# EPILEPSYFM: FOUNDATION MODEL FOR LEARNING GENERALIZED EPILEPTIC REPRESENTATIONS FROM EEG AND SEEG SIGNALS

Anonymous authors

006

008 009 010

011

013

014

015

016

017

018

019

021

023

024

025

026

027

028

029

031

032

034

037 038 Paper under double-blind review

### ABSTRACT

Extracranial electroencephalography (EEG) and intracranial stereoelectroencephalography (SEEG) are crucial for epilepsy diagnosis. However, existing deep learning models often limit themselves to specific signal types and application scenarios, leading to challenges in generalization and perception capabilities. While large language models excel in natural language processing, they cannot effectively capture the disease-specific signal features in the highly specialized field of epilepsy, and the lack of pre-training data restricts their generalization ability. To address these issues, we propose a Epilepsy Foundation Model (EpilepsyFM), a domain-specific foundational model that considers the mechanisms of seizure and propagation in epilepsy. EpilepsyFM learns a generalized representation of epilepsy through unsupervised pre-training across various signal types, data formats, and sources, and optimizes multiple epilepsy-related downstream tasks through fine-tuning. We collected clinical EEG and SEEG data from multiple patients at a first-class hospital, as well as the currently largest publicly available epilepsy dataset, the TUH series, ensuring diversity in representation learning. First, the neural activity signals are segmented into multiple patches, and a discrete EEG and SEEG neural tokenizer is trained to construct a domain-specific neural codebook for epilepsy. Then, EpilepsyFM takes into account the mechanisms of clustered neuronal discharges in epilepsy and designs a channel set masking strategy to enhance the model's ability to capture the spatiotemporal characteristics of the signals. The model fully utilizes the multi-dimensional propagation characteristics of seizures through temporal, spectral, and spatial encoder modules, achieving comprehensive representation of complex neural signals. Extensive experiments show that EpilepsyFM achieves state-of-the-art performance in a variety of domain-specific tasks, including seizure detection and both short-term and long-term predictions of neural signals, demonstrating strong generalization ability and broad clinical application potential.

- 039 1 INTRODUCTION
- 040

041 Epilepsy is a chronic neurological disorder caused by abnormal discharges of brain neurons, affecting 042 approximately 70 million people worldwide, making it one of the most common neurological diseases 043 (Asadi-Pooya et al., 2023; Specchio et al., 2022). Its pathological mechanisms are complex and 044 varied, with diverse clinical manifestations. Persistent epileptic seizures not only endanger patients' health but also place a heavy burden on medical resources (Kanner & Bicchi, 2022). Extracranial electroencephalography (EEG) and intracranial stereoelectroencephalography (SEEG) are effective 046 neurosurgical methods for recording brain activity in epilepsy patients during ictal and interictal 047 periods, with intracranial SEEG widely regarded as the gold standard for diagnosing epileptic seizures 048 (Supriya et al., 2021; Narasimhan et al., 2020; Yan et al., 2024). 049

Currently, deep learning models for epilepsy are typically designed for specific neural signals and
application scenarios (Shoeibi et al., 2021; Usman et al., 2021; Peh et al., 2023). Although there has
been some progress in seizure detection and prediction using EEG, such as using graph convolutional
networks (Li et al., 2020) and generative adversarial networks (Rasheed et al., 2021), two core issues
remain: first, clinical epilepsy EEG and SEEG are scarce, particularly the limited availability of

SEEG as the gold standard, which prevents many models from being thoroughly validated in clinical applications; second, most existing models are designed for specific tasks, performing exceptionally well on single tasks but lacking generalization capability in real clinical settings.

057 In recent years, the rise of general-purpose large language models (LLMs) (Vaswani, 2017; 058 Thirunavukarasu et al., 2023; Yuan et al., 2024a), such as ChatGPT, has demonstrated their pow-059 erful capabilities in addressing multi-task and multi-domain problems. Compared to traditional 060 deep learning models, LLMs effectively overcome the limitations of single-task models through 061 pre-training and fine-tuning. In the field of medical information, LLMs provide intelligent support for 062 disease diagnosis by processing vast datasets, significantly improving healthcare efficiency (Li et al., 063 2023; Thirunavukarasu et al., 2023). Recent studies have introduced various foundational models 064 for analyzing brain neural activity, showcasing exceptional performance. For example, Cui et al. (Cui et al., 2024) proposed Neuro-GPT, which combines an EEG encoder with a GPT architecture 065 to address challenges related to the scarcity and heterogeneity of EEG data. Yi et al., (Yi et al., 066 2023) aimed to bridge gaps between different EEG resources, introducing the MMM framework 067 for cross-dataset pre-training, while Jiang et al. (Jiang et al., 2024) presented a unified model that 068 successfully navigates issues such as mismatched EEG electrode counts. 069

Although general-purpose LLMs perform exceptionally well, domain-specific models often demon-071 strate superior performance in specialized fields (Pal et al., 2024; Zhang et al., 2024; Arefeen et al., 2024). These models are optimized to capture unique features relevant to their domains. For example, 072 Gu et al. (Gu et al., 2021) highlighted the advantages of training domain-specific language models 073 from scratch in biomedical natural language processing, while Xie et al. (Xie et al., 2023) proposed a 074 cost-effective strategy that successfully built domain-specific models, achieving significant improve-075 ments in financial tasks. For epilepsy, a clinical neurological disorder, we hope to break through the 076 problem that deep learning can only solve a single task, and avoid the limitation of relying only on 077 LLMs related to generalized brain activity to deal with various data sources (e.g., motor imagery, sleep detection, etc.). Our goal is to design a clinical-focused foundation model that concentrates 079 on epilepsy. This model will learn the generalized representations of intra- and extracranial signals associated with epilepsy, aiming to effectively tackle various tasks related to the condition. However, 081 we still face several challenges in this process:

1) Lack of sufficient EEG and SEEG data for epilepsy: Although several publicly available
 EEG datasets for epilepsy exist, SEEG data remains extremely scarce. Collecting EEG data poses
 significant challenges due to individual differences in seizures and the considerable time and effort
 required for expert annotation.

2) Lack of consideration for the mechanisms of seizure: The mechanisms of epilepsy are complex and involve abnormal discharges from various brain neurons. Improving the model's understanding of the dynamic changes during seizure processes remains a critical challenge.

3) Lack of consideration for the propagation process of epilepsy: After a seizure, the propagation process spreads over time to different brain regions, accompanied by low-amplitude, high-frequency oscillations. This process involves information across time, space, and frequency. Capturing signal changes during this propagation and learning the features of EEG activity related to epilepsy are key to building an efficient model.

To address aforementioned challenges, our goal is to design a domain-specific foundational model 095 for epilepsy, termed EpilepsyFM, which considers seizure onset and its propagation mechanisms, 096 effectively handling EEG and SEEG data while achieving excellent performance across different epilepsy task states. We first segment neural activity signals into multiple patches and, prior to 098 pre-training, leverage the concept of VQ-VAE (Van Den Oord et al., 2017) to train a discrete neural tokenizer, thereby constructing a neural vocabulary specific to the epilepsy domain. Subsequently, 100 during the pre-training phase, we design a channel set masking strategy that reflects the characteristics 101 of clustered epileptic discharges, randomly masking sets of neural patches and predicting the masked 102 patches from visible ones, enhancing the model's learning capability and accelerating the training 103 speed through a symmetric masking strategy. Additionally, EpilepsyFM incorporates temporal, 104 spatial, and spectral encoding methods aimed at capturing the dynamic features during the propagation 105 process of epilepsy. The main contributions of this work are summarized as follows:

1) Robust data foundation: We collected clinical EEG and SEEG epilepsy data from a a first-class hospital's neurosurgery department and combined it with the largest publicly available epilepsy

EEG dataset, the TUH series, to serve as the pre-training foundation for EpilepsyFM. The labels for downstream tasks are meticulously annotated by two specialized physicians from the hospital.

2) Channel set masking pre-training strategy: In response to the clustered discharge mechanisms observed during epileptic seizures, we designed a self-supervised pre-training strategy utilizing localized channel set masking. This approach enables more accurate capture of neural activity both within and across channels while improving computational efficiency through symmetric masking.

3) Temporal-spatial-spectral encoding: To gain deeper insights into the propagation mechanisms of epilepsy, we developed temporal, spatial, and spectral encoding methods. These techniques dy-namically capture the features of neural activity associated with epilepsy, providing a comprehensive understanding of its complex dynamics.

4) Diverse downstream tasks: We conducted comprehensive experimental studies on several key tasks related to epilepsy, including seizure detection, as well as short-term and long-term signal prediction. The results indicate EpilepsyFM's adaptability across different tasks, highlighting its broad potential for clinical applications.

### 2 Method

The overall architecture of EpilepsyFM is depicted in Figure 1. In this architecture, EEG and SEEG
 signals with varying channel counts and temporal lengths are input into the model to learn generalized
 representations of epileptic neural activity, which are then encoded for downstream tasks.



Figure 1: The overall architecture of EpilepsyFM, the Epileptic Neural Transformer. EEG or SEEG signals are first segmented into fixed-size patches. A temporal encoder and a spectral encoder extract time and frequency features from each patch, which are then concatenated. The spatial encoder captures channel information and integrates the temporal and spatial embeddings into patch embeddings. Finally, these embeddings are fed into the transformer encoder in patch order to produce the final output.

147 148

123 124

125

### 2.1 DEFINITIONS AND NOTATIONS

149 The large publicly available epilepsy EEG datasets TUEP, TUSL, and TUSZ, along with the collected 150 clinical long-term EEG and SEEG datasets (XJSZ-EEG and XJSZ-SEEG), are utilized as pre-training 151 data (for more details, see Appendix B). The multi-channel EEG and SEEG signals are represented 152 as  $X \in \mathbb{R}^{C \times P}$ , where C represents the number of EEG and SEEG channels, and P represents 153 the number of sampling points with a sampling rate of S. The set of channels used in EEG and 154 SEEG is defined as  $C_X = \{c_{i1}, c_{i2}, \dots, c_{iC}\}$  for EEG and  $C'_X = \{c_1, c_2, \dots, c_{C'}\}$  for SEEG, 155 where  $C_X \subseteq C_{EEG}$  and  $C'_X \subseteq C_{SEEG}$ , with  $C_{EEG}$  being the complete set of channels defined by 156 the international 10-20 system (Jurcak et al., 2007), and  $C_{SEEG}$  representing the set of all SEEG 157 channels actually used in clinical experiments.

