}^l, \quad (5)$$

$$\mathcal{D}_c = \text{Softmax}(\hat{h}_{t,c}^l \cdot \Phi), \quad (6)$$

$$\Psi(\bar{h}_{t,c}^l) = \text{TopK}(\mathcal{D}_c, k), \quad (7)$$

where $\Phi \in \mathbb{R}^{d \times N_x}$ represents the trainable weight matrix of the router, and $\text{TopK}(\cdot, k)$ selects the top k largest values, setting all others to zero. Unlike conventional MoE modules commonly employed in large language models, where the dispatch matrix \mathcal{D} is computed based solely on individual tokens, Neuro-MoBRE aggregates neural feature information across the entire channel before routing, ensuring coherent channel-specific expert specialization. An auxiliary loss is adopted to ensure the load balance among experts (Lepikhin et al., 2020; Fedus et al., 2022; Shazeer et al., 2017).

Task-disentangled Information Aggregation. In Transformer-based architectures, a global learnable CLS token is typically employed to aggregate information from other tokens, subsequently processed by a feed-forward network (FFN) for further feature extraction. However, complex neural activities in humans involve distinct brain regions operating collaboratively under diverse functional patterns. Consequently, different neural decoding tasks require tailored neural feature aggregation strategies. To address this, we propose utilizing task-disentangled neural CLS tokens (denoted as $\text{CLS}_{\mathcal{T}_j}$) for each decoding task \mathcal{T}_j , each having a dimensionality J times larger than standard local neural tokens. This approach enhances global neural representation capacity and facilitates more effective neural feature aggregation across brain regions (Fuller et al., 2025; Darcet et al., 2024). Formally, given an input neural token sequence $E \in \mathbb{R}^{\lfloor \frac{T}{k} \rfloor C \times d}$, we initialize n distinct task-specific CLS tokens, each with dimensions $J \cdot d$. Prior to applying the multi-head self-attention (MSA) operation, each neural CLS token is decomposed into J separate d -dimensional tokens. These are concatenated with the existing local neural tokens, forming the expanded token sequence $E^{\text{CLS}} \in \mathbb{R}^{(\lfloor \frac{T}{k} \rfloor C + nJ) \times d}$. Following the MSA operation, neural CLS tokens are separated from the combined sequence to yield:

$$E^{\text{CLS}} \rightarrow (\text{CLS}_{\mathcal{T}_1} \in \mathbb{R}^{Jd}, \dots, \text{CLS}_{\mathcal{T}_n} \in \mathbb{R}^{Jd}, \mathbb{R}^{\lfloor \frac{T}{k} \rfloor C \times d}). \quad (8)$$

Given the fundamental difference between extracting neural patterns from individual brain regions and aggregating neural features for distinct decoding tasks, the neural CLS tokens are processed using a specialized FFN. Conversely, local neural tokens are routed through the FFNs contained within the Brain-regional MoE module. Finally, we employ lightweight task-specific neural decoders that map the processed CLS tokens from the final layer to the corresponding decoding class labels.

3.2 REGIONAL MASKED AUTOENCODING AS CO-UPCYCLING INITIALIZATION

Thus far, we have managed to address the neural heterogeneity from diverse electrode implantation strategies and variations in neural patterns corresponding to distinct neural activities. To further mitigate inter-subject neural variability and establish a unified, generalizable neural feature extractor, we propose a regional masked autoencoding (RMAE) approach, shown in Fig. 2. Specifically, we first pre-train a set of m subject-specific models parameterized by $\{\theta_i\}_{i=1}^m$ via the proposed RMAE task. These models share the original network structure described earlier, except that the brain-regional mixture-of-experts (MoE) block is replaced by a conventional FFN. Subsequently, we ‘‘co-upcycle’’

these individual subject-specific pre-trained models as initialization for the proposed Neuro-MoBRE framework, enhancing its cross-subject generalization capability. Given a neural token sequence E , we randomly mask all tokens of the same brain region with a masking probability ratio r . During pre-training, the masked tokens are replaced by learnable embedding tokens of the same feature dimension d . A lightweight prediction head is then employed to predict the frequency-domain representation, with the training objective defined as the mean squared error between the original and predicted time-domain signals of the masked regions (Wu et al., 2024). We set $r = 0.2$ by default in this work. We provide a detailed calculation of the prediction target in the Appendix D. We then apply “co-upcycle” to the neural feature extraction network components while randomly initializing the brain-region-specific experts to increase their diversity and degrees of freedom, enhancing representational distinctions across different brain regions and tasks during feature extraction.

For Co-upcycle, we consider two variants: naïve averaging and sign-uniformity aggregation (Yadav et al., 2023). Naïve averaging computes the element-wise mean of all subject-specific models, given by $\sum_{i=1}^m \theta_i / m$. In contrast, sign-uniformity co-upcycle first sparsifies each pre-trained model by pruning the lowest-magnitude parameters, setting the smallest $k\%$ to zero, yielding trimmed parameters $\{\hat{\theta}_i\}_{i=1}^m$. For each parameter p , we then establish a consensus sign $\gamma^p = \text{sgn}(\sum_{i=1}^m \hat{\theta}_i^p)$, where sgn denotes the sign function. Defining the index set $\mathcal{I}^p = \{i \mid \text{sgn}(\hat{\theta}_i^p) = \gamma^p\}$, we initialize the parameter p in the “co-upcycled” network as $\theta_{\text{init}}^p = \frac{1}{|\mathcal{I}^p|} \sum_{i \in \mathcal{I}^p} \hat{\theta}_i^p$. The proposed initialization algorithm does not incur any extra computational overhead. A detailed complexity analysis is presented in the Appendix E.

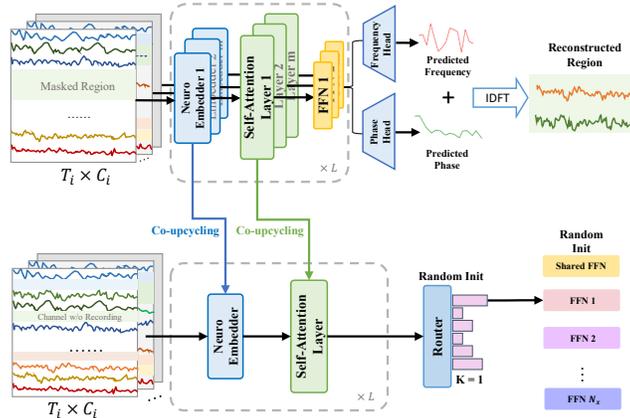


Figure 2: Illustration of the RMAE initialization strategy. Subject-specific neuro-embeddings and attention layers pre-trained via RMAE, utilizing frequency and phase distributions as prediction objectives, are jointly leveraged to initialize the Neuro-MoBRE model parameters. The shared FFN and BrMoE FFN remain randomly initialized to enable effective feature specialization during fine-tuning.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Decoding tasks and data acquisition. In this study, we focus on neural decoding tasks relevant to language decoding and epileptic seizure diagnosis. We recruited 11 participants who were undergoing epilepsy monitoring with implanted stereo-electroencephalography (sEEG) electrodes at anonymous hospitals (hospital names omitted to maintain compatibility with double-blind review). Figure A1 illustrates the electrode distributions across subjects, while Table A1 summarizes participant demographic information. Experimental protocols were approved by institutional review boards, and participants provided written informed consent. 1) **Language decoding task:** Participants read aloud 407 monosyllabic Mandarin characters repeated three times, capturing common Mandarin pronunciations. Mandarin syllables contain three phonological components: initials (23 classes), finals (11 classes), and tones (4 classes), each decoded as separate classification tasks. Given the acoustic similarity and complexity among Mandarin’s 35 phonetic finals, we grouped them into 11 clusters based on acoustic features via k-means clustering Feng et al. (2023), simplifying decoding. For analyses, we excluded the electrodes located in the visual cortex or white matter. 2) **Epileptic seizure diagnosis tasks:** In the epilepsy diagnosis tasks, we addressed both detection and prediction. Seizure detection is formulated as a classification task of differentiating ictal (seizure-active) from interictal (seizure-free) brain states. Seizure prediction task aims to identify electrophysiological patterns or biomarkers indicative of imminent seizure onset, allowing for estimation of seizure likelihood

Table 1: Performance comparison for language decoding tasks. Multi-sub denotes multi-subject. The best and second-best results are highlighted in **bold** and underlined, respectively. “Avg” and “Sgn” to the averaging and sign-uniformity variants of Co-upcycle, respectively.

