LEAD: An EEG Foundation Model for Alzheimer's Disease Detection

Anonymous authors
Paper under double-blind review

Abstract

Electroencephalography (EEG) provides a non-invasive, highly accessible, and cost-effective approach for detecting Alzheimer's disease (AD). However, existing methods, whether based on handcrafted feature engineering or standard deep learning, face three major challenges: 1) the lack of large-scale EEG-based AD datasets for robust representation learning and evaluation; 2) limited cross-subject generalizability; and 3) difficulty in adapting to highly heterogeneous data. To address these challenges, we curate the world's largest EEG-AD corpus to date, comprising 2,238 subjects. Leveraging this unique resource, we propose LEAD, the first foundation model for EEG-based AD detection. Specifically, we design a gated temporal-spatial Transformer that can adapt to EEG recordings with arbitrary lengths, channel configurations, and sampling rates. In addition, we introduce a subject-regularized training strategy to enhance end-to-end subject-level detection. We further employ medical contrastive learning to pre-train on 13 datasets, including 4 AD datasets and 9 non-AD neurological disorder datasets, and fine-tune/test the model on the other 5 AD datasets. LEAD achieves the best average ranking across all 20 evaluations on 5 downstream datasets, substantially outperforming existing approaches, including state-of-the-art (SOTA) EEG foundation models. These results strongly demonstrate the effectiveness and practical potential of the proposed method for real-world EEG-based AD detection. Source code: https://anonymous.4open.science/r/LEAD-3B51

1 Introduction

Table 1: **Comparison of Different AD Detection Modalities.** Different modalities present distinct trade-offs with respect to accessibility, safety, affordability, and detection accuracy.

Technologies	Accessibility↑	Safety↑	Affordability↑	Accuracy↑
Cerebrospinal Fluid (Salvadó et al., 2024)	Low	Low	Low	High
Neuroimaging (Ottoy et al., 2025)	High	High	Low	High
Blood Test (Grande et al., 2025)	Low	Moderate	Moderate	Moderate
Neurological Exam (Vöglein et al., 2023)	High	High	Moderate	Moderate
EEG (Wang et al., 2024d)	High	High	High	Low

Alzheimer's Disease (AD) is the most common neurodegenerative disorder in the elderly, affecting 38 million individuals with $1.3 trillion annual financial cost (Breijyeh & Karaman, 2020; Masters et al., 2015; National Academies of Sciences, Engineering, and Medicine, 2021). Early detection of AD is a significant step to slow symptom progression and increase patients' life expectancy (Nelson & Tabet, 2015; Chu, 2012). Compared with other AD detection modalities, electroencephalography (EEG) offers appealingly unique advantages (Table 1). These include high accessibility through portable devices, strong safety due to its non-invasive nature, and high affordability with low hardware and operational costs (Ieracitano et al., 2019). In addition, EEG enables long-term monitoring of disease progression with high temporal resolution and real-time capability (Ahmed et al., 2025). Existing EEG-based AD detection methods generally fall into two main research directions. The first relies on handcrafted biomarkers, such as phase shift (Wang et al., 2017), power spectral density (Fahimi et al., 2017), and Shannon entropy (Azami et al., 2019). The second direction leverages deep learning models, including convolutional neural networks (Roncero-Parra et al., 2024), graph neural networks (Klepl et al., 2023), and Transformer-based architectures (Wang et al., 2024d).