158 159

160

2.2 MODEL ARCHITECTURE

161 We introduce a general architecture for learning epileptic neural activity — Epileptic Neural Transformer, which can handle scalp EEG signals and intracranial SEEG signals with different data formats,



Figure 2: Training of Epilepsy EEG&SEEG Tokenizers and EpilepsyFM Pre-training. (a) EpilepsyFM Tokenizer training: The tokenizer discretizes epileptic EEG and SEEG signals into discrete neural tokens by decoding epileptic neural activity. (b) EpilepsyFM Pre-training: During pretraining, patches of epileptic EEG and SEEG are masked according to the channel set, with the objective of predicting the masked neural tokens from the unmasked patches.

189

190 191

197

208 209

183

184

185

channel numbers, and sampling points. The model framework is illustrated in Figure 1. For each sample x, we divide it into N patches using a non-overlapping time window W (Bao et al., 2021), resulting in  $x = \{x_{c_{ij},k} \in \mathbb{R}^{C \times W} | j = 1, 2, ..., C, k = 1, 2, ..., \lfloor \frac{P}{W} \rfloor\}$ , where  $N = C \lfloor \frac{P}{W} \rfloor$ .

Temporal Encoder. Due to the abnormal temporal variations in EEG and SEEG signals during
 epileptic seizures, we propose a temporal encoder to learn temporal features before the self-attention
 mechanism interacts between patches. The encoder consists of multiple temporal convolution blocks,
 each composed of a convolution layer, a group normalization layer, and a GELU activation function
 (Hendrycks & Gimpel, 2016). The output of the temporal encoder can be expressed as:

$$\varepsilon_t = \left\{ e_{c_{ij},k}^t \in \mathbb{R}^d | j = 1, 2, \dots, C, k = 1, 2, \dots, \lfloor \frac{P}{W} \rfloor \right\}$$
(1)

where d is the dimension of the embeddings.

**Spectral Encoder.** Due to epileptic ictal or interictal periods, patients' brain activity signals exhibit characteristic waves such as spikes, sharp waves, and sharp-slow waves, which are high-frequency, low-amplitude oscillations, all accompanied by changes in frequency. It is therefore necessary to analyze the signal variations at different frequency components during epileptic ictal or interictal periods. To this end, we propose a wavelet packet convolution (WavepConv) to extract frequency features from EEG and SEEG signals (Wang et al., 2021). In the aforementioned approach, the input to the spectral encoder can be expressed as  $e_i^f$ , where:

$$\varepsilon_w = e_i^f (N - \frac{R}{2} + 1), \dots, e_i^f (N - 1) \widehat{\mathbb{C}} e_i^f(0), \qquad (2)$$
$$\dots, e_i^f (N - 1) \widehat{\mathbb{C}} e_i^f(0), \dots, e_i^f (\frac{R}{2} - 2)$$

211  
212  
213  

$$y_A(t) = \sum_{r=0}^R \varepsilon_w(s \times i - r) \times g(r)$$
  
 $y_D(t) = \sum_{r=0}^R \varepsilon_w(s \times i - r) \times h(r)$ 
(3)

where,  $\varepsilon_w$  represents the output of the periodic padding. Periodic padding ensures the minimum length of wavelet packet decomposition, preserving the original characteristics of the input signals and avoiding boundary effects. © represents the concatenating operation. R, s represent the size of the convolution kernel and stride, respectively.  $y_A(t)$  and  $y_D(t)$  represent the approximate coefficients and fine coefficients, respectively. And g, h are a pair of convolution kernels (Shi et al., 2020).

**Spatial Encoder.** Due to the characteristic of abnormal discharge signals spreading from the epileptic focus to other brain regions over time during a seizure, we propose a spatial encoder to effectively capture this spatial diffusion pattern (Zhang et al., 2019). The spatial encoder is designed to learn spatial features both within and across channels in EEG and SEEG signals. This encoder is composed of multiple spatial convolution blocks, similar to the temporal encoder, but with a greater focus on capturing the spatial correlations between channels. We concatenate the features encoded in both time and frequency from EEG and SEEG signals ( $\varepsilon p = \varepsilon_t \odot \varepsilon_w$ ), and the output patch embedding of the spatial encoder can be expressed as:

$$\varepsilon_p = \left\{ e_{c_{ij},k}^p \in \mathbb{R}^d | j = 1, 2, \dots, C, k = 1, 2, \dots, \lfloor \frac{P}{W} \rfloor \right\}$$
(4)

where d is the dimension of the embeddings.

230

238 239

241

242

243

244

245 246 247

251

253

263 264

Temporal & Spatial Embedding. To enable the model to capture both temporal and spatial information of patch embeddings (Petukhova et al., 2024), we constructed two learnable embedding lists: temporal embeddings  $te = \{te_1, te_2, ..., te_{t_{max}}\}$  and spatial embeddings  $se = \{se_1, se_2, ..., se_C\}$ , both of which have a dimension of d. In this process, the hyperparameter  $t_{max}$  determines the maximum number of temporal patches, satisfying the condition  $t_{max} \ge N$ . Given arbitrary patch embeddings can be expressed as:

$$\varepsilon_{ineb} = \{ e^p_{c_{ij},k} + te_k + se_c | j = 1, \dots, C, k = 1, 2, \dots, \lfloor \frac{P}{W} \rfloor \}$$

$$(5)$$

240 where,  $\varepsilon_{ineb}$  is the input to the Transformer encoder.

**Transformer Encoder.** The sequence of embedding  $\varepsilon_{\text{ineb}}$  is input into the Transformer encoder (Vaswani, 2017), resulting in the output embedding  $e_i \in \mathbb{R}^d$ . To prevent excessively large values in the attention scores, we apply layer normalization to the queries and keys before the dot-product attention mechanism.

Attention
$$(Q, K, V) = \operatorname{softmax}\left(\frac{\operatorname{LN}(Q)\operatorname{LN}(K)^{T}}{\sqrt{d_{head}}}\right)V$$
 (6)

where, LN represents layer normalization, and  $d_{head}$  represents the dimension of each attention head. For downstream prediction tasks, the output embeddings are flattened and processed using classification or prediction heads.

### 2.3 NEURAL TOKENIZER TRAINING

**EpilepsyFM Neural Tokenizer.** As shown in Figure 2(a), we are inspired by LaBraM (Van 254 Den Oord et al., 2017; Jiang et al., 2024); however, they only trained the tokenizer on EEG sig-255 nals, which makes it incompatible with SEEG signals. To address the different signal sources 256 in epilepsy (such as intracranial and extracranial signals), we define an epilepsy neural tokenizer 257  $B = \{b_i | i = 1, ..., N_{\text{code}}\} \in \mathbb{R}^{N_{\text{code}} \times D_{\text{code}}}$ , where  $N_{\text{code}}$  represents the number of discrete epilepsy 258 neural embeddings and  $D_{\text{code}}$  represents the dimensionality of the embeddings. Given an epilepsy 259 EEG or SEEG signal, the tokenizer encodes it into patches  $q = \{q_i \in \mathbb{R}^{d_{code}} | i = 1, 2, \dots, N\}$ , 260 where N indicates the number of patches. Subsequently, the quantizer converts  $q_i$  into embeddings in 261 the epilepsy codebook. Finally, the nearest neighbor for each  $q_i$  is found using the epilepsy codebook 262 B (Peng et al., 2022), which can be expressed as:

$$z_i = \arg\min_{i} \|\ell_2(q_i) - \ell_2(b_j)\|^2$$
(7)

where,  $\ell_2$  is the  $\ell_2$ -normalization, and  $z_i$  is the vector after being quantized by the quantizer.

267 EpilepsyFM Neural Decoder. Due to the low signal-to-noise ratio of epilepsy EEG and SEEG
 268 signals, especially with EEG signals exhibiting strong randomness and nonstationarity, directly
 269 reconstructing these signals is highly challenging. Instead of reconstructing the raw signals, we
 predict the amplitude and phase from their Fourier spectrum, as these features more accurately reflect

270 brain activity. We apply the Discrete Fourier Transform (DFT) to EEG or SEEG segments to extract 271 amplitude and phase information (Gao et al., 2023), followed by normalization to ensure stable model 272 convergence. Then, using a epilepsy neural tokenizer and decoder, we train the model to predict these 273 features using the mean squared error loss function, with the total loss defined as follows:

277 278

279

280

281 282

283

284

 $\mathcal{L}_{T} = \sum_{\boldsymbol{x}\in\mathcal{D}}\sum_{i=1}^{N} \|o_{i}^{A} - A_{i}\|_{2}^{2} + \|o_{i}^{\phi} - \phi_{i}\|_{2}^{2} + \|\boldsymbol{sg}(\ell_{2}(q_{i})) - \ell_{2}(v_{b_{i}})\|_{2}^{2} + \|\ell_{2}(q_{i}) - \boldsymbol{sg}(\ell_{2}(b_{z_{i}}))\|_{2}^{2}$ (8)

where,  $o_i^A$  and  $o_i^{\phi}$  represent the predicted amplitude and phase from the epilepsy neural decoder, respectively, while  $A_i$  and  $\phi_i$  represent the true amplitude and phase obtained from the Fourier transform of the EEG or SEEG signals, and  $sq(\cdot)$  indicates the stop-gradient operation.