Methods	Multi-sub	Multi-task	Initial Decoding		Finals Decoding		Tone Decoding	
			Accuracy	Cohen’s Kappa	Accuracy	Cohen’s Kappa	Accuracy	Cohen’s Kappa
ST-Transformer (Song et al., 2021)	✗	✗	0.1178 ± 0.0639	0.0777 ± 0.0668	0.1434 ± 0.0819	0.0983 ± 0.0862	0.2864 ± 0.0910	0.2792 ± 0.0919
FFCL (Li et al., 2022)	✗	✗	0.1055 ± 0.0711	0.0649 ± 0.0744	0.1277 ± 0.0748	0.0818 ± 0.0787	0.2857 ± 0.0884	0.2785 ± 0.0893
CoST (Woo et al., 2022)	✗	✗	0.1443 ± 0.0796	0.1054 ± 0.0832	0.1583 ± 0.0984	0.1140 ± 0.1035	0.2973 ± 0.0736	0.2902 ± 0.0743
SPaRCNet (Jing et al., 2023)	✗	✗	0.0921 ± 0.0761	0.0508 ± 0.0796	0.1087 ± 0.0871	0.0618 ± 0.0917	0.2609 ± 0.0826	0.2534 ± 0.0834
BIOT (Yang et al., 2023)	✓	✗	0.1478 ± 0.0683	0.1091 ± 0.0715	0.1736 ± 0.0899	0.1301 ± 0.0946	0.3209 ± 0.0817	0.3140 ± 0.0825
Brant (Zhang et al., 2023)	✓	✗	0.1423 ± 0.0923	0.1033 ± 0.0964	0.1963 ± 0.0699	0.1540 ± 0.0736	0.3548 ± 0.0837	0.3483 ± 0.0845
NeuroBERT (Wu et al., 2024)	✗	✗	0.1312 ± 0.0620	0.0917 ± 0.0648	0.2029 ± 0.0936	0.1609 ± 0.0985	0.3223 ± 0.08851	0.3154 ± 0.0894
LaBraM (Jiang et al., 2024)	✓	✗	0.1300 ± 0.0660	0.0905 ± 0.0690	0.1942 ± 0.0754	0.1518 ± 0.0794	0.3334 ± 0.0873	0.3267 ± 0.0882
H2DiLR (Wu et al., 2025)	✗	✗	0.1561 ± 0.1229	0.1178 ± 0.1285	0.2169 ± 0.0696	0.1757 ± 0.0733	0.3277 ± 0.0843	0.3209 ± 0.0852
PopT (Chau et al., 2025)	✓	✗	0.0818 ± 0.0751	0.0335 ± 0.0790	0.1686 ± 0.0931	0.1248 ± 0.0980	0.2632 ± 0.0847	0.2557 ± 0.0856
NeuroLM (Jiang et al., 2025)	✓	✗	0.0621 ± 0.0826	0.0194 ± 0.0864	0.1409 ± 0.0889	0.0957 ± 0.0936	0.2766 ± 0.0873	0.2693 ± 0.0882
Neuro-MoBRE (Avg Co-upcycle)	✓	✓	0.2810 ± 0.0898	0.2011 ± 0.0998	0.2860 ± 0.0973	0.2484 ± 0.1024	0.4125 ± 0.0926	0.4051 ± 0.0938
Neuro-MoBRE (Sgn Co-upcycle)	✓	✓	0.2901 ± 0.0933	0.2112 ± 0.1037	0.3161 ± 0.0901	0.2801 ± 0.0949	0.4341 ± 0.0813	0.4284 ± 0.0821

Table 2: Performance comparison for seizure prediction and detection tasks. The best and second-best results are highlighted in **bold** and underlined, respectively. “Avg” and “Sgn” to the averaging and sign-uniformity variants of Co-upcycle, respectively.

Methods	Multi-sub	Multi-task	Seizure Prediction			Seizure Detection		
			Accuracy	Sensitivity	Weighted F1	Accuracy	Sensitivity	Weighted F1
ST-Transformer (Song et al., 2021)	✗	✗	0.6357 ± 0.1002	0.6414 ± 0.1077	0.6356 ± 0.1001	0.7170 ± 0.1003	0.7127 ± 0.1051	0.7170 ± 0.1002
FFCL (Li et al., 2022)	✗	✗	0.6295 ± 0.1154	0.6355 ± 0.1120	0.6293 ± 0.1157	0.6902 ± 0.1121	0.6900 ± 0.1028	0.6901 ± 0.1122
CoST (Woo et al., 2022)	✗	✗	0.6734 ± 0.1022	0.6795 ± 0.1116	0.6731 ± 0.1024	0.7482 ± 0.1122	0.7491 ± 0.1178	0.7480 ± 0.1123
SPaRCNet (Jing et al., 2023)	✗	✗	0.6361 ± 0.1097	0.6255 ± 0.1129	0.6358 ± 0.1097	0.6845 ± 0.0982	0.6745 ± 0.1014	0.6843 ± 0.0984
BIOT (Yang et al., 2023)	✓	✗	0.6595 ± 0.1139	0.6627 ± 0.1161	0.6594 ± 0.1140	0.8011 ± 0.1078	0.7959 ± 0.1107	0.8011 ± 0.1078
Brant (Zhang et al., 2023)	✓	✗	0.6602 ± 0.1143	0.6695 ± 0.1108	0.6600 ± 0.1146	0.7850 ± 0.1000	0.7945 ± 0.0963	0.7848 ± 0.1001
NeuroBERT (Wu et al., 2024)	✗	✗	0.6518 ± 0.1024	0.6536 ± 0.1123	0.6517 ± 0.1025	0.7614 ± 0.1046	0.7550 ± 0.0994	0.7612 ± 0.1046
LaBraM (Jiang et al., 2024)	✓	✗	0.6775 ± 0.1101	0.6795 ± 0.1169	0.6773 ± 0.1102	0.7989 ± 0.1130	0.7991 ± 0.1076	0.7988 ± 0.1131
H2DiLR (Wu et al., 2025)	✗	✗	0.6734 ± 0.1008	0.6827 ± 0.1041	0.6733 ± 0.1008	0.7682 ± 0.1044	0.7655 ± 0.1126	0.7681 ± 0.1045
PopT (Chau et al., 2025)	✓	✗	0.6614 ± 0.1150	0.6695 ± 0.1146	0.6612 ± 0.1149	0.7652 ± 0.0887	0.7600 ± 0.0908	0.7652 ± 0.0887
NeuroLM (Jiang et al., 2025)	✓	✗	0.6495 ± 0.1057	0.6364 ± 0.1166	0.6492 ± 0.1060	0.6927 ± 0.1192	0.6864 ± 0.1319	0.6925 ± 0.1193
Neuro-MoBRE (Avg co-upcycle)	✓	✓	<u>0.7636 ± 0.1068</u>	<u>0.7655 ± 0.1140</u>	<u>0.7636 ± 0.1068</u>	<u>0.8730 ± 0.1069</u>	<u>0.8627 ± 0.1117</u>	<u>0.8729 ± 0.1070</u>
Neuro-MoBRE (Sgn Co-upcycle)	✓	✓	0.7877 ± 0.1044	0.7741 ± 0.1092	0.7877 ± 0.1044	0.8957 ± 0.0900	0.8991 ± 0.0882	0.8957 ± 0.0900

within a predefined future time window. We set a seizure prediction horizon (SPH) at 5 minutes, a preictal interval length (PIL) at 35 minutes Wu et al. (2022b), and defined interictal intervals from 30 minutes after seizure offset to the next preictal interval. Clinical experts annotated seizure onset and offset via visual inspection of sEEG signals. Interictal samples were generated from non-overlapping 20-second sliding windows. For preictal and ictal samples, we employed overlapping windows (50% overlap) to address sample imbalance. We provide detailed signal preprocessing in Appendix C.

Evaluation Protocols. For the language decoding tasks (formulated as multi-class classification problems), decoding performance was evaluated using top-1 accuracy (Acc) and Cohen’s Kappa coefficient. For the epilepsy diagnosis tasks, we assessed top-1 accuracy (Acc), sensitivity (Sen), and weighted F1-score. Regarding dataset partitioning, we employed a balanced sampling strategy to ensure uniform representation across classes within the test set. Specifically, for language decoding tasks, we randomly sampled 10, 20, and 80 instances per class from each participant to construct the test datasets for initial decoding, final decoding, and tone decoding, respectively. This sampling approach resulted in test sets comprising approximately 20% of the total available data. The remaining samples formed the training datasets, from which we further allocated 20% to create the validation sets. For epileptic seizure diagnosis tasks, participant data were partitioned into training (80%), and testing (20%) subsets. From the training subset, an additional 20% was reserved as a validation set. The sliding-window segmentation approach with overlapping windows applied during preprocessing inherently addressed potential class imbalance. All experiments were conducted five times using distinct random seeds. We report performance results as mean values with corresponding standard deviations across these repetitions. We provide implementation details in Appendix H.