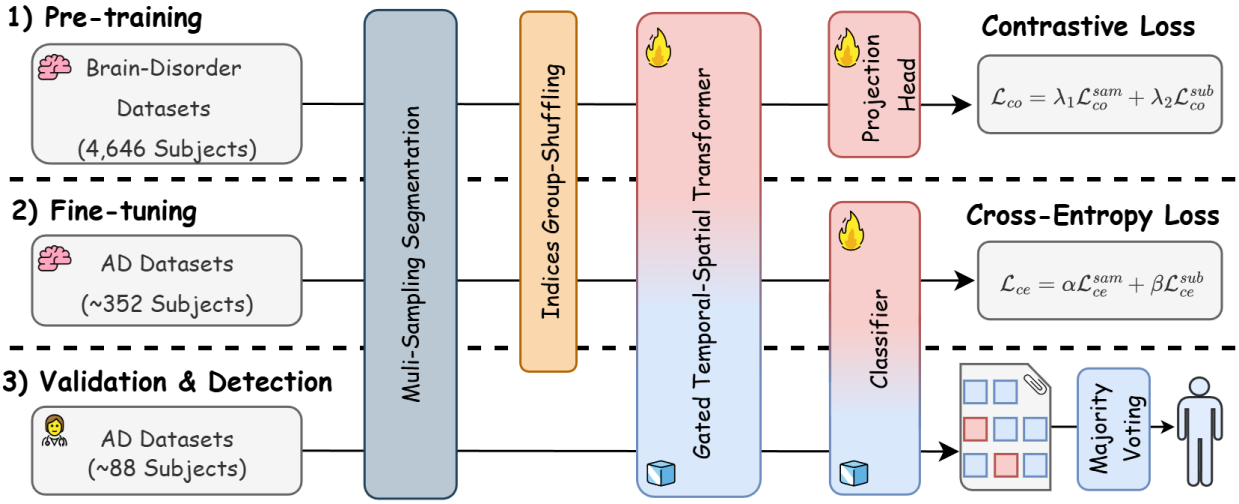

Figure 1: **LEAD overview.** We pre-train LEAD on 13 neurological disorder datasets using medical contrastive learning, and subsequently fine-tune it on five AD datasets under a **subject-independent cross-validation** protocol, with training, validation, and test splits of 80%, 10%, and 10%, respectively. A gated temporal-spatial Transformer is employed to jointly capture temporal and spatial features. In addition, we propose a subject-level cross-entropy loss $\mathcal{L}_{\text{ce}}^{\text{sub}}$ to enhance subject-level representation learning. Engineering tricks, including multi-scale segmentation and index group shuffling, are further introduced to facilitate subject-level training. Finally, majority voting is applied to obtain subject-level detection results.

However, existing EEG-based AD detection methods, whether using manual feature extraction or deep learning, still suffer from several limitations. **First,** large and high-quality datasets remain scarce. The cost and complexity of collecting EEG data from patients with AD result in most studies being conducted on relatively small cohorts, typically ranging from 20 to 100 subjects (Aviles et al., 2024). The limited diversity and scale of existing EEG datasets pose a fundamental challenge to learn robust models. **Second,** model generalizability across subjects remains a critical bottleneck. The core challenge for real-world clinical deployment is generalization to unseen subjects (subject-independent evaluation). Current approaches, while often strong in controlled, at-lab settings (subject-dependent), suffer a substantial performance drop when moving to subject-independent evaluation (Wang et al., 2024b). This degradation is largely caused by inter-subject variability and shortcut learning between diagnostic labels and subject-specific characteristics (Wang et al., 2024c). Although subjects diagnosed with AD are expected to exhibit consistent disease-related patterns, confounding factors such as age, gender, and other personal attributes may obscure these patterns. **Third,** existing studies are difficult to adapt to heterogeneous data. Most prior works train models on a single dataset or on datasets with identical acquisition configurations (Nayana et al., 2025). In practice, EEG data are highly heterogeneous, not only due to demographic differences among subjects but also because of variations in recording environments, acquisition devices, channel montages, sampling rates, and recording durations. This heterogeneity makes it challenging to transfer the detection capability of existing methods across datasets, which is critical for real-world deployment.

To address the aforementioned limitations, we propose **LEAD**, the first EEG foundation model for AD detection. We conduct a comprehensive review of relevant publications and public EEG repositories (by 2026), including OpenNeuro, Dryad, and Figshare, and identify 9 highly heterogeneous EEG datasets for AD and dementia detection. These datasets comprise **2,238 subjects and 427.81 hours** of recordings. To the best of our knowledge, this constitutes the world's **largest** EEG dataset for AD detection, which is approximately **20 times larger** than those commonly used in prior publications. Building on this unique resource, we train an EEG foundation model designed for **subject-level** Alzheimer's detection that can handle EEG signals of **arbitrary length, channel montages, and sampling rates**. To this end, we design a gated temporal-spatial Transformer as the backbone. Univariate patch embedding is used to adapt to downstream tasks with varying input lengths and channel counts. In addition, 3D channel embeddings and sampling rate embeddings allow the model to capture spatial relationships and to distinguish EEG recorded