### 2.4 EPILEPSYFM PRE-TRAINING

**Masked EEG and SEEG Modeling.** As shown in Figure 2(b), to enhance the general representation learning of EpilepsyFM with extensive data, we propose a channel set masking strategy that 285 specifically considers the mechanisms of epileptic seizures

286 (Wei et al., 2022). Specifically, for the case where the num-287 ber of channels is divisible by 3, we apply random masking 288 to group the channels into sets of three, forming the mask 289  $M = \{m_i | i = 1, ..., N\}$ , where  $m_i \in \{0, 1\}$  and a propor-290 tion r of m is set to 1. For the remaining channels, we use 291 single-channel random masking, as shown in Figure 3. The 292 rationale for using 3 channels is detailed in Appendix F. We 293 replace the random mask M with a learnable mask  $e_M \in \mathbb{R}^d$ ; these masked patches are added to the temporal and spatial embeddings to serve as input to the Transformer encoder. Using 295 the output hidden vectors, we predict the neural tokens related 296 to EEG and SEEG. Ultimately, our objective training loss is 297 defined as follows: 298





Figure 3: Epileptic EEG and SEEG electrodes are masked according to channel sets.

Symmetric Masking. Inspired by LaBraM, we propose a symmetric masking strategy to enhance training efficiency. We utilize the inverse of the mask M as a new masking approach for pre-training, thereby obtaining the corresponding loss  $L_{sym}^M$ . Consequently, the total loss for our epilepsy base model is defined as follows:

$$L_{\rm mae} = L_M + L_{\rm sym}^M \tag{10}$$

306 307

299 300 301

302

303 304

305

308

3.1 DATASETS 310

3

**EXPERIMENTS** 

311 Due to the current lack of publicly available SEEG datasets related to epilepsy, we have collected 312 long-term clinical SEEG data from 20 patients in the neurosurgery department of a first-class hospital, 313 as well as extensive long-term EEG data, all acquired using Neuracle equipment (i.e., XJSZ-EEG 314 and XJSZ-SEEG). In these datasets, the EEG and SEEG signals are time-locked .

315 Pre-training Datasets: The pre-training dataset includes the largest publicly available epilepsy-316 related TUH EEG datasets, such as TUEP, TUSL, and TUSZ (Harati et al., 2014), as well as data from 317 4 patients for XJSZ-EEG and 12 patients for XJSZ-SEEG that we collected. To cover all training 318 data and improve training efficiency, we set the time length of each sample to 4 seconds.

319 Downstream Datasets: The downstream task dataset includes TUAB, TUEV, (Harati et al., 2014) 320 CHB-MIT (Shoeb, 2009), and data from 8 patients with XJSZ-EEG and XJSZ-SEEG. The tasks 321 across these datasets vary. TUAB is segmented into 10-second samples, TUEV into 5-second samples, 322 and CHB-MIT, XJSZ-EEG, and XJSZ-SEEG into 2-second samples. 323

For more detailed information about the data, please refer to Appendix B.

# 324 3.2 IMPLEMENTATION DETAILS

Preprocessing: We downsampled the SEEG signals from 1000Hz to 200Hz and similarly downsampled the EEG signals from various datasets to 200Hz to retain essential information. Meanwhile, we applied a 0.5-70Hz bandpass filter to the EEG and SEEG signals to remove noise, along with a 50Hz notch filter to eliminate power line interference. Finally, z-score normalization is used to reduce nonstationarity and fluctuations in the signals (Hosseini et al., 2020).

Model Configuration: Each EEG and SEEG patch has a length of 1 second, with the number of
 patches limited to 256. The temporal and spatial encoders consist of 3 layers of 1-D convolution,
 GroupNorm, and GELU, while the frequency encoder is composed of 3 layers of wavelet packet
 convolution, GroupNorm, and GELU, with the wavelet packet convolution used to extract frequency
 band features from the signals. Subsequently, the EEG or SEEG patches are transformed into patch
 embeddings and fed into a Transformer encoder with a depth of 8 layers and 10 attention heads. For
 more details, please refer to Appendix C.

338 Pre-training and Fine-tuning: To train the epilepsy neural tokenizer and EpilepsyFM pre-training, we utilized the largest publicly available EEG epilepsy TUH series dataset, along with most of the 339 340 XJSZ-EEG and XJSZ-SEEG datasets as training data. This dataset includes approximately 9,000 patients and 2,000 hours of data, with specific details provided in Appendix B. The task records 341 vary across different datasets. We used TUAB, TUEV, CHB-MIT, and a small portion of XJSZ-EEG 342 and XJSZ-SEEG data from different patients to evaluate downstream epilepsy tasks. The training, 343 validation, and testing sets are divided in a ratio of 80%, 10%, and 10%. Our experiments are 344 conducted on 8 A100-PCIE-40GB GPUs using Python 3.8 (Ubuntu 20.04) and PyTorch 1.11.0 + 345 CUDA 11.3. During training, binary classification tasks employed binary cross-entropy loss, while 346 multi-class classification tasks used cross-entropy loss. 347

348 349 350

351

357

358

359

360

361 362

364

365

366

367 368

369

370

371 372 373

374

3.3 EVALUATION RESULTS

3.4 MAIN RESULTS

The performance of EpilepsyFM across all downstream tasks is presented in Figure 4. As a foundational model in the field of epilepsy, EpilepsyFM demonstrates SOTA (state-of-the-art) performance in both extracranial EEG and intracranial SEEG detection and prediction tasks, compared to other baseline models. In the following paragraphs, we will provide a detailed discussion of the experimental results for each task.



Figure 4: Performance comparison of downstream tasks. **Left:** Epilepsy detection task (Metrics: Balance Accuracy); **Center:** Epilepsy short-term prediction task (Metrics: Mean Absolute Error); **Right:** Epilepsy long-term prediction task (Metrics: Mean Absolute Error).

### 3.5 SEIZURE DETECTION

Epileptic seizure detection is of significant importance in the clinical application of EEG and SEEG (Siddiqui et al., 2020; Tran et al., 2022; Farooq et al., 2023). The aim of seizure detection is to identify or classify the discharge phenomena and types of discharges in the intracranial SEEG and extracranial EEG signals of patients with epilepsy. We present the seizure detection performance of

scalp EEG and intracranial SEEG on different datasets. We use Balance Accuracy, AUC-PR, and
 AUROC as evaluation metrics, with the definitions of these metrics provided in Appendix D.

Seizure Detection on Extracranial EEG: We present the detection performance of EpilepsyFM against other advanced methods on various public epilepsy EEG datasets (TUAB, TUEV, CHB-MIT) and a private EEG dataset (XJSZ-EEG). We compared not only self-supervised pre-training methods (TF-C, SimMTM) (Zhang et al., 2022; Dong et al., 2024) and temporal mask modeling methods (BrainBERT, LaBraM) (Wang et al., 2023; Jiang et al., 2024) as unsupervised approaches but also a supervised learning method specifically for EEG, namely ST-Transformer (Song et al., 2021).

As shown in Table 1, EpilepsyFM outperformed other methods across multiple evaluation metrics, with LaBraM and BrainBERT following closely behind. This is largely attributed to LaBraM and BrainBERT's ability to effectively consider contextual information and integrate embeddings of neural signals related to brain activity, particularly as BrainBERT shares a similar approach to our model in integrating temporal and spectral information.

Table 1: The performance(%) of epilepsy detection tasks on EEG.

Metrics		TUAB			CHB-MIT	
Methods	Balance Acc.	AUC-PR	AUROC	Balance Acc.	AUC-PR	AUROC
TF-C	59.98±0.47	58.08±0.44	57.45±0.68	80.68±0.66	62.74±0.50	58.65±0.88
SimMTM	58.68±0.40	59.33±0.44	$62.15 \pm 0.72$	75.74±0.63	58.02±0.81	81.38±0.92
ST-Transformer	79.66±0.23	85.21±0.26	$87.07 \pm 0.19$	82.45±1.27	60.32±1.46	86.28±0.68
BrainBERT	78.97±0.75	85.14±0.54	87.33±0.71	82.31±0.72	60.45±0.45	86.36±0.72
LaBraM	81.37±0.21	89.33±0.24	89.72±0.16	84.51±0.05	62.45±0.73	90.47±0.14
EpilepsyFM	82.27±0.18	89.77±0.17	91.91±0.11	85.12±0.34	64.25±0.45	92.01±0.52
Metrics		XJSZ-EEG			TUEV	
Metrics Methods	Balance Acc.	XJSZ-EEG AUC-PR	AUROC	Balance Acc.	TUEV Cohen Kappa	Weighted F1
Metrics Methods TF-C	Balance Acc.	XJSZ-EEG AUC-PR 69.35±0.34	AUROC 73.35±0.86	Balance Acc.	TUEV Cohen Kappa 37.19±0.12	Weighted F1 67.19±0.12
Metrics Methods TF-C SimMTM	Balance Acc. 68.54±0.43 74.67±0.87	XJSZ-EEG AUC-PR 69.35±0.34 75.34±0.56	AUROC 73.35±0.86 75.85±0.45	Balance Acc. 40.46±0.67 40.12±0.96	TUEV Cohen Kappa 37.19±0.12 37.94±0.73	Weighted F1 67.19±0.12 37.94±0.73
Metrics Methods TF-C SimMTM ST-Transformer	Balance Acc. 68.54±0.43 74.67±0.87 76.27±1.64	XJSZ-EEG AUC-PR 69.35±0.34 75.34±0.56 76.79±1.58	AUROC 73.35±0.86 75.85±0.45 79.36±0.99	Balance Acc. 40.46±0.67 40.12±0.96 39.84±2.28	TUEV Cohen Kappa 37.19±0.12 37.94±0.73 37.65±3.06	Weighted F1 67.19±0.12 37.94±0.73 68.23±1.90
Metrics Methods TF-C SimMTM ST-Transformer BrainBERT	Balance Acc. 68.54±0.43 74.67±0.87 76.27±1.64 75.90±0.36	XJSZ-EEG AUC-PR 69.35±0.34 75.34±0.56 76.79±1.58 76.45±0.53	AUROC 73.35±0.86 75.85±0.45 79.36±0.99 78.34±0.57	Balance Acc. 40.46±0.67 40.12±0.96 39.84±2.28 47.55±0.75	TUEV Cohen Kappa 37.19±0.12 37.94±0.73 37.65±3.06 48.76±0.58	Weighted F1 67.19±0.12 37.94±0.73 68.23±1.90 60.64±0.56
Metrics Methods TF-C SimMTM ST-Transformer BrainBERT LaBraM	Balance Acc. 68.54±0.43 74.67±0.87 76.27±1.64 75.90±0.36 76.74±0.45	XJSZ-EEG AUC-PR 69.35±0.34 75.34±0.56 76.79±1.58 76.45±0.53 77.21±0.84	AUROC 73.35±0.86 75.85±0.45 79.36±0.99 78.34±0.57 79.98±0.56	Balance Acc. 40.46±0.67 40.12±0.96 39.84±2.28 47.55±0.75 61.26±0.48	TUEV Cohen Kappa 37.19±0.12 37.94±0.73 37.65±3.06 48.76±0.58 61.76±1.21	Weighted F1 67.19±0.12 37.94±0.73 68.23±1.90 60.64±0.56 80.46±0.45