4.2 DECODING PERFORMANCE COMPARISON

We compare the performance of our proposed Neuro-MoBRE with baseline methods in Tab. 1 and 2. In language decoding tasks, conventional supervised pre-training baselines achieve decoding performance slightly above chance levels. Baselines employing pre-training consistently outperform those

378
379
380
381
382
383
384
385
386
387

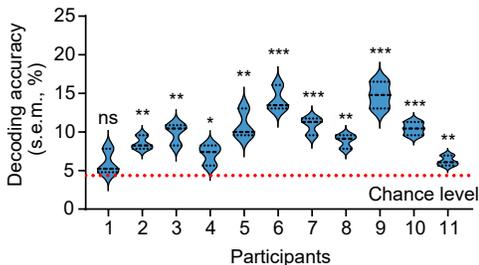


Figure 3: Initial decoding performance of Neuro-MoBRE on unseen subjects.

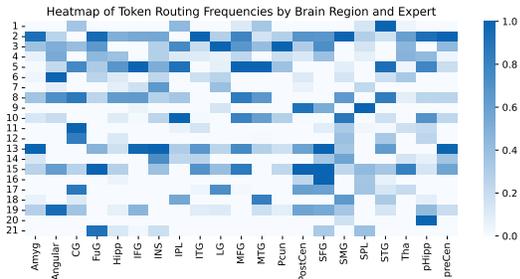


Figure 4: Distribution of brain regions processed by different experts.

391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412

without pre-training, primarily due to enhanced generic feature extraction capabilities. Models such as BIOT, LaBraM, and H2DiLR exhibit superior decoding performance, benefiting from sophisticated architectures designed at both micro and macro levels. However, most existing methods assume fixed electrode placements across subjects to facilitate cross-subject decoding. This assumption limits their practical application in intracranial implant scenarios, where electrode configurations vary significantly between subjects. Consequently, these approaches are marked as incapable of multi-subject decoding in realistic scenarios. Originally developed for tasks like abnormal detection [Harati et al. \(2015\)](#) and sleep stage classification [Alvarez-Estevéz & Rijsman \(2021\)](#), which feature relatively consistent neural patterns (e.g., specific frequency activations), these methods struggle with tasks such as language decoding that involve higher inter-subject neural variability. Although frameworks like BIOT, LaBraM, and NeuroLM address data heterogeneity related to differences in channel counts, sampling rates, and data lengths, they fall short in handling variability stemming from distinct neural activity patterns and substantial inter-subject differences. Despite NeuroLM establishing a multi-task decoding paradigm using a question-answer interaction, it still inadequately captures subtle neural pattern differences, thereby limiting its effectiveness due to unresolved neural data heterogeneity. Our proposed Neuro-MoBRE framework explicitly addresses these sources of neural data heterogeneity throughout its design, leading to significantly improved decoding performance compared to existing approaches. Notably, performance variability in language decoding tasks largely arises from subject-specific differences in phoneme articulation. For example, Participant 4’s lower decoding accuracy (~20%) is primarily due to dialect-influenced articulation challenges, specifically difficulty distinguishing phonemes such as ‘s’ and ‘sh’. In contrast, Participant 10 demonstrates markedly higher decoding accuracy (~60%), reflecting lower phoneme articulation variability.

4.3 ANALYSIS ON NEURO-MOBRE

413
414
415
416
417
418
419
420
421
422
423
424
425

Neuro-MoBRE generalizes to unseen subject. A key advantage of multi-subject, multi-task decoding models over conventional single-task approaches is their improved ability to learn subject-invariant representations and effectively handle inter-subject variability. To systematically evaluate Neuro-MoBRE in terms of generalization to unseen subjects, we conducted experiments following a leave-one-subject-out (LOSO) validation scheme. Specifically, we trained Neuro-MoBRE exclusively on neural data from a subset of subjects and subsequently assessed performance on subjects whose data were completely withheld from training. Our experimental protocol closely follows the implementation details outlined earlier, and we present the decoding performance results for this LOSO evaluation in Fig. 3. Our results demonstrate that Neuro-MoBRE consistently achieves decoding performance significantly above chance-level on previously unseen subjects, underscoring its effectiveness in learning robust, subject-agnostic neural features. Crucially, by successfully generalizing across subject-specific variability, Neuro-MoBRE highlights its potential as a powerful tool for reliable, real-world deployment across diverse, unseen populations in neural decoding tasks.

426
427
428
429
430
431

Regional experts focus on specialized brain regions. To examine whether regional experts within our proposed model exhibit clear preferences toward particular brain regions, we analyzed and visualized the distributions of tokens from these regions as processed by each expert. As illustrated in Fig. 4, the token distribution significantly varies across different experts, demonstrating specialized functional roles. A darker color for a given brain region–expert pair indicates a higher routing intensity, signifying that more tokens from the corresponding brain region are routed to that expert. For instance, Expert One predominantly focuses on processing neural signals from the superior temporal gyrus

Table 3: Ablation study on different components of Neuro-MoBRE.

Component	Decoding Task		
	Initial Decoding	Final Decoding	Tone Decoding
Tokenizer	0.1798	0.2256	0.3270
+ BrMoE	0.2483	0.2798	0.3639
+ TIA	0.2719	0.3025	0.4045
+ RMAE	0.2901	0.3161	0.4341

Table 4: Ablation study on different number of experts.

Number of Experts	Decoding Task		
	Initial Decoding	Final Decoding	Tone Decoding
4	0.2336	0.2696	0.3736
8	0.2526	0.2905	0.3955
16	0.2751	0.3112	0.4185
21	0.2901	0.3161	0.4341

(STG) related to auditory processing related to speech and language, while Experts 2, 8, 13, and 15 are more involved with pre-central and post-central gyrus associated with speech production, assisting motor areas in fine-tuning and coordinating speech movements. Such selective processing further validates that regional experts function by attending to specialized brain areas, aligning effectively with the functional modularity observed in neurobiological studies. This specialization enables accurate extraction of region-specific features and facilitates targeted decoding of neural activities.

4.4 ABLATION STUDY

This section ablates three key designs and the effect of different numbers of experts on the task of language decoding. The average top-1 accuracy for initial decoding is reported, and we follow the same experimental setup as described in Appendix H. Additional ablation study on the scaling effect of network parameters is provided in Appendix I.1.

Effectiveness of the proposed components. Next, we conduct an ablation study to verify the effectiveness of different components proposed in Neuro-MoBRE. The brain-regional-temporal tokenizer is fundamental to accommodating varied electrode implantation schemes and generating unified neural embeddings; thus, we treat it as non-removable and do not ablate this component. Specifically, we assess the contributions of the brain-regional Mixture-of-Experts block (BrMoE), the task-disentangled information aggregation mechanism (TIA), and the regional masked autoencoding strategy (RMAE). As illustrated in Tab. 3, both BrMoE and TIA substantially improve the decoding performance. Moreover, incorporating the proposed RMAE strategy further enhances the decoding results. These findings empirically support our hypothesis that effectively addressing data heterogeneity is essential for multi-subject and multi-task neural decoding.

Ablation on the number of experts. Ideally, neural features from distinct brain regions should be modeled by dedicated experts. We investigate how varying the number of experts impacts overall decoding performance. As observed in Tab. 4, increasing the number of experts generally leads to improved performance, whereas using fewer experts results in performance degradation. Interestingly, comparable decoding accuracies were obtained from 8 and 16 experts. We hypothesize that, with fewer experts than brain regions, Neuro-MoBRE captures collaborative patterns among brain regions that function jointly as coherent groups during specific neural activities. Conversely, when the number of experts exceeds the number of brain regions, multiple experts could be dedicated adaptively to brain regions that contribute more significantly to underlying neural processes.

5 CONCLUSION AND LIMITATION

This paper presents the Neural Mixture of Brain Regional Experts (Neuro-MoBRE), a novel framework designed to explicitly resolve the ubiquitous data heterogeneity encountered in multi-task and multi-subject neurophysiological decoding. Extensive experimental evaluations conducted on stereo-electroencephalography (sEEG) data collected from a cohort of 11 subjects and spanning five distinct decoding tasks, including language decoding and seizure diagnosis, showed that Neuro-MoBRE substantially outperforms existing decoding methods. We list three potential limitations of this work: (1) We have currently verified the effectiveness of Neuro-MoBRE on language decoding and epileptic seizure diagnosis; extending it to more diverse tasks that capture broader neural behaviors could facilitate a deeper understanding of general brain functionality and neural representations. (2) Currently, Neuro-MoBRE treats the neural patterns of an entire brain region as the smallest modeling unit; however, more detailed functional modeling at the level of smaller anatomical areas within each region has yet to be investigated. (3) The inherent complexity and invasiveness involved in acquiring intracranial recordings constrain our capacity to expand the participant cohort at this time.

REFERENCES

- 486
487
488 Diego Alvarez-Estevéz and Roselyne M Rijsman. Inter-database validation of a deep learning
489 approach for automatic sleep scoring. *PLoS one*, 16(8):e0256111, 2021.
- 490 Miguel Angrick, Christian Herff, Emily Mugler, Matthew C Tate, Marc W Slutzky, Dean J Krusienski,
491 and Tanja Schultz. Speech synthesis from ecog using densely connected 3d convolutional neural
492 networks. *Journal of neural engineering*, 16(3):036019, 2019.
- 493
494 Alim Louis Benabid, Thomas Costecalde, Andrey Eliseyev, Guillaume Charvet, Alexandre Verney,
495 Serpil Karakas, Michael Foerster, Aurélien Lambert, Boris Morinière, Neil Abroug, et al. An
496 exoskeleton controlled by an epidural wireless brain-machine interface in a tetraplegic patient: a
497 proof-of-concept demonstration. *The Lancet Neurology*, 18(12):1112–1122, 2019.
- 498
499 Chad E Bouton, Ammar Shaikhouni, Nicholas V Annetta, Marcia A Bockbrader, David A Friedenberg,
500 Dylan M Nielson, Gaurav Sharma, Per B Sederberg, Bradley C Glenn, W Jerry Mysiw, et al.
501 Restoring cortical control of functional movement in a human with quadriplegia. *Nature*, 533
502 (7602):247–250, 2016.
- 503 Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal,
504 Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are
505 few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.
- 506 Geeling Chau, Christopher Wang, Sabera J Talukder, Vighnesh Subramaniam, Saraswati Soedarmadji,
507 Yisong Yue, Boris Katz, and Andrei Barbu. Population transformer: Learning population-level
508 representations of neural activity. In *The Thirteenth International Conference on Learning Repre-*
509 *sentations*, 2025. URL <https://openreview.net/forum?id=FVuqJt3c4L>.
- 510
511 Timothée Darcet, Maxime Oquab, Julien Mairal, and Piotr Bojanowski. Vision transformers need
512 registers. In *The Twelfth International Conference on Learning Representations*, 2024. URL
513 <https://openreview.net/forum?id=2dnO3LLiJl>.
- 514 Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and
515 Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. In
516 *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pp.
517 2352–2359, 2021.
- 518
519 William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter
520 models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39,
521 2022.
- 522 Chen Feng, Lu Cao, Di Wu, En Zhang, Ting Wang, Xiaowei Jiang, Jinbo Chen, Hui Wu, Siyu Lin,
523 Qiming Hou, et al. Acoustic inspired brain-to-sentence decoder for logosyllabic language. *Cyborg*
524 *and Bionic Systems*, 2023.
- 525
526 Anthony Fuller, Yousef Yassin, Daniel G Kyrollos, Evan Shelhamer, and James R Green. Simpler
527 fast vision transformers with a jumbo cls token. *arXiv preprint arXiv:2502.15021*, 2025.
- 528
529 Amir Harati, Meysam Golmohammadi, Silvia Lopez, Iyad Obeid, and Joseph Picone. Improved eeg
530 event classification using differential energy. In *2015 IEEE Signal Processing in Medicine and*
531 *Biology Symposium (SPMB)*, pp. 1–4. IEEE, 2015.
- 532
533 Lauren Julius Harris and Jason B Almerigi. Probing the human brain with stimulating electrodes:
534 the story of roberts bartholow’s (1874) experiment on mary rafferty. *Brain and Cognition*, 70(1):
92–115, 2009.
- 535
536 Leigh R Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Y Masse, John D Simeral, Joern Vogel,
537 Sami Haddadin, Jie Liu, Sydney S Cash, Patrick Van Der Smagt, et al. Reach and grasp by people
538 with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372–375, 2012.
- 539
Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic represen-
tations with tremendous eeg data in bci. *arXiv preprint arXiv:2405.18765*, 2024.

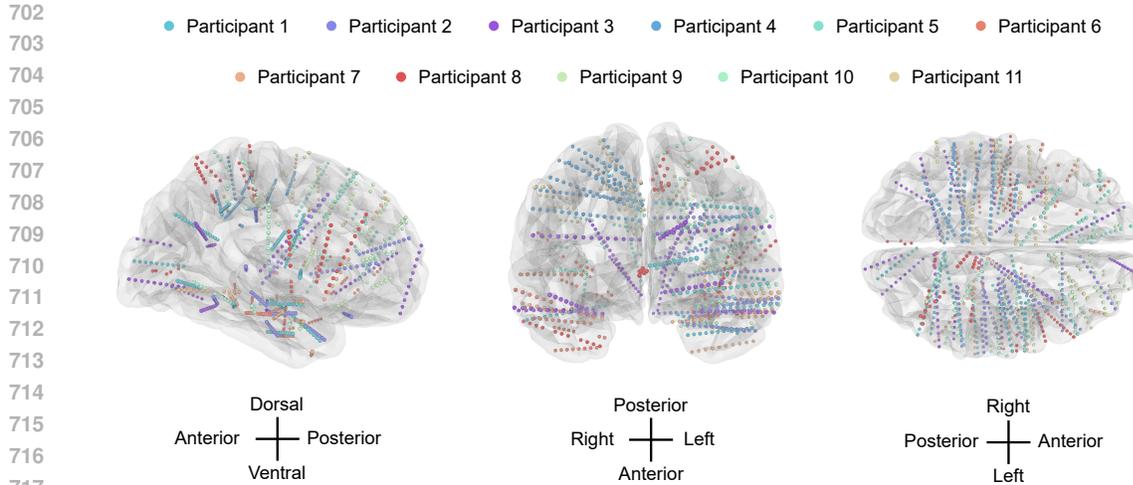
- 540 Weibang Jiang, Yansen Wang, Bao liang Lu, and Dongsheng Li. NeuroLM: A universal multi-task
541 foundation model for bridging the gap between language and EEG signals. In *The Thirteenth*
542 *International Conference on Learning Representations*, 2025. URL [https://openreview.](https://openreview.net/forum?id=Io9yFt7XH7)
543 [net/forum?id=Io9yFt7XH7](https://openreview.net/forum?id=Io9yFt7XH7).
544
- 545 Jin Jing, Wendong Ge, Shenda Hong, Marta Bento Fernandes, Zhen Lin, Chaoqi Yang, Sungtae An,
546 Aaron F Struck, Aline Herlopian, Ioannis Karakis, et al. Development of expert-level classification
547 of seizures and rhythmic and periodic patterns during eeg interpretation. *Neurology*, 100(17):
548 e1750–e1762, 2023.
- 549 Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott
550 Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models.
551 *arXiv preprint arXiv:2001.08361*, 2020.
- 552 Dmitry Lepikhin, HyoukJoong Lee, Yuanzhong Xu, Dehao Chen, Orhan Firat, Yanping Huang,
553 Maxim Krikun, Noam Shazeer, and Zhifeng Chen. Gshard: Scaling giant models with conditional
554 computation and automatic sharding. *arXiv preprint arXiv:2006.16668*, 2020.
- 555
556 Hongli Li, Man Ding, Ronghua Zhang, and Chunbo Xiu. Motor imagery eeg classification algorithm
557 based on cnn-lstm feature fusion network. *Biomedical signal processing and control*, 72:103342,
558 2022.
- 559 Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Confer-*
560 *ence on Learning Representations (ICLR)*, 2019.
- 561
562 Sean L Metzger, Kaylo T Littlejohn, Alexander B Silva, David A Moses, Margaret P Seaton,
563 Ran Wang, Maximilian E Dougherty, Jessie R Liu, Peter Wu, Michael A Berger, et al. A
564 high-performance neuroprosthesis for speech decoding and avatar control. *Nature*, 620(7976):
565 1037–1046, 2023.
- 566 David A Moses, Sean L Metzger, Jessie R Liu, Gopala K Anumanchipalli, Joseph G Makin, Pengfei F
567 Sun, Josh Chartier, Maximilian E Dougherty, Patricia M Liu, Gary M Abrams, et al. Neuroprosthe-
568 sis for decoding speech in a paralyzed person with anarthria. *New England Journal of Medicine*,
569 385(3):217–227, 2021.
- 570 Paul L Nunez and Ramesh Srinivasan. *Electric fields of the brain: the neurophysics of EEG*. Oxford
571 university press, 2006.
- 572
573 Jacob T Robinson, Sumner L Norman, Matthew R Angle, Timothy G Constandinou, Timothy
574 Denison, John P Donoghue, Ryan M Field, Andreas Forsland, Sid Kouider, José del R Millán, et al.
575 An application-based taxonomy for brain–computer interfaces. *Nature Biomedical Engineering*,
576 pp. 1–3, 2024.
- 577
578 Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and
579 Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv*
580 *preprint arXiv:1701.06538*, 2017.
- 581 Yonghao Song, Xueyu Jia, Lie Yang, and Longhan Xie. Transformer-based spatial-temporal feature
582 learning for eeg decoding. *arXiv preprint arXiv:2106.11170*, 2021.
- 583
584 Christopher Wang, Vighnesh Subramaniam, Adam Uri Yaari, Gabriel Kreiman, Boris Katz, Ignacio
585 Cases, and Andrei Barbu. BrainBERT: Self-supervised representation learning for intracranial
586 recordings. In *The Eleventh International Conference on Learning Representations*, 2023. URL
587 https://openreview.net/forum?id=xmcYx_reUn6.
- 588 Francis R Willett, Erin M Kunz, Chaofei Fan, Donald T Avansino, Guy H Wilson, Eun Young
589 Choi, Foram Kamdar, Matthew F Glasser, Leigh R Hochberg, Shaul Druckmann, et al. A high-
590 performance speech neuroprosthesis. *Nature*, 620(7976):1031–1036, 2023.
- 591
592 Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven C. H. Hoi. Cost: Contrastive
593 learning of disentangled seasonal-trend representations for time series forecasting. In *International*
Conference on Learning Representations (ICLR), 2022.