410 411

387

388

389

390

391

392

396 397

412 413

414

**Seizure Detection on Intracranial SEEG:** The detection results for SEEG signals are presented in Table 2. This study found that the detection performance of SEEG signals significantly outperforms that of EEG signals, primarily due to the high signal-to-noise ratio of SEEG signals. This abarrateristic analysis our EpilepsyEM

This characteristic enables our EpilepsyFM 415 model to excel across multiple evalua-416 tion metrics. Additionally, the design of 417 the EpilepsyFM model fully considers the 418 unique properties of SEEG signals, result-419 ing in enhanced robustness across different 420 intracranial channel locations. These find-421 ings underscore the critical role of Epilep-422 syFM in SEEG analysis and further support 423 its effectiveness in clinical applications.

Ultimately, our detection results outperform other methods due to our model's design, which specifically addresses the characteristics of seizure onset and propagation in epilepsy. By utilizing channel Table 2: The performance(%) of epilepsy detection tasks on SEEG.

Metrics	Σ	KJSZ-SEEG	
Methods	Balance Acc.	AUC-PR	AUROC
TF-C	78.10±0.50	76.50±0.60	80.04±0.55
SimMTM	80.25±0.65	78.80±0.70	$82.85 \pm 0.60$
ST-Transformer	83.50±0.70	81.56±0.50	$85.34 \pm 0.65$
BrainBERT	86.30±0.60	84.50±0.75	89.00±0.70
LaBraM	87.50±0.75	83.67±0.61	85.46±0.55
EpilepsyFM	89.74±0.74	86.05±0.57	92.07±0.56

set masking and spatio-temporal frequency encoders, we achieve a broader receptive field. This
enables the model to effectively capture complex patterns and relationships within the EEG signals,
thereby improving the accuracy of epilepsy detection (further experimental evidence can be found in Appendix G).

#### 432 3.6 EPILEPSY SHORT- AND LONG-TERM SIGNAL PREDICTION 433

434 Predicting epilepsy signals is essential for developing early warning systems for patients in need of preventive measures (Usman et al., 2020; Savadkoohi et al., 2020; Dissanayake et al., 2021). We 435 established both short-term and long-term predictions for seizures, focusing on signals of varying 436 durations. Specifically, we set the historical sequence length to 90 epochs (90 seconds), with short-437 term predictions covering 12 epochs (12 seconds) and long-term predictions spanning 20 epochs 438 (20 seconds). A linear prediction head is utilized to forecast future signals, and we employed Mean 439 Absolute Error (MAE) and Mean Squared Error (MSE) as evaluation metrics, with the definitions of 440 these metrics provided in Appendix D. The prediction results are shown in Table 3. 441

Short- and Long-term Signal Prediction on Extracrania EEG: EpilepsyFM has demonstrated 442 outstanding performance in short-term prediction across multiple EEG datasets. For example, in the 443 XJSZ-EEG dataset, the MAE is 0.6421 and the MSE is 0.9380, which is significantly better than 444 other baseline methods. This superiority is attributed to the model's multi-dimensional design, which 445 enables it to analyze signals across various dimensions, allowing for flexible capture of dynamic 446 changes and long-range dependencies in the signals. In long-term prediction, EpilepsyFM also 447 exhibits strong performance, with a MAE of 0.6567 and an MSE of 1.1123 in the XJSZ-EEG dataset, 448 further validating its potential and reliability in epileptic signal prediction. 449

Short- and Long-term Signal Prediction on Intracrania SEEG: The EpilepsyFM method has 450 again demonstrated its outstanding capabilities in short-term prediction, achieving a MAE of 0.5587 451 and an MSE of 0.8592 in the XJSZ-SEEG dataset, highlighting its effectiveness in capturing key 452 features in SEEG signal processing. In contrast, the LaBraM and SimMTM methods recorded MAE 453 values of 0.7556 and 0.9621, respectively. While these methods performed well, they did not match 454 the accuracy of EpilepsyFM. In long-term signal prediction, EpilepsyFM continued to maintain its 455 advantage, with a MAE of 0.6024 and an MSE of 0.9502, further emphasizing its adaptability and 456 accuracy across different signal processing scenarios. 457

	1	1 1 2	υ	υ	
Dataset	Metrics	Short-term Sig	gnal Prediction	Long-term Sig	nal Prediction
2 4440 00	Methods	MAE	MSE	MAE	MSE
	TF-C	1.2413±0.0013	1.8411±0.0037	$1.4242 \pm 0.0012$	$1.9720 \pm 0.0021$
	SimMTM	1.5213±0.0015	2.2561±0.0063	1.6892±0.0010	2.4720±0.0023
TUAB	LaBraM	0.1907±0.0025	0.3136±0.0081	$0.2237 \pm 0.0030$	0.3321±0.0013
	EpilepsyFM	0.1498±0.0010	0.3868±0.0076	0.1482±0.0026	0.2868±0.0016
	TF-C	1.3251±0.0011	$2.2802 \pm 0.0053$	$1.4436 \pm 0.0017$	$2.7676 \pm 0.0011$
	SimMTM	$1.3765 \pm 0.0014$	2.3215±0.0062	$1.4212 \pm 0.0012$	2.8675±0.0010
CHB-MIT	LaBraM	0.9823±0.0013	1.4657±0.0076	0.9701±0.0014	$1.5246 \pm 0.0018$
	EpilepsyFM	0.8231±0.0014	$1.2002 \pm 0.0048$	$0.8550 \pm 0.0012$	1.2319±0.0016
	TF-C	1.4075±0.0008	1.6802±0.0110	1.4436±0.0016	1.6789±0.0031
	SimMTM	1.4356±0.0024	1.7721±0.0143	$1.3665 \pm 0.0033$	1.6213±0.0030
TUEV	LaBraM	0.4112±0.0023	$1.6092 \pm 0.0064$	$0.4231 \pm 0.0012$	$1.6294 \pm 0.0013$
	EpilepsyFM	0.3489±0.0046	1.5272±0.0044	$0.3489 \pm 0.0014$	$1.5257 \pm 0.0043$
	TF-C	0.8265±0.0014	$1.2588 \pm 0.0022$	$0.8541 \pm 0.0015$	1.3187±0.0025
	SimMTM	0.8531±0.0012	$1.3065 \pm 0.0025$	0.9779±0.0016	$1.4231 \pm 0.0030$
XJSZ-EEG	LaBraM	0.7231±0.0010	$1.1214 \pm 0.0015$	$0.7583 \pm 0.0015$	$1.2780 \pm 0.0021$
	EpilepsyFM	0.6421±0.0010	0.9380±0.0010	$0.6567 \pm 0.0012$	1.1123±0.0015
	TF-C	0.8513±0.0011	1.3034±0.0023	0.9332±0.0014	1.4015±0.0026
	SimMTM	0.9621±0.0015	$1.4295 \pm 0.0025$	$0.9577 \pm 0.0018$	$1.5389 \pm 0.0028$
XJSZ-SEEG	LaBraM	0.7556±0.0012	$1.2674 \pm 0.0020$	$0.8206 \pm 0.0015$	$1.3423 \pm 0.0023$
	EpilepsyFM	0.5587±0.0010	0.8592±0.0012	0.6024±0.0014	0.9502±0.0018

Table 3: The performance of epilepsy short- and long-term Signal Prediction.

481 482 483

458

In summary, the EpilepsyFM method outperformed other methods in both short-term and long-term 484 prediction tasks on the EEG and SEEG datasets, exhibiting significant accuracy advantages. This 485 result underscores the potential of EpilepsyFM in predicting epilepsy signals.

489

490

491

492

493

494

523

528

# 486 4 EPILEPSYLM WITH/WITHOUT PRE-TRAINING

In the experiments, we compared the impact of pre-training on the detection and prediction performance across downstream datasets. For the without pre-training experiments, we trained the EpilespyFM model from scratch on each downstream task dataset, and the results are shown in Figure 5. Here, we only present the performance of publicly available datasets in the downstream tasks. The experimental results indicate a decline in performance on the TUAB, CHB-MIT, and TUEV datasets, with a particularly noticeable drop on the TUEV dataset, thereby validating the effectiveness of channel set mask pre-training.



Figure 5: Performance comparison with/without pre-training. Left: Epilepsy detection task; Right: Epilepsy short- and long-term prediction tasks

### 5 CONCLUSION

529 Inspired by LLMs, we designed a foundational model—EpilepsyFM, specifically for understanding 530 the mechanisms of epilepsy onset and propagation. This model is capable of learning generalized 531 representations from extracranial EEG and intracranial SEEG signals across different signal types, 532 data formats and sources to facilitate various epilepsy-related tasks. To enhance data diversity and 533 address the scarcity of intracranial SEEG signals, we collected time-locked clinical EEG and SEEG 534 data from a first-class hospital's neurosurgery department. Initially, we trained an epilepsy neural 535 tokenizer to learn comprehensive EEG and SEEG data, thereby creating a codebook for the epilepsy 536 neural tokenizer. We then employed unsupervised pre-training using a channel set masking approach 537 to capture generalized representations of epilepsy. Subsequently, we fine-tuned the model on multiple downstream tasks related to epilepsy, achieving state-of-the-art performance across different task 538 settings. Extensive experiments demonstrated that EpilepsyFM achieved leading performance in tasks such as seizure detection, and short- and long-term neural signal prediction.