- 594 Di Wu, Siyuan Li, Jie Yang, and Mohamad Sawan. neuro2vec: Masked fourier spectrum prediction
595 for neurophysiological representation learning. *ArXiv*, abs/2204.12440, 2022a.
596
- 597 Di Wu, Jie Yang, and Mohamad Sawan. Bridging the gap between patient-specific and patient-
598 independent seizure prediction via knowledge distillation. *Journal of Neural Engineering*, 19(3):
599 036035, 2022b.
- 600 Di Wu, Siyuan Li, Jie Yang, and Mohamad Sawan. Neuro-bert: Rethinking masked autoencoding
601 for self-supervised neurological pretraining. *IEEE Journal of Biomedical and Health Informatics*,
602 2024.
603
- 604 Di Wu, Siyuan Li, Chen Feng, Lu Cao, Yue Zhang, Jie Yang, and Mohamad Sawan. Towards
605 homogeneous lexical tone decoding from heterogeneous intracranial recordings. In *The Thirteenth
606 International Conference on Learning Representations*, 2025. URL [https://openreview.
607 net/forum?id=cWEfRkYj46](https://openreview.net/forum?id=cWEfRkYj46).
- 608 Prateek Yadav, Derek Tam, Leshem Choshen, Colin A Raffel, and Mohit Bansal. Ties-merging:
609 Resolving interference when merging models. *Advances in Neural Information Processing Systems*,
610 36:7093–7115, 2023.
- 611 Chaoqi Yang, M Brandon Westover, and Jimeng Sun. Biot: Biosignal transformer for cross-data
612 learning in the wild. In *Thirty-seventh Conference on Neural Information Processing Systems*,
613 2023. URL <https://openreview.net/forum?id=c2LZyTyddi>.
614
- 615 Ke Yi, Yansen Wang, Kan Ren, and Dongsheng Li. Learning topology-agnostic eeg representations
616 with geometry-aware modeling. In *Thirty-seventh Conference on Neural Information Processing
617 Systems*, 2023.
- 618 Daoze Zhang, Zhizhang Yuan, Yang Yang, Junru Chen, Jingjing Wang, and Yafeng Li. Brant: Foun-
619 dation model for intracranial neural signal. In *Thirty-seventh Conference on Neural Information
620 Processing Systems*, 2023. URL <https://openreview.net/forum?id=DDk19vaJyE>.
621
- 622 Daoze Zhang, Zhizhang Yuan, Yang Yang, Junru Chen, Jingjing Wang, and Yafeng Li. Brant:
623 Foundation model for intracranial neural signal. *Advances in Neural Information Processing
624 Systems*, 36, 2024.
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647

A DETAILED RELATED WORK

Intracranial Neurophysiological Decoding. Intracranial brain recordings provide an unparalleled view into the brain’s electrical activity, enabling detailed analysis of complex neural processes inaccessible to non-invasive methods such as EEG. These recordings typically involve electrodes placed directly on the brain’s surface (electrocorticography, ECoG) or inserted into deeper brain structures (stereo-electroencephalography, sEEG). Unlike scalp EEG, which is significantly attenuated by the skull and scalp, intracranial recordings offer superior spatial and temporal resolution. This makes them particularly suited for pinpointing localized neural activity and investigating the neural substrates of cognitive, sensory, and motor functions (Bouton et al., 2016; Benabid et al., 2019; Hochberg et al., 2012; Willett et al., 2023; Metzger et al., 2023). The investigation of intracranial brain recordings began as early as 1874 when Roberts Bartholow first documented the effects of electrical stimulation on human brain tissue and its subsequent impact on motor functions (Harris & Almerigi, 2009). More recently, Hochberg et al. (Hochberg et al., 2012) advanced this field by developing a neural interface that enabled two individuals with long-standing tetraplegia to control robotic arms using signals collected from a 96-channel microelectrode array implanted in the motor cortex. This groundbreaking work demonstrated participants’ ability to execute complex tasks, including three-dimensional reach-and-grasp movements, with one participant even successfully drinking from a bottle. Beyond motor restoration, intracranial brain recordings have been instrumental in developing speech prostheses aimed at restoring rapid communication for individuals experiencing paralysis (Willett et al., 2023; Metzger et al., 2023; Moses et al., 2021; Angrick et al., 2019). By decoding neural activity associated with attempted speech, these systems convert neural signals into text or audible speech. Notably, Willett et al. introduced a speech prosthesis utilizing intracranial recordings from microelectrode arrays implanted in the ventral premotor cortex (area 6v) and Broca’s area (area 44). Their system achieved impressive word error rates, 9.1% on a 50-word vocabulary and 23.8% on a 125,000-word vocabulary, representing substantial progress in speech decoding capabilities. Despite these notable advances, most studies have predominantly focused on decoding neural signals from individual subjects engaged in single-task scenarios, limiting the generalizability and broader applicability of these methods. To address this limitation, we propose a unified neurophysiological decoding framework capable of multi-task decoding by integrating intracranial recordings collected from a diverse cohort of 11 subjects, thereby enhancing the model’s versatility and robustness across various neural decoding tasks.

Large-scale Neurophysiological Model. Large language models (LLMs) have demonstrated remarkable abilities as general-purpose task solvers when trained on extensive corpora (Brown et al., 2020). Motivated by these achievements, several researchers have explored methods for learning robust neurophysiological representations by integrating EEG signals collected from multiple subjects across diverse decoding tasks. For instance, MMM (Yi et al., 2023) introduced a topology-agnostic pre-training framework, remapping EEG recordings from various sensor configurations into a standardized 10-10 system topology through multi-dimensional positional encodings, effectively incorporating geometric information. MMM employs a masked autoencoding paradigm at both channel and regional levels as its pretext task. Similarly, BIOT (Yang et al., 2023) proposed a handcrafted tokenizer capable of converting biosignals of arbitrary lengths and channel arrangements into a unified neural representation across subjects, also leveraging channel-wise and temporal-wise masking within a masked autoencoding framework. Furthermore, both Brant (Zhang et al., 2024) and BrainBERT (Wang et al., 2023) adopt masked autoencoding strategies in the frequency domain, utilizing intracranial recordings as inputs. LaBraM (Jiang et al., 2024) introduced discrete tokenization of continuous EEG signals to serve as prediction targets within their masked autoencoding framework. Despite leveraging neurophysiological signals from multiple subjects across varied decoding scenarios, these approaches require extensive fine-tuning for each downstream task and generally lack the capability for effective multi-task decoding. NeuroLM (Jiang et al., 2025) addressed this limitation by reframing EEG decoding as a question-answering problem facilitated by large language models, demonstrating multi-task capabilities. Nevertheless, existing approaches predominantly focus on designing encoder architectures to reconcile differences in data format but often neglect the explicit modeling of data heterogeneity arising from variations between subjects and tasks. Recently, H2DiLR (Wu et al., 2025) addressed heterogeneity explicitly by disentangling homogeneous and heterogeneous components in EEG signals, enabling unified decoding across subjects but restricted to single-task decoding scenarios. In this paper, we introduce a novel multi-task, multi-subject neurophysiological decoding framework equipped with explicit disentanglement mechanisms to



722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739

Figure A1: The anatomy of 11 participants mapped onto the standard Montreal Neurological Institute template brain, with directions indicated.

740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755

Table A1: Demographic and experimental information of participants, comprising sex, age, handedness, and the number of implanted electrodes.

Participant	Sex	Age	Handedness	Number of electrodes	Channel count
1	Male	20	Right-handed	12	63
2	Male	25	Right-handed	8	36
3	Female	19	Right-handed	11	30
4	Female	19	Right-handed	13	66
5	Male	14	Right-handed	9	73
6	Female	17	Left-handed	11	78
7	Female	15	Right-handed	11	74
8	Male	15	Right-handed	12	68
9	Female	26	Right-handed	12	61
10	Male	35	Right-handed	12	88
11	Male	22	Right-handed	10	84

address data heterogeneity stemming from both task-specific and subject-specific factors, enhancing decoding versatility and generalizability.

B DATA ACQUISITION ON LANGUAGE DECODING TASK

Participants were instructed to read aloud a set of 407 monosyllabic Mandarin characters, repeated three times. These characters were selected to comprehensively cover the range of common pronunciations in Mandarin Chinese. Each Mandarin syllable comprises three distinct phonological elements: an initial consonant, a final vowel (or vowel combination), and a lexical tone. Accordingly, we defined the decoding procedure in terms of three classification tasks: initials (23 classes), finals (11 classes), and tones (4 classes). Given that Mandarin contains 35 final phonemes (6 simple finals and 29 compound finals), the high acoustic similarity and dynamic articulation involved rendered direct classification challenging. To overcome this issue, we employed k-means clustering on acoustic formant-based features derived from the original 35 final phonemes [Feng et al. \(2023\)](#). This process grouped acoustically and articulatorily similar finals, resulting in 11 distinct final clusters conducive to improved decoding performance. For all language decoding analyses, we specifically selected channels associated with speech production regions and excluded electrodes located in the visual

cortex or within the white matter. During recording, the participants read aloud the target character displayed on a screen following a countdown visual cue. Each reading trial consisted of a preparation phase lasting 0.8–1.2 seconds, followed by an articulation phase of 2 seconds, and concluded with a rest period of 0.5–1.0 seconds. A comprehensive list of characters along with their corresponding syllables is presented in Tab. A5–A8. These characters were selected to comprehensively cover the range of common pronunciations in Mandarin Chinese. Our reading material has been carefully designed by Mandarin linguists to encompass as many pronunciation phenomena as possible, thereby ensuring that the brain decoding algorithm achieves broad generalizability.

C NEURAL RECORDING PREPROCESSING.

The neural recordings were first converted to microvolts (μV) to ensure consistency in magnitude. The raw data were then filtered using a low-pass filter with a cutoff frequency of 200 Hz to prevent aliasing artifacts. Additionally, notch filters centered at 50 Hz and its harmonics at 100 Hz were applied to remove power-line interference. Following these filtering procedures, the neural signals were downsampled to 512 Hz to reduce computational overhead. Finally, each channel’s data was individually normalized by subtracting its mean and dividing by its standard deviation, standardizing the signals to facilitate subsequent analyses.

D PREDICTION TARGET OF RMAE.

Given a neural token sequence E , we randomly select and mask tokens that belong to the same brain region with a masking ratio r sampled uniformly. During pre-training, the masked tokens are replaced by learnable embedding tokens of the same feature dimension d . Suppose that x is the corresponding discrete neurological signal of the masked patch, we apply the discrete Fourier transform (DFT) as:

$$X_m = \sum_{n=1}^{n=N} x_n * e^{-\frac{2\pi j}{N}mn}, \quad (9)$$

where $m \in [1, N]$. We further decompose equation 9 into its real and imaginary components by applying Euler’s formula:

$$X_m = \sum_{n=1}^{n=N} x_n * \underbrace{\cos\left(\frac{2\pi}{N}mn\right)}_{\text{real}} - j * \underbrace{\sin\left(\frac{2\pi}{N}mn\right)}_{\text{imaginary}}, \quad (10)$$

where j denotes the imaginary unit, defined by $j^2 = -1$. Specifically, X_m represents the spectrum of the sequence x_n evaluated at frequency $\omega_m = 2\pi m/N$. Consequently, we can compute the magnitude $\|X_m\|^2$ and phase θ_m as follows:

$$\begin{aligned} \|X_m\|^2 &= \frac{1}{N} \sqrt{\text{Re}(X_m)^2 + \text{Im}(X_m)^2} \\ \theta_m &= \text{atan2}(\text{Re}(X_m)^2, \text{Im}(X_m)^2), \end{aligned} \quad (11)$$

where $\text{atan2}(\cdot, \cdot)$ denotes the two-argument arctangent function, and $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ represent the real and imaginary parts, respectively. We define our prediction target as the missing amplitude and phase components of the signal. Subsequently, we apply the inverse discrete Fourier transform (IDFT) to convert the predicted amplitude and phase back into the temporal domain:

$$\begin{aligned} \hat{X}_m &= \|\hat{X}_m\|^2 * e^{j*\theta_m} \\ \hat{x}_n &= \sum_{m=1}^{m=N} \hat{X}_m * e^{\frac{2\pi j}{N}mn}. \end{aligned} \quad (12)$$

A lightweight prediction head is then employed to predict the frequency-domain representation, with the training objective defined as the mean squared error between the original and predicted time-domain signals of the masked regions Wu et al. (2024). We set $r = 0.2$ by default in this work.

E TRAINING COMPLEXITY ANALYSIS OF CO-UPCYCLE INITIALIZATION

Please note that the proposed Co-upcycle mechanism is motivated by the hypothesis that corresponding brain regions across individuals exhibit functional homogeneity. Accordingly, Co-upcycle is applied only to the neural embedding layer and attention module parameters pretrained on different subjects via RMAE. In what follows, we analyze the computational and spatial complexity of Co-upcycle.

The overall RMAE Co-upcycle initialization can be divided into two parts: (i) the RMAE pretraining stage and (ii) the parameter aggregation (Co-upcycle) stage. We first examine the computational complexity of the RMAE component. Assume data are collected from m subjects, each with N data samples and let the model contain P parameters. For simplicity, we assume that a single forward-backward training step incurs a constant factor number of FLOPs per parameter per sample. Under the conventional homogeneous pretraining setting, where data from all subjects are pooled and used to train one universal model, as adopted in prior work (Chau et al., 2025; Jiang et al., 2024; Zhang et al., 2023; Yang et al., 2023), the total training complexity is $O(mN \cdot P)$.

In our proposed Co-upcycle initialization, we instead pretrain m subject-specific models, each with complexity $O(NP)$. The total pretraining cost is therefore also $O(m \cdot NP)$, assuming an equal number of epochs in both approaches. In practice, subject-specific models often converge in fewer steps than a single universal model, further reducing the effective training burden. For the Co-upcycle stage, the naïve averaging and sign-uniformity Co-upcycle require only $O(P \cdot m)$ FLOPs in total, which is negligible compared with the $O(m \cdot NP)$ cost of pretraining. Hence, Co-upcycle introduces no additional asymptotic training complexity beyond that of homogeneous pretraining.

Regarding spatial complexity, homogeneous pretraining requires $O(P)$ memory. The naïve averaging variant of Co-upcycle incurs the same spatial cost, as it only computes the element-wise mean across models. In contrast, the sign-uniformity variant in its naïve form requires $O(P \cdot m)$ storage, since it must hold all m subject-specific models simultaneously. However, in realistic scenarios, where data from all m subjects have already been collected, an efficient implementation can be adopted. By maintaining only running aggregates, namely the positive/negative sign counts and corresponding sums, one needs to store only four additional matrices of size P . Consequently, the spatial complexity reduces to $O(P)$, independent of the number of subjects m . Thus, Co-upcycle can be implemented with spatial cost asymptotically identical to homogeneous pretraining, i.e., $O(P)$. This demonstrates that scalability is not constrained by memory, as the framework avoids persistent storage of all subject-specific models.

F NETWORK ARCHITECTURE SETTINGS

We list the detailed description of network architectures in Tab. A2. We set J and K to 4 and 2 by default, respectively. Since most of the baseline methods are not inherently designed to handle data heterogeneity arising from varying channel counts across subjects, we introduce a convolutional layer specifically to project the differing channel dimensions into a standardized channel count. This mapping allows for consistent integration and fair comparison across baseline architectures with varying design specifications.