# 540 REFERENCES

558

565

- Md Adnan Arefeen, Biplob Debnath, and Srimat Chakradhar. Leancontext: Cost-efficient domain specific question answering using llms. *Natural Language Processing Journal*, 7:100065, 2024.
- Ali A Asadi-Pooya, Francesco Brigo, Simona Lattanzi, and Ingmar Blumcke. Adult epilepsy. *The Lancet*, 402(10399):412–424, 2023.
- Hangbo Bao, Li Dong, Songhao Piao, and Furu Wei. Beit: Bert pre-training of image transformers.
   *arXiv preprint arXiv:2106.08254*, 2021.
- <sup>549</sup> Donghong Cai, Junru Chen, Yang Yang, Teng Liu, and Yafeng Li. Mbrain: A multi-channel self-supervised learning framework for brain signals. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, August 2023. doi: 10.1145/3580305.
  <sup>552</sup> 3599426. URL http://dx.doi.org/10.1145/3580305.3599426.
- Junru Chen, Yang Yang, Tao Yu, Yingying Fan, Xiaolong Mo, and Carl Yang. Brainnet: Epileptic wave detection from seeg with hierarchical graph diffusion learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2741–2751. ACM, August 2022. doi: 10.1145/3534678.3539178. URL http://dx.doi.org/10.1145/3534678.3539178.
- Wenhui Cui, Woojae Jeong, Philipp Thölke, Takfarinas Medani, Karim Jerbi, Anand A Joshi, and
  Richard M Leahy. Neuro-gpt: Towards a foundation model for eeg. In 2024 IEEE International
  Symposium on Biomedical Imaging (ISBI), pp. 1–5. IEEE, 2024.
- Theekshana Dissanayake, Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes.
   Deep learning for patient-independent epileptic seizure prediction using scalp eeg signals. *IEEE* Sensors Journal, 21(7):9377–9388, 2021.
- Jiaxiang Dong, Haixu Wu, Haoran Zhang, Li Zhang, Jianmin Wang, and Mingsheng Long. Simmtm:
   A simple pre-training framework for masked time-series modeling. Advances in Neural Information Processing Systems, 36, 2024.
- Muhammad Shoaib Farooq, Aimen Zulfiqar, and Shamyla Riaz. Epileptic seizure detection using
   machine learning: Taxonomy, opportunities, and challenges. *Diagnostics*, 13(6):1058, 2023.
- Qiang Gao, Alaa Hamza Omran, Yasamin Baghersad, Omid Mohammadi, Mohammed Ayad Alkhafaji, Abdul Kareem J Al-Azzawi, Sara Hakem Al-Khafaji, Nafiseh Emami, Davood Toghraie, and Mohammad Javad Golkar. Electroencephalogram signal classification based on fourier transform and pattern recognition network for epilepsy diagnosis. *Engineering Applications of Artificial Intelligence*, 123:106479, 2023.
- Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare*, 3(1):1–23, October 2021. ISSN 2637-8051. doi: 10.1145/3458754. URL http://dx.doi.org/10.1145/3458754.
- A. Harati, S. López, I. Obeid, J. Picone, M. P. Jacobson, and S. Tobochnik. The tuh eeg corpus: A
   big data resource for automated eeg interpretation. In 2014 IEEE Signal Processing in Medicine
   and Biology Symposium (SPMB), pp. 1–5, 2014.
- Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415, 2016.
- 588 Mohammad-Parsa Hosseini, Amin Hosseini, and Kiarash Ahi. A review on machine learning for eeg signal processing in bioengineering. *IEEE reviews in biomedical engineering*, 14:204–218, 2020.
- Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci. *arXiv preprint arXiv:2405.18765*, 2024.
- <sup>593</sup> Valer Jurcak, Daisuke Tsuzuki, and Ippeita Dan. 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage*, 34(4):1600–1611, 2007.

594 595 596	Andres M Kanner and Manuel Melo Bicchi. Antiseizure medications for adults with epilepsy: a review. <i>Jama</i> , 327(13):1269–1281, 2022.
597 598	Binbin Li, Tianxin Meng, Xiaoming Shi, Jie Zhai, and Tong Ruan. Meddm: Llm-executable clinical guidance tree for clinical decision-making. <i>arXiv preprint arXiv:2312.02441</i> , 2023.
599 600 601 602	Yang Li, Yu Liu, Wei-Gang Cui, Yu-Zhu Guo, Hui Huang, and Zhong-Yi Hu. Epileptic seizure detection in eeg signals using a unified temporal-spectral squeeze-and-excitation network. <i>IEEE Transactions on Neural Systems and Rehabilitation Engineering</i> , 28(4):782–794, 2020.
603 604	Pejman Memar and Farhad Faradji. A novel multi-class eeg-based sleep stage classification system. <i>IEEE Transactions on Neural Systems and Rehabilitation Engineering</i> , 26(1):84–95, 2017.
605 606 607 608 609	Saramati Narasimhan, Keshav B Kundassery, Kanupriya Gupta, Graham W Johnson, Kristin E Wills, Sarah E Goodale, Kevin Haas, John D Rolston, Robert P Naftel, Victoria L Morgan, et al. Seizure-onset regions demonstrate high inward directed connectivity during resting-state: An seeg study in focal epilepsy. <i>Epilepsia</i> , 61(11):2534–2544, 2020.
610 611	Iyad Obeid and Joseph Picone. The temple university hospital eeg data corpus. <i>Frontiers in neuroscience</i> , 10:196, 2016.
612 613 614 615	Soumen Pal, Manojit Bhattacharya, Sang-Soo Lee, and Chiranjib Chakraborty. A domain-specific next-generation large language model (llm) or chatgpt is required for biomedical engineering and research. <i>Annals of biomedical engineering</i> , 52(3):451–454, 2024.
616 617 618	Wei Yan Peh, Prasanth Thangavel, Yuanyuan Yao, John Thomas, Yee-Leng Tan, and Justin Dauwels. Six-center assessment of cnn-transformer with belief matching loss for patient-independent seizure detection in eeg. <i>International Journal of Neural Systems</i> , 33(03):2350012, 2023.
619 620 621	Zhiliang Peng, Li Dong, Hangbo Bao, Qixiang Ye, and Furu Wei. Beit v2: Masked image modeling with vector-quantized visual tokenizers. <i>arXiv preprint arXiv:2208.06366</i> , 2022.
622 623	Alina Petukhova, Joao P Matos-Carvalho, and Nuno Fachada. Text clustering with llm embeddings. <i>arXiv preprint arXiv:2403.15112</i> , 2024.
625 626 627	Khansa Rasheed, Junaid Qadir, Terence J. O'Brien, Levin Kuhlmann, and Adeel Razi. A generative model to synthesize eeg data for epileptic seizure prediction. <i>IEEE Transactions on Neural Systems and Rehabilitation Engineering</i> , 29:2322–2332, 2021.
628 629 630	Marzieh Savadkoohi, Timothy Oladunni, and Lara Thompson. A machine learning approach to epilep- tic seizure prediction using electroencephalogram (eeg) signal. <i>Biocybernetics and Biomedical</i> <i>Engineering</i> , 40(3):1328–1341, 2020.
631 632 633 634	Vinit Shah, Eva Von Weltin, Silvia Lopez, James Riley McHugh, Lillian Veloso, Meysam Golmo- hammadi, Iyad Obeid, and Joseph Picone. The temple university hospital seizure detection corpus. <i>Frontiers in neuroinformatics</i> , 12:83, 2018.
635 636 637 638	Jun Shi, Xiaoping Liu, Wei Xiang, Mo Han, and Qinyu Zhang. Novel fractional wavelet packet transform: theory, implementation, and applications. <i>IEEE Transactions on Signal Processing</i> , 68: 4041–4054, 2020.
639 640	Ali Hossam Shoeb. <i>Application of machine learning to epileptic seizure onset detection and treatment</i> . PhD thesis, Massachusetts Institute of Technology, 2009.
641 642 643 644 645	Afshin Shoeibi, Marjane Khodatars, Navid Ghassemi, Mahboobeh Jafari, Parisa Moridian, Roohallah Alizadehsani, Maryam Panahiazar, Fahime Khozeimeh, Assef Zare, Hossein Hosseini-Nejad, et al. Epileptic seizures detection using deep learning techniques: A review. <i>Int. J. Environ. Res. Public Health</i> , 18(11):5780, 2021.
646 647	Mohammad Khubeb Siddiqui, Ruben Morales-Menendez, Xiaodi Huang, and Nasir Hussain. A review of epileptic seizure detection using machine learning classifiers. <i>Brain Informatics</i> , 7(1): 1–18, 2020.