G BASELINE

We first examine three supervised decoding approaches employing diverse backbone architectures without pre-training (Jing et al., 2023; Li et al., 2022; Song et al., 2021). Additionally, we include two methods utilizing either contrastive learning or masked modeling for pre-training, specifically tailored to single-task decoding scenarios, enabling generic feature extraction prior to downstream decoding tasks (Woo et al., 2022; Wu et al., 2022a). Notably, these baseline approaches exclusively support single-subject and single-task decoding. Subsequently, we explore representative large-scale modeling techniques that integrate neural recordings across multiple subjects or tasks. Specifically, we discuss BIOT (Yang et al., 2023), H2DiLR (Wu et al., 2025), and LaBraM (Jiang et al., 2024), which are general-purpose neural representation learning frameworks pre-trained on extensive EEG data. While these frameworks perform pre-training using multi-subject and multi-task data, they necessitate downstream fine-tuning for specific single-task decoding scenarios. Finally, we consider

Table A2: Detailed architecture specifications for models utilized in this work.

Architecture	Hyperparameters	Values
Tokenizer	Number of filters	{8, 16, 16, 32, 64}
	Kernel size	{15, 7, 5, 3, 3}
	Stride	{7, 4, 3, 2, 2}
Brain-regional MoE	Number of Blocks	4
	Hidden Size	64
	MLP Size	128
	Number of Head	8
Prediction Head	Number of Layer	2
	Hidden Size	256
Reconstructor	Number of Layer	2
	Hidden Size	256

NEUROLM (Jiang et al., 2025), a multi-task neural decoding framework that formulates neural decoding as a question-answering task, leveraging large language models (LLMs) to facilitate simultaneous decoding across multiple tasks.

Table A3: Task instruction designed for NeuroLM implementation on our tasks.

Task	Instruction Description
Initial Decoding	[SEP] Question: Which initial is being articulated? Options: (A): ‘b’, (B): ‘p’, (C): ‘m’, (D): ‘f’, (E): ‘d’, (F): ‘t’, (G): ‘n’, (H): ‘l’, (I): ‘g’, (J): ‘h’, (K): ‘j’, (L): ‘q’, (M): ‘x’, (N): ‘z’, (O): ‘c’, (P): ‘s’, (Q): ‘zh’, (R): ‘ch’, (S): ‘sh’, (T): ‘r’, (U): ‘w’, (V): ‘y’, (W): ‘k’. Answer: {(A), (B), (C), (D), (E), (F), (G), (H), (I), (J), (K), (L), (M), (N), (O), (P), (Q), (R), (S), (T), (U), (V), (W)} [END]
Final Decoding	[SEP] Question: Which final cluster is being articulated? Options: (A): ‘a’, ‘ua’, ‘ia’. (B): ‘uai’, ‘uan’. (C): ‘an’, ‘ai’. (D): ‘en’, ‘eng’, ‘un’. (E): ‘ing’, ‘in’, ‘ian’, ‘ie’. (F): ‘iu’, ‘iao’, ‘iang’. (G): ‘ui’, ‘ei’. (H): ‘ve’, ‘ue’, ‘v’, ‘i’. (I): ‘o’, ‘u’, ‘uo’, ‘ou’. (J): ‘iong’, ‘ong’. (K): ‘ang’, ‘ao’, ‘uang’. (L): ‘e’, ‘er’. Answer: {(A), (B), (C), (D), (E), (F), (G), (H), (I), (J), (K), (L)} [END]
Tone Decoding	[SEP] Question: Which tone is being articulated? Options: (A): ‘First’. (B): ‘Second’ (C): ‘Third’ (D): ‘Fourth’. Answer: {(A), (B), (C), (D), (E), (F), (G), (H), (I), (J), (K), (L)} [END]
Seizure Detection	Question: Is this EEG segment ictal? Answer: {Yes, No} [END]
Seizure Prediction	Question: Is this EEG segment pre-ictal? Answer: {Yes, No} [END]

H IMPLEMENTATION DETAILS

All experiments were performed on GPU workstations equipped with NVIDIA H800 GPUs, utilizing PyTorch-based implementations. For baseline methods with publicly available code (Woo et al., 2022; Eldele et al., 2021; Wu et al., 2022a; Yang et al., 2023; Jiang et al., 2025), we reproduced their experimental results by strictly following the official implementations and configurations provided by the original authors. Specifically, we crafted the task instructions detailed in Tab. A3 to facilitate the implementation of NeuroLM. For baselines lacking publicly released implementations (Jing et al., 2023; Li et al., 2022; Song et al., 2021), we employed the re-implementations offered within the BOIT framework (Yang et al., 2023). Specifically, baseline approaches without pre-training were trained for a fixed duration of 200 epochs. For optimization, we used the AdamW optimizer (Loshchilov & Hutter, 2019) with a base learning rate of 5e-4, momentum parameters of $\beta_1, \beta_2 = 0.9, 0.999$, and a weight decay of 0.01. The learning rate decayed following a cosine annealing schedule. Training employed a batch size equivalent to eight times the number of subjects, and a dropout probability

Table A4: Ablation on scaling-up network parameters (embedding dimension and blocks).

Architecture	Initial Decoding	Final Decoding	Tone Decoding	Seizure Prediction	Seizure Detection
Transformer-4-64	0.2901	0.3161	0.4341	0.7877	0.8957
Transformer-4-128	0.2881	0.3337	0.4384	0.7923	0.8909
Transformer-8-64	0.2846	0.3258	0.4348	0.8032	0.9011
Transformer-16-512	0.1597	0.2208	0.3216	0.6393	0.7059

of 0.1 was consistently applied. For baseline models incorporating pre-training, we strictly adhered to the pre-training setups and experimental protocols specified by their original authors. During the fine-tuning phase for these pre-trained methods, we utilized the same exact training configuration as detailed above. We now describe the pre-training procedures of the proposed Neuro-MoBRE model. During our regional masked autoencoding pre-training phase, optimization was conducted using the AdamW (Loshchilov & Hutter, 2019) optimizer with a base learning rate of $5e-5$, momentum parameters $\beta_1, \beta_2 = 0.9, 0.95$, and weight decay set to 0.05. The learning rate was scheduled via cosine annealing over a fixed training duration of 800 epochs, with a batch size of 32 for each task. Additionally, “co-upcycling” was implemented by pruning 50% of parameters having the lowest magnitudes. Following pre-training, our multi-task neural decoding stage consisted of training for a fixed total of 200 epochs, employing a batch size equal to four per subject per task. In this phase, we used the AdamW optimizer, setting the base learning rate to $5e-4$, momentum parameters $\beta_1, \beta_2 = 0.9, 0.95$, and a higher weight decay of 0.1.

I ADDITIONAL ABLATION STUDY

I.1 ABLATION ON NETWORK PARAMETERS.

We further investigate the scaling behavior of our model by systematically varying the embedding dimension and number of layers (blocks). Neuro-MoBRE-N-D denotes a model configuration consisting of N layers and embedding dimension D , with Transformer-4-64 serving as the baseline configuration. The average top-1 accuracy is reported, and we follow the same experimental setup as described in Appendix H. As illustrated in Tab. A4, modest increases in network depth and embedding dimension consistently yield improvements in decoding performance. However, excessively scaling the model size, such as utilizing 16 blocks combined with a 512-dimensional embedding, results in a marked degradation in performance, indicative of model collapse. We hypothesize this phenomenon arises due to insufficient training data relative to the increased number of model parameters. Such an observation aligns with our assertion that, given an effective resolution of data heterogeneity, excessively large models are not required to achieve high-quality representations in multi-subject, multi-task neurophysiological decoding scenarios.

J BROADER IMPACT

The Neuro-MoBRE model proposed in this study holds significant potential implications across scientific, clinical, and societal domains. By explicitly addressing data heterogeneity challenges inherent to neurophysiological decoding, Neuro-MoBRE substantially enhances the robustness and precision of brain-computer interfaces (BCIs). This advancement contributes directly to fundamental neuroscience research, offering deeper insights into neural network organization and functional dynamics across different brain regions and tasks.

Clinically, Neuro-MoBRE’s ability to accurately and reliably decode intracranial neurophysiological signals across diverse subjects and tasks has crucial implications. Improved decoding accuracy can substantially enhance diagnostic precision, particularly in detecting epileptic seizures, thereby facilitating timely and effective medical interventions. Moreover, decoding complex neural activities such as language production and motor functions can significantly advance rehabilitation strategies, improve prosthetic device control for patients with motor impairments, and enhance communication assistance technologies for individuals experiencing speech difficulties.

Table A5: The first set of 407 Mandarin Chinese characters used as reading material, with corresponding syllables organized by initial phonetic order. For infrequently used characters, participants were presented directly with the syllable to pronounce.