667

668

669

670

678

- Yonghao Song, Xueyu Jia, Lie Yang, and Longhan Xie. Transformer-based spatial-temporal feature learning for eeg decoding. *arXiv preprint arXiv:2106.11170*, 2021.
- Nicola Specchio, Elaine C Wirrell, Ingrid E Scheffer, Rima Nabbout, Kate Riney, Pauline Samia, Marilisa Guerreiro, Sam Gwer, Sameer M Zuberi, Jo M Wilmshurst, et al. International league against epilepsy classification and definition of epilepsy syndromes with onset in childhood: Position paper by the ilae task force on nosology and definitions. *Epilepsia*, 63(6):1398–1442, 2022.
- Supriya Supriya, Siuly Siuly, Hua Wang, and Yanchun Zhang. Epilepsy detection from eeg using
   complex network techniques: A review. *IEEE Reviews in Biomedical Engineering*, 16:292–306,
   2021.
- Amara Tariq, Man Luo, Aisha Urooj, Avisha Das, Jiwoong Jeong, Shubham Trivedi, Bhavik Patel, and Imon Banerjee. Domain-specific llm development and evaluation–a case-study for prostate cancer. *medRxiv*, pp. 2024–03, 2024.
- Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang
   Tan, and Daniel Shu Wei Ting. Large language models in medicine. *Nature medicine*, 29(8):
   1930–1940, 2023.
  - Ly V Tran, Hieu M Tran, Tuan M Le, Tri TM Huynh, Hung T Tran, and Son VT Dao. Application of machine learning in epileptic seizure detection. *Diagnostics*, 12(11):2879, 2022.
  - Syed Muhammad Usman, Shehzad Khalid, and Muhammad Haseeb Aslam. Epileptic seizures prediction using deep learning techniques. *Ieee Access*, 8:39998–40007, 2020.
- 671 Syed Muhammad Usman, Shehzad Khalid, and Sadaf Bashir. A deep learning based ensemble
  672 learning method for epileptic seizure prediction. *Computers in Biology and Medicine*, 136:104710, 2021.
- Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. Advances in neural information processing systems, 30, 2017.
- 677 A Vaswani. Attention is all you need. Advances in Neural Information Processing Systems, 2017.
- L Veloso, J McHugh, E Von Weltin, S Lopez, I Obeid, and J Picone. Big data resources for eegs: Enabling deep learning research. In 2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), pp. 1–3. IEEE, 2017.
- Eva von Weltin, Tameem Ahsan, Vinit Shah, Dawer Jamshed, Meysam Golmohammadi, Iyad Obeid, and Joseph Picone. Electroencephalographic slowing: A primary source of error in automatic seizure detection. In 2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), pp. 1–5. IEEE, 2017.
- <sup>686</sup> Christopher Wang, Vighnesh Subramaniam, Adam Uri Yaari, Gabriel Kreiman, Boris Katz, Ignacio
   <sup>687</sup> Cases, and Andrei Barbu. Brainbert: Self-supervised representation learning for intracranial
   <sup>688</sup> recordings. *arXiv preprint arXiv:2302.14367*, 2023.
- <sup>690</sup> Dong Wang, Zhian Liu, Yi Tao, Wenjing Chen, Badong Chen, Qiang Wang, Xiangguo Yan, and
   <sup>691</sup> Gang Wang. Improvement in eeg source imaging accuracy by means of wavelet packet transform
   <sup>692</sup> and subspace component selection. *IEEE Transactions on Neural Systems and Rehabilitation* <sup>693</sup> Engineering, 29:650–661, 2021.
- Chen Wei, Haoqi Fan, Saining Xie, Chao-Yuan Wu, Alan Yuille, and Christoph Feichtenhofer.
   Masked feature prediction for self-supervised visual pre-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14668–14678, 2022.
- Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting, 2022. URL https://arxiv.org/abs/2202.01575.
- 701 Yong Xie, Karan Aggarwal, and Aitzaz Ahmad. Efficient continual pre-training for building domain specific large language models, 2023. URL https://arxiv.org/abs/2311.08545.

702	Zhaofen Yan, Yujiao Yang, Jing Wang, Oin Deng, Liping Zhang, Minghui Wang, Jian Zhou, YuGuang
703	Guan Guoming Luan and Mengyang Wang. Posterior cingulate enlensy: Seizure semiology and
704	intercarial electrical stimulation using spage Sairura: Europage Journal of Enilossy 110-28 35
705	and a contrained electrical stimulation using seeg. Seizure. European Journal of Epitepsy, 119.26–55,
CU1	2024.

- Ke Yi, Yansen Wang, Kan Ren, and Dongsheng Li. Learning topology-agnostic eeg representations with geometry-aware modeling. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 53875–53891. Curran Associates, Inc., 2023.
- Dong Yuan, Eti Rastogi, Gautam Naik, Jai Chintagunta, Sree Prasanna Rajagopal, Fen Zhao, Sagar
   Goyal, and Jeff Ward. A continued pretrained llm approach for automatic medical note generation.
   *arXiv preprint arXiv:2403.09057*, 2024a.
- Zhizhang Yuan, Daoze Zhang, Junru Chen, Geifei Gu, and Yang Yang. Brant-2: Foundation model for brain signals. *arXiv preprint arXiv:2402.10251*, 2024b.
- Daoze Zhang, Zhizhang Yuan, Yang Yang, Junru Chen, Jingjing Wang, and Yafeng Li. Brant: Foun dation model for intracranial neural signal. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- Wenxuan Zhang, Hongzhi Liu, Zhijin Dong, Yingpeng Du, Chen Zhu, Yang Song, Hengshu Zhu, and Zhonghai Wu. Bridging the information gap between domain-specific model and general llm for personalized recommendation. In Wenjie Zhang, Anthony Tung, Zhonglong Zheng, Zhengyi Yang, Xiaoyang Wang, and Hongjie Guo (eds.), *Web and Big Data*, pp. 280–294, Singapore, 2024. Springer Nature Singapore.
- Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency. In *Proceedings of Neural Information Processing Systems, NeurIPS*, 2022.
- Yuan Zhang, Yao Guo, Po Yang, Wei Chen, and Benny Lo. Epilepsy seizure prediction on eeg using
   common spatial pattern and convolutional neural network. *IEEE Journal of Biomedical and Health Informatics*, 24(2):465–474, 2019.
  - Xuyang Zhao, Qibin Zhao, and Toshihisa Tanaka. Epilepsyllm: Domain-specific large language model fine-tuned with epilepsy medical knowledge. *arXiv preprint arXiv:2401.05908*, 2024.

#### 756 **RELATED WORK** А

758 Self-supervised Pre-training: Self-supervised pre-training has made significant progress in natural 759 language processing and computer vision, yet its potential in the brain-computer interface (BCI) 760 field remains underexplored, providing an opportunity to leverage large-scale EEG data for training 761 large models. Currently, several foundational models have emerged: Neuro-GPT (Cui et al., 2024) 762 aims to address the scarcity and heterogeneity of EEG data by combining EEG encoders with GPT models; MMM (Yi et al., 2023) facilitates cross-dataset brainwave pre-training through a unified 764 topological structure, employing multidimensional position encoding and a multi-level channel hierarchy; LaBraM (Jiang et al., 2024) segments raw EEG signals into slices and trains a neural 765 tokenizer via spectral prediction to generate neural vocabulary, masking portions of slices during 766 pre-training for the Neural Transformer to predict from visible slices. The Brant (Zhang et al., 767 2023) and Brant-2 (Yuan et al., 2024b) models are specifically designed for intracranial recordings, 768 with Brant-2 expanding upon its predecessor to demonstrate robustness against data variations and 769 modeling scales, applicable to a broader range of neural data and validating the effectiveness and 770 application potential of self-supervised learning in the BCI domain. The introduction of these models 771 marks a significant step toward utilizing self-supervised learning and large-scale datasets in BCI.

772 **Domain-specific Foundation Model:** Existing brain signal analysis models are primarily general-773 purpose, excelling in various applications but lacking focus on epilepsy analysis. For instance, 774 domain-specific models have been successfully utilized in other medical fields, such as Tariq et al.'s 775 work on prostate cancer, where they collected over 1.8 million clinical notes and developed a model 776 fine-tuned for clinical information prediction and question-answering tasks. Their model, enhanced 777 by UMLS-guided training and domain-specific tokenizers, significantly outperformed the general-778 purpose model GPT-2 and even surpassed the specialized BioGPT in certain tasks (Tarig et al., 2024). 779 However, similar methodologies have not been widely applied to epilepsy analysis. While there have been efforts like Zhao et al.'s EpilepsyLLM, which fine-tuned models on epilepsy-related datasets, this work focused mainly on textual knowledge representation (Zhao et al., 2024). In contrast, our 781 research advances a model specifically tailored for epilepsy, leveraging extensive specialized data 782 to analyze seizure dynamics and addressing clinical needs in EEG analysis. Our work not only fills 783 a gap in LLM field but also provides insights for applying large-scale EEG data to specific disease 784 research, further advancing epilepsy studies. 785

786 787

788

789

791

792

793

794

797

798

799

800

801

802

803

804

805

### В DATASET DESCRIPTION

- **TUEP** (Veloso et al., 2017): This corpus is a subset of the TUH EEG Corpus that contains 100 sessions from patients with epilepsy and 100 sessions from patients without epilepsy with EEG recorded (19-23 channels, 256 Hz). These data are based on clinical history, EEG features, etc., reviewed and verified by a board-certified neurologist. (total time: 473 hours)
- TUSL (von Weltin et al., 2017): This is another subset of TUH EEG that contains sessions that are known to contain seizure events, slowing events, and complex background events (23 channels, 256 Hz). The corpus has been used in studies to distinguish between seizures and slowing events. (total time: 20.59 hours)
  - TUSZ (Shah et al., 2018): This corpus contains EEG signals that are annotated to contain in addition to a start and stop time, a channel index with EEG recorded (19-23 channels, 256 Hz). (total time: 1138.53 hours)
- TUAB (Obeid & Picone, 2016): The corpus contains statistical data (about demographics and related information) from the TUH EEG anomaly Corpus, which contains EEG recordings is classified as clinically normal or abnormal. (total time: 1142 hours)
- TUEV (Obeid & Picone, 2016): This is a subset of TUH EEG that contains annotations of EEG segments as one of six classes: (1) spike and sharp wave (SPSW), (2) generalized periodic epileptiform discharges (GPED), (3) periodic lateralized epileptiform discharges (PLED), (4) eye movement (EYEM), (5) artifact (ARTF) and (6) background (BCKG).
- CHB-MIT (Shoeb, 2009): This dataset consists of scalp EEG signals from 23 pediatric epilepsy patients, including 24 EEG sets, each subset corresponding to several days of 808 records from one epilepsy patient, and each epileptic EEG set contains 9 to 42 consecutive 809 sets of multilead epileptic EEGs (256 Hz) (total time: 969hours)

 • **XJSZ-EEG:** The dataset is collected from the Department of Neurosurgery at a first-class hospital, monitoring the EEG signals of 12 clinical epilepsy patients. Each patient have different electrode positions and channel numbers. We selected the data from 4 patients for self-supervised pre-training and used the data from 8 patients for downstream tasks. The labels are manually annotated by two clinically experienced physicians. The sampling frequency is 1000 Hz, synchronized with SEEG signals. We chose seizure subfolders (6 hours per subfolder) for training, with a total duration of approximately 38 hours, as shown in Table 4.

Table 4: General information of epileptic patients of the XJSZ-EEG dataset.

Patient ID	Gender	Age	EEG Channels	Seizure Events	Total Seizure Time (s)	Total Time (h)
01	Male	24	19	14	420	98.3
02	Male	16	19	13	390	125.5
03	Male	34	18	32	960	100.5
04	Male	14	19	18	540	244.7
05	Male	16	19	13	390	125.5
06	Male	13	18	32	962	100.5
07	Male	14	19	18	540	244.7
08	Male	8	18	41	1231	401.5
09	Female	17	18	6	362	24
10	Female	16	18	32	960	237
11	Female	5	17	43	1295	166
12	Male	34	19	3	107	72

• **XJSZ-SEEG:** The dataset is collected from the Department of Neurosurgery at a firstclass hospital, monitoring the SEEG signals of 20 clinical epilepsy patients. All signals are recorded using the Neruacle digital SEEG machines (NSH0256). Each patient have a personalized surgical plan, and the number of SEEG contacts varied, with each contact representing an SEEG channel. The dataset is manually annotated by two clinically experienced physicians. The sampling frequency is 1000 Hz, with a total duration of 388 hours. We used the data from 12 patients for pre-training, not by summing the total duration of each patient's data but by selecting seizure subfolder data from each patient. The remaining 8 patients are used for downstream tasks, as shown in Table 5.

Table 5: General information of epileptic patients of the XJSZ-SEEG dataset.

Patient ID	Gender	Age	SEEG Electrodes	SEEG Channels	Seizure Events	Total Seizure Time (s)	Total Time (h)
01	Female	33	9	127	146	4380	384
02	Female	5	10	132	65	10549	154
03	Female	11	8	126	138	2438	94
04	Male	15	13	196	12	2430	237
05	Female	10	11	142	16	1067	92
06	Male	28	13	166	26	1222	141
07	Male	7	10	130	17	1020	169
08	Female	17	11	141	6	362	24
09	Female	34	14	190	5	202	120
10	Female	11	10	141	14	427	67
11	Female	28	10	141	11	385	102
12	Male	34	13	200	3	107	72
13	Male	16	13	166	13	390	125.5
14	Male	14	13	188	18	540	244.7
15	Male	8	9	106	41	1231	401.5
16	Male	34	14	194	32	962	100.5
17	Female	5	10	120	43	1295	166
18	Female	11	10	118	138	4142	93.5
19	Female	34	11	127	44	1326	118.5
20	Male	24	13	195	5	157	98.3

#### 864 С MODEL DETAILS 865

866

867

868

869

870

871

873 874 875

903 904

The EpilepsyFM Tokenizer Training model comprises three key components: (1) the Epilepsy Neural Tokenizer, (2) the Vector Quantizer, and (3) the EpilepsyFM Neural Decoder, as shown in Table 6. As illustrated in Figure 2, the architecture of the Neural Tokenizer efficiently processes and tokenizes epilepsy-related data. The Vector Quantizer is designed similarly to the LaBraM model, ensuring effective data compression and representation. The EpilepsyFM Neural Decoder features a Transformer decoder with stacked layers and a temporal regression head, enabling accurate modeling 872 of time series data. Additionally, the pre-training of EpilepsyFM aligns with the Epilepsy Neural Transformer module, as shown in Table 7.

876ModuleSub-ModuleNameValue877ModuleSub-ModuleNameValue878Temporal Encoder $\#$ of Input Channels $\{1,8,8\}$ 879EpilepsySpectral Encoder $\#$ of Output Channels $\{1,2,2\}$ 881Neural TokenizerSpectral Encoder $\#$ of Output Channels $\{1,2,2\}$ 882EpilepsySpatial Encoder $\#$ of Output Channels $\{10,16,16\}$ 883Neural Transformer)Fransformer Encoder $\#$ of Transformer Layers8884Transformer Encoder $\#$ of Attention Heads10885Vector Quantizer-Codex Size $8192 \times 64$ 886Vector Quantizer-Embedding-to-Codex $200 \rightarrow 200(Tanh) \rightarrow 6$ 888Vector Quantizer- $Embedding-to-Codex$ $200 \rightarrow 200(Tanh) \rightarrow 6$ 889 $\#$ of Transformer Layers3
Temporal Encoder# of Input Channels $\{1,8,8\}$ 879EpilepsySpectral Encoder# of Output Channels $\{1,8,8\}$ 880EpilepsySpectral Encoder# of Input Channels $\{1,2,2\}$ 881Neural Tokenizer (EpilepsySpatial Encoder# of Input Channels $\{1,2,2\}$ 883Neural Transformer)Spatial Encoder# of Output Channels $\{10,16,16\}$ 884Transformer Encoder# of Transformer Layers8884Transformer Encoder# of Attention Heads10886Vector Quantizer-Codex Size Codex-to-Embedding $8192 \times 64$ 887Vector Quantizer-Codex Size Codex-to-Embedding $8192 \times 64$ 888Vector Quantizer-Codex Size Codex-to-Embedding $8192 \times 64$ 889Vector Quantizer-State Size Codex Size $8192 \times 64$ 889Vector Quantizer-State Size Codex Size $8192 \times 64$ 889Xetter Quantizer-State Size Codex Size $8192 \times 64$ 889Xetter Quantizer-State Size Codex Size $8192 \times 64$
879# of Output Channels $\{8,8,8\}$ 880Epilepsy881Spectral Encoder# of Input Channels $\{1,2,2\}$ 881Neural Tokenizer (EpilepsySpectral Encoder# of Input Channels $\{1,2,2\}$ 882(EpilepsySpatial Encoder# of Input Channels $\{10,16,16\}$ 883Neural Transformer)# of Transformer Layers8884Transformer EncoderHidden Size $200$ 885MLP Size $800$ 886Codex Size $8192 \times 64$ 887Vector Quantizer-Codex Size $8192 \times 64$ 888Transformer Layers $3$
880Epilepsy Neural Tokenizer (EpilepsySpectral Encoder# of Input Channels $\{1,2,2\}$ # of Output Channels882Neural Tokenizer (EpilepsySpatial Encoder# of Input Channels $\{2,2,2\}$ 883Neural Transformer)Fransformer Encoder# of Output Channels $\{1,6,16,8\}$ 884Transformer Encoder# of Transformer Layers8885Fransformer EncoderHidden Size200886Codex Size8192×64887Vector Quantizer-Codex Size8192×64888Vector Quantizer-Embedding-to-Codex $200 \rightarrow 200(Tanh) \rightarrow 6$ Codex-to-Embedding $64 \rightarrow 200$ 889# of Transformer Layers3
Reprint procession881Implicity processionNeural Tokenizer (EpilepsyFortial Encoder882Neural Transformer)Spatial Encoder# of Input Channels $\{10,16,16\}$ 883Neural Transformer)# of Transformer Layers8884Transformer EncoderHidden Size MLP Size $200$ MLP Size886Codex Size Embedding-to-Codex $200 \rightarrow 200(Tanh) \rightarrow 6$ Codex-to-Embedding887Codex Size Embedding-to-Codex $200 \rightarrow 200(Tanh) \rightarrow 6$ Codex-to-Embedding888Transformer Layers $3$
882InterferenceSpatial Encoder# of Transformer Layers(Fortest, 6)883Neural Transformer) $=$ # of Output Channels{16,16,8}884Transformer Encoder# of Transformer Layers8885 $=$ Transformer EncoderHidden Size200886 $=$ Odd Attention Heads10886 $=$ Codex Size8192×64887Vector Quantizer-Embedding-to-Codex888 $=$ Codex-to-Embedding $64 \rightarrow 200$ 889 $=$ Transformer Layers3
883Neural Transformer)# of Transformer Layers8884Transformer Encoder# of Transformer Layers8885Transformer EncoderHidden Size200886MLP Size800886# of Attention Heads10886Codex Size8192×64888Vector Quantizer-Embedding-to-Codex Codex-to-Embedding $200 \rightarrow 200(Tanh) \rightarrow 6$ $64 \rightarrow 200889# of Transformer Layers3$
884 885Transformer EncoderHidden Size MLP Size200 800 800 # of Attention Heads886 887 888 888 889Vector Quantizer Codex Size Embedding-to-Codex Codex-to-Embedding $8192 \times 64$ $200 \rightarrow 200(Tanh) \rightarrow 6$ $64 \rightarrow 200$ 889# of Transformer Layers3
885MLP Size800886# of Attention Heads10886Codex Size $8192 \times 64$ 887Vector Quantizer-Embedding-to-Codex Codex-to-Embedding $200 \rightarrow 200(Tanh) \rightarrow 6$ $64 \rightarrow 200$ 889# of Transformer Layers3
# of Attention Heads10886# of Attention Heads10887Codex Size $8192 \times 64$ 888Vector Quantizer-Embedding-to-Codex $200 \rightarrow 200(Tanh) \rightarrow 6$ 889# of Transformer Layers3
$\begin{array}{c} 887 \\ 888 \\ 888 \\ 889 \end{array} \qquad - \qquad \begin{array}{c} Codex Size \\ Embedding-to-Codex \\ Codex-to-Embedding \\ 64 \rightarrow 200 \end{array}$
Vector Quantizer - Embedding-to-Codex $200 \rightarrow 200(Tanh) \rightarrow 6$ Codex-to-Embedding $64 \rightarrow 200$
889 # of Transformer Layers 3
# of Transformer Lavers 4
890 Enilopey Hidden Size 200
891 Neural Decoder Transformer Decoder MI P Size 800
892# of Attention Heads10
Batch Size 512
894 Maximum Learning Rate 5e-5
895 Minimum Learning Rate 1e-6
896 Learning Rate Scheduler Cosine
897 Optimizer - Optimizer Type AdamW
$\begin{array}{c} Adam & (0.9,0.99) \\ W = 1 + D & 0.01 \\ \end{array}$
Weight Decay 0.01
900 Warm-up Epochs 30
901
902