Characters	Syllables	Characters	Syllables	Characters	Syllables
阿	ā	扯	chě	吊	diào
爱	ài	趁	chèn	叠	dié
安	ān	城	chéng	顶	dǐng
昂	áng	痴	chī	丢	diū
袄	ǎo	虫	chóng	东	dōng
拔	bá	愁	chóu	豆	dòu
白	bái	初	chū	读	dú
板	bǎn	chuā	chuā	短	duǎn
帮	bāng	揣	chuāi	对	duì
保	bǎo	穿	chuān	蹲	dūn
杯	bēi	床	chuáng	夺	duó
本	běn	吹	chuī	鹅	é
崩	bēng	春	chūn	恩	ēn
鼻	bí	戳	chuō	ēng	ēng
编	biān	次	cì	耳	ěr
标	biāo	葱	cōng	罚	fá
瘪	biě	凑	còu	反	fǎn
宾	bīn	粗	cū	方	fāng
病	bìng	窜	cuàn	肥	féi
伯	bó	脆	cùi	粉	fěn
补	bǔ	存	cún	风	fēng
擦	cā	搓	cuō	福	fú
财	cái	打	dǎ	否	fǒu
残	cán	带	dài	付	fù
仓	cāng	胆	dǎn	尬	gà
槽	cáo	党	dǎng	盖	gài
测	cè	到	dào	敢	gǎn
参	cān	德	dé	缸	gāng
层	céng	得	děi	告	gào
茶	chá	dèn	dèn	割	gē
柴	chái	等	děng	给	gěi
产	chǎn	底	dǐ	根	gēn
唱	chàng	爹	diē	耕	gēng
超	chāo	电	diàn	共	gòng

Table A6: The second set of 407 Mandarin Chinese characters used as reading material, with corresponding syllables organized by initial phonetic order.

Characters	Syllables	Characters	Syllables	Characters	Syllables
狗	gǒu	金	jīn	冷	lěng
估	gū	镜	jìng	力	lì
瓜	guā	窘	jiǒng	俩	liǎ
怪	guài	酒	jiǔ	连	lián
关	guān	举	jǔ	良	liáng
广	guǎng	捐	juān	料	liào
贵	guì	决	jué	列	liè
滚	gǔn	俊	jùn	林	lín
裹	guǒ	卡	kǎ	领	lǐng
哈	hā	开	kāi	柳	liǔ
孩	hái	砍	kǎn	龙	lóng
寒	hán	糠	kāng	楼	lóu
航	háng	靠	kào	路	lù
耗	hào	科	kē	卵	luǎn
河	hé	克	kè	轮	lún
黑	hēi	肯	kěn	罗	luó
恨	hèn	坑	kēng	滤	lǜ
恒	héng	孔	kǒng	略	lüè
烘	hōng	口	kǒu	马	mǎ
吼	hǒu	库	kù	买	mǎi
虎	hǔ	夸	kuā	慢	màn
画	huà	快	kuài	忙	máng
坏	huài	款	kuǎn	毛	máo
换	huàn	狂	kuáng	美	měi
慌	huāng	亏	kuī	门	mén
悔	huǐ	捆	kǔn	猛	měng
昏	hūn	阔	kuò	米	mǐ
货	huò	拉	lā	面	miàn
机	jī	来	lái	苗	miáo
价	jià	懒	lǎn	灭	miè
尖	jiān	浪	làng	民	mín
浆	jiāng	老	lǎo	命	mìng
交	jiāo	勒	lēi	谬	miù
姐	jiě	雷	léi	魔	mó

Table A7: The third set of 407 Mandarin Chinese characters used as reading material, with corresponding syllables organized by initial phonetic order. For infrequently used characters, participants were presented directly with the syllable to pronounce.

Characters	Syllables	Characters	Syllables	Characters	Syllables
谋	móu	盆	pén	乳	rǔ
木	mù	碰	pèng	刖	ruá
拿	ná	皮	pí	软	ruǎn
奶	nǎi	偏	piān	锐	ruì
男	nán	票	piào	润	rùn
囊	náng	瞥	piē	弱	ruò
闹	nào	品	pǐn	洒	sǎ
讷	nè	瓶	píng	赛	sài
内	nèi	破	pò	伞	sǎn
嫩	nèn	剖	pōu	噪	sǎng
能	néng	普	pǔ	骚	sāo
泥	ní	骑	qí	涩	sè
年	nián	掐	qiā	森	sēn
娘	niáng	浅	qiǎn	僧	sēng
尿	niào	枪	qiāng	沙	shā
捏	niē	桥	qiáo	筛	shāi
您	nín	窃	qiè	闪	shǎn
凝	níng	琴	qín	伤	shāng
牛	niú	情	qíng	烧	shāo
农	nóng	穷	qióng	赊	shē
糲	nòu	球	qiú	谁	shuí
奴	nú	取	qǔ	神	shén
暖	nuǎn	劝	quàn	绳	shéng
挪	nuó	缺	quē	石	shí
女	nǚ	群	qún	手	shǒu
虐	nüè	然	rán	书	shū
哦	ó	让	ràng	刷	shuā
藕	ǒu	饶	ráo	帅	shuài
爬	pá	热	rè	栓	shuān
牌	pái	忍	rěn	爽	shuǎng
盘	pán	扔	rēng	水	shuǐ
旁	páng	日	rì	顺	shùn
抛	pāo	容	róng	硕	shuò
赔	péi	肉	ròu	死	sǐ

1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187

Table A8: The fourth set of 407 Mandarin Chinese characters used as reading material, with corresponding syllables organized by initial phonetic order.

Characters	Syllables	Characters	Syllables	Characters	Syllables
松	sōng	午	wǔ	早	zǎo
搜	sōu	洗	xǐ	则	zé
酥	sū	霞	xiá	贼	zéi
算	suàn	险	xiǎn	怎	zěn
岁	suì	向	xiàng	增	zēng
损	sǔn	笑	xiào	渣	zhā
锁	suǒ	写	xiě	债	zhài
塔	tǎ	心	xīn	展	zhǎn
抬	tái	形	xíng	涨	zhǎng
谈	tán	熊	xióng	找	zhǎo
躺	tǎng	修	xiū	遮	zhē
讨	tǎo	许	xǔ	这	zhè
特	tè	选	xuǎn	针	zhēn
藤	téng	学	xué	蒸	zhēng
替	tì	寻	xún	直	zhí
田	tián	芽	yá	肿	zhǒng
条	tiáo	烟	yān	州	zhōu
铁	tiě	养	yǎng	煮	zhǔ
停	tíng	药	yào	抓	zhuā
桶	tǒng	野	yě	拽	zhuāi
偷	tōu	衣	yī	砖	zhuān
图	tú	银	yín	撞	zhuàng
团	tuán	鹰	yīng	追	zhuī
腿	tuǐ	哟	yō	准	zhǔn
吞	tūn	永	yǒng	捉	zhuō
拖	tuō	油	yóu	字	zì
挖	wā	雨	yǔ	总	zǒng
外	wài	元	yuán	走	zǒu
万	wàn	月	yuè	组	zǔ
忘	wàng	云	yún	钻	zuān
围	wéi	杂	zá	醉	zuì
文	wén	栽	zāi	尊	zūn
翁	wēng	暂	zàn	左	zuǒ
窝	wō	葬	zàng		

1188 From an applicability and accessibility standpoint, the generalized decoding capabilities of Neuro-
1189 MoBRE across multiple tasks and subjects have the potential to significantly expand the accessibility
1190 of neurotechnological solutions. Improved cross-subject decoding performance on previously unseen
1191 individuals greatly reduces system calibration requirements, allowing more effective personalization
1192 and broader practical deployment.

1193 Nevertheless, despite the promising advancements presented, several ethical and societal considera-
1194 tions must be addressed. Enhanced decoding capabilities pose risks concerning patient privacy and
1195 data security, necessitating rigorous ethical guidelines, robust data protection measures, and trans-
1196 parent informed consent protocols. Additionally, the increased accuracy and potential widespread
1197 adoption of BCIs could lead to misuse or unauthorized surveillance, highlighting the need for compre-
1198 hensive policy frameworks, advanced encryption methods, and clear guidelines for responsible usage.
1199 Finally, the current reliance on invasive intracranial data collection methods introduces inherent proced-
1200 ural risks. Future research should prioritize validating similar models with high-quality non-invasive
1201 data, reducing invasiveness, and improving safety and accessibility for broader populations.

1202

1203

1204 K DATA AVAILABILITY

1205

1206 Data access will be provided to qualified researchers upon reasonable request to the corresponding
1207 author, subject to standard data use agreements to ensure ethical and legal compliance.

1208

1209

1210

1211

1212

1213

1214

1215

1216

1217

1218

1219

1220

1221

1222

1223

1224

1225

1226

1227

1228

1229

1230

1231

1232

1233

1234

1235

1236

1237

1238

1239

1240

1241