Table 7. The hyperparameters for the Ephepsyl Wi	Table 7:	The hyperparameters	for the	EpilepsyFM.
--	----------	---------------------	---------	-------------

905	Hyperparameters	Values
906	Batch Size	64
907	Maximum Learning Rate	5e-4
908	Minimal Learning Rate	1e-6
909	Learning Rate Scheduler	cosine
910	Öptimizer	adamw
911	$Adam \beta$	(0.9, 0.99)
912	Weight Decay	0.05
913	Total Epochs	30
914	Warmup Epochs	5
915	Drop Path	0.1
916	Layer-wise Learning Rate Decay	0.9
917	Label Smoothing	0.1

# 918 D EVALUATION METRICS

In the task of epileptic seizure detection(Cai et al., 2023; Chen et al., 2022), we follow existing research and employ balanced accuracy, AUC-PR, and AUROC as evaluation metrics for binary classification tasks. For multi-class classification tasks(Memar & Faradji, 2017), we use balanced accuracy, Cohen Kappa, and weighted F1 as evaluation metrics. In the short-term and long-term prediction tasks for epilepsy (Woo et al., 2022), we also adhere to existing research by utilizing MAE, MSE as evaluation metrics.

- **Balanced accuracy:** Balanced accuracy is a method of calculating accuracy that accounts for class imbalance. It is the average of the recall for each class. In epilepsy detection, balanced accuracy effectively reflects the model's classification performance in the face of imbalanced data, avoiding the bias that can arise from simple accuracy.
- AUC-PR: AUC-PR measures the area under the curve plotted between precision and recall, primarily used to evaluate the performance of classification models on imbalanced data. In epilepsy datasets, AUC-PR better assesses model performance in scenarios with scarce positive samples, particularly focusing on the precision and recall of the model.
- AUROC: AUROC measures the area under the receiver operating characteristic curve, primarily used to evaluate the performance of classification models across various threshold levels. In epilepsy datasets, AUROC better assesses the overall performance of the model under different thresholds.
- Cohen Kappa: Cohen Kappa is a statistic that measures the agreement between two classifiers while accounting for the effect of random chance. Its value ranges from -1 to 1, where 1 indicates perfect agreement, 0 indicates no better than random agreement, and negative values indicate agreement below random levels. In multi-class tasks, Cohen Kappa effectively evaluates the consistency of the model's classifications, especially in situations of class imbalance.
- Weighted F1 Score: Weighted F1 Score is an extension of the F1 score that considers the importance of each class. It calculates the F1 score for each class and weighs it according to the frequency of that class in the dataset. The Weighted F1 Score is particularly important in multi-class tasks, as it balances the performance of all classes, making it suitable for addressing class imbalance issues and ensuring effective evaluation of the model across all categories.
  - Mean Absolute Error (MAE): MAE calculates the average of the absolute differences between predicted values and actual values. MAE treats all errors equally, avoiding the bias introduced by MSE towards extreme values, thus providing an intuitive assessment of model performance. In epilepsy prediction, MAE quantifies the overall predictive accuracy of the model clearly, particularly when ensuring the reliability of each prediction is critical.
    - **Mean Squared Error** (**MSE**): MSE refers to the average of the squared differences between predicted values and actual values. The primary advantage of MSE lies in its sensitivity to larger errors, making it useful in model optimization to identify and reduce significant biases. In the context of epilepsy prediction, MSE effectively reflects how well the model captures rare but severe seizure events.

### E VISUALIZATION OF MASK MODELING

In Figure 6, we present the convergence curves of the total pre-training loss and the masked modeling
 accuracy of EpliespyFM. It is evident from the figure that the training method utilizing channel set
 masked modeling effectively facilitates the learning process of the model. Specifically, the training
 loss steadily decreases with each iteration, indicating that the model is progressively optimizing its
 parameters to enhance performance. At the same time, the masked accuracy shows a consistent
 upward trend, further validating the effectiveness of this approach in capturing the features of EEG and SEEG data.



Figure 6: The loss curve and accuracy curve during the training process of the EpliespyFM.

### F ABLATION ON CHANNEL-SET MASKING

Due to the locality and specificity of brain computation, research into more challenging epilepsy tasks (such as the clustering discharges and propagation circuits of different brain regions in different patients) remains to be explored. Our designed channel masking strategy can cover the neural activity signals of multiple neighboring channels, which exhibit signal similarity, as shown in Figure 7. SEEG signal channels are continuous, so there is no need to consider channel reordering. For EEG signals, we reorder the electrode signals based on the left and right hemispheres, grouping nearby brain region channels together to implement the masking.

In the experiments, we evaluated the performance of EEG seizure detection using different numbers of masked channels, as shown in Table 8. From the figure, it can be observed that in most datasets, the detection performance is best with three channels, and the masking methods using three or four channels perform slightly better than the single-channel masking approach. Utilizing multiple channels for masking can more effectively capture and analyze the characteristic information of seizure events. Compared to single-channel masking, this approach may provide richer spatial and temporal information, enhancing the model's ability to learn complex signal patterns.





983

984 985

Figure 7: The layouts and information of EEG and SEEG electrode channels.

	Table 8: 7	The ablation	study of chai	nnel set maskin	ıg.	
Channel-set Mask		TUAB			CHB-MIT	
	Balance Acc.	AUC-PR	AUROC	Balance Acc.	AUC-PR	AUROC
1	81.77±0.54	86.73±0.37	89.53±0.56	83.45±0.07	61.98±0.35	93.32±0.15
3	82.27±0.18	89.77±0.17	91.91±0.11	85.12±0.34	64.25±0.45	92.01±0.52
4	82.10±0.27	89.66±0.12	90.91±0.21	85.42±0.51	64.00±0.57	91.41±0.47
Channel-set Mask		XJSZ-EEG			TUEV	
	Balance Acc.	AUC-PR	AUROC	Balance Acc.	AUC-PR	AUROC
1	79.35±0.87	81.58±0.67	84.69±0.48	61.76±0.58	61.33±0.65	80.34±0.68
3	80.35±0.83	82.05±0.68	84.25±0.73	62.74±0.76	64.35±1.01	81.45±0.96
4	81.57±0.26	81.45±0.73	86.37±0.36	62.25±0.38	67.11±0.46	81.79±0.47

### G ABLATION ON ENCODER MODULE

In the experiments, we found that the temporal encoder (TEME), spectral encoder (SPEE), and spatial encoder (SPAE) significantly impact the performance of the epilepsy model, as shown in Table 9. Removing the temporal encoder leads to a noticeable decline in balanced accuracy and AUC metrics, indicating that temporal information is crucial for capturing the dynamic characteristics of epileptic seizures. The absence of the spectral encoder also affects predictive performance, particularly in the analysis of long time series. Although the impact of the spatial encoder is relatively smaller, its removal still results in a performance drop, highlighting the importance of spatial features. Ultimately, the EpilepsyFM model, which combines all encoder modules, demonstrates the best performance. 

Table 9: The ablation study of encoder module							
_		TUAB			CHB-MIT		
	Balance Acc.	AUC-PR	AUROC	Balance Acc.	AUC-PR	AUROC	
w/o TEME	81.47±0.15	89.32±0.10	89.88±0.08	82.12±0.20	65.23±0.15	90.12±0.12	
w/o SPEE	80.45±0.18	88.75±0.12	$89.00 \pm 0.09$	81.50±0.25	$64.10 \pm 0.20$	89.50±0.15	
w/o SPAE	79.65±0.20	87.40±0.15	88.50±0.10	80.00±0.30	$63.55 \pm 0.25$	88.80±0.20	
EpilepsyFM	82.27±0.18	89.77±0.17	91.91±0.11	85.12±0.34	$64.25 \pm 0.45$	92.01±0.52	
_		XJSZ-EEG			TUEV		
	Balance Acc.	AUC-PR	AUROC	Balance Acc.	AUC-PR	AUROC	
w/o TEME	80.13±0.23	81.86±0.43	83.32±0.66	62.76±0.33	61.76±1.23	80.45±0.45	
w/o SPEE	79.00±0.51	81.20±0.34	83.00±0.41	61.50±0.60	63.00±1.00	80.20±0.90	
w/o SPAE	78.50±0.43	80.00±0.25	82.00±0.32	60.80±0.55	62.50±0.80	79.80±0.75	
EpilepsyFM	80.35±0.83	82.05±0.68	84.25±0.73	62.74±0.76	64.35±1.01	81.45±0.96	

## 1069 H DISCUSSION

Limitations. Although EpilspyFM has made significant progress in common clinical tasks for epilepsy by training on multiple publicly available EEG datasets and private data, the availability of SEEG data in clinical settings remains relatively limited. This somewhat constrains our in-depth exploration in the field of foundational models for epilepsy. If we can further enrich task settings, particularly in the precise localization of epileptic foci and avoidance of functional areas, it will greatly enhance the clinical applicability of the model, allowing for more effective guidance in treatments such as thermal coagulation.

1078 Outlook. Based on the aforementioned limitations, our research outlook encompasses several key
 1079 areas: First, we plan to expand the collection of SEEG data, particularly from different devices, to enhance the diversity and representativeness of the data; we are currently actively promoting

this effort. Second, we will extend the task settings to consider the relationship between epileptic lesion areas and brain functional regions, thereby improving the comprehensiveness and depth of our research. Lastly, we will explore the potential of multimodal integration. Our existing model already includes EEG and SEEG as two types of neural signals, and in the future, we may consider incorporating fMRI or PET data. This direction not only poses technical challenges but also represents an innovative approach with significant clinical implications. Through these efforts, we aim to achieve a deeper understanding and breakthroughs in the field of epilepsy research.