at different sampling rates. A parallel temporal-spatial self-attention mechanism is employed to extract both temporal and spatial features, with a learnable gated fusion module to adaptively control their relative importance. Furthermore, we introduce a subject-regularized training strategy. A subject-level cross-entropy loss is proposed to propagate subject-level errors back to the model during training. Useful engineering tricks, such as multi-scale segmentation and index group shuffling, are used to further enhance subject-level learning. Finally, we employ a domain-inspired self-supervised medical contrastive learning strategy for pre-training on 4 of the 9 AD datasets, along with 9 additional neurological disorder datasets, comprising a total of 4,646 subjects and 1,185.84 hours of EEG recordings. The remaining 5 AD datasets are used for fine-tuning and evaluation under a realistic yet challenging subject-independent cross-validation protocol. Figure 1 illustrates the overall pipeline of our proposed method. We compare our method against 16 baseline methods, including manual feature extraction, supervised deep learning methods, and SOTA EEG foundation models. LEAD achieves the best average ranking across all 20 evaluations on 5 datasets, far surpassing the second-best model LaBraM while using significantly fewer pre-training resources.

The key contributions of this work are summarized as follows: 1) **The largest EEG-based AD detection dataset.** We curate the largest dataset for EEG-based AD detection to date, far surpassing prior work. 2) **First EEG Foundation Model for AD detection.** Built on this unique dataset, we develop the first foundation model for end-to-end subject-level EEG-based AD detection, adaptable to diverse recording lengths, channel topologies, and sampling rates. **3) Superior performance.** LEAD achieves SOTA subject-level results on 5 downstream AD datasets under challenging subject-independent cross-validation.

2 Related Work

2.1 EEG-Based AD Detection

In the last two decades, EEG-based AD detection has followed two main research directions: manual biomarker extraction and deep learning methods.

Manual Biomarker Extraction: This research direction aims to identify potential biomarkers in EEG signals of AD patients and use simple classifiers, such as Multi-Layer Perceptrons (MLP) and Support Vector Machines (SVM), to differentiate these features from normal healthy subjects. Different types of EEG features are used, including statistical features like Mean, Skewness, Kurtosis, and Standard Deviation (Tzimourta et al., 2019b;a; Kulkarni & Bairagi, 2017; Kanda et al., 2014; Waser et al., 2013; Tylova et al., 2013; Mora-Sánchez et al., 2019), spectral features like Phase Shift, Phase Coherence, Bispectrum, and Bicoherence (Wang et al., 2017; Cassani et al., 2014; Wang et al., 2015; Fraga et al., 2013; Tait et al., 2019; Waser et al., 2016; Trambaiolli et al., 2011), power features like Power Spectrum Density, Relative Band Power, Ratio of EEG Rhythm, and Energy (Fahimi et al., 2017; Schmidt et al., 2013; Liu et al., 2016; Kanda et al., 2014), as well as complexity features like Shannon Entropy, Tsallis Entropy, and Permutation Entropy (Garn et al., 2015; Azami et al., 2019; Tylová et al., 2018; Coronel et al., 2017; Al-Nuaimi et al., 2018). The main advantage of this approach is its interpretability, but it suffers from limited performance.

Deep Learning: Compared to manual biomarker extraction, deep learning offers an alternative approach by automatically extracting useful representations for AD detection. Models such as Convolutional Neural Networks (CNNs) (Li et al., 2022; Cura et al., 2022), Graph Neural Networks (GNNs) (Shan et al., 2022; Klepl et al., 2023), and Transformers (Wang et al., 2024d) are widely used for representation learning. Some researchers still perform manual feature extraction or transform the data before applying deep learning models. For example, the method in (Ieracitano et al., 2019) converts 5-second EEG intervals into Power Spectral Density (PSD) images and uses 2D convolutional layers on the images for feature extraction. DICE-net (Miltiadous et al., 2023a) extracts relative band power and spectral coherence connectivity across five frequency bands and applies convolutional layers followed by transformers. In contrast, some studies apply deep learning methods directly to EEG data. For instance, the method in (Gallego-Viñarás et al., 2024) uses semi-supervised spatiotemporal representation learning with deep learning models for AD detection based on different sleep-stage EEG data. STEADYNet (Kachare et al., 2024) designs low-complexity convolutional models for AD and dementia detection, focusing on fast inference times. Research in (Watanabe et al., 2024) using MNet that applies convolutional networks for feature extraction and concatenates with the relative

power spectrum for AD and other dementia detection. ADformer (Wang et al., 2024d) uses a multi-granularity transformer for AD detection and widely tests on 4 AD datasets. Research in (Nayana et al., 2025) compares biomarker extraction and deep learning methods for the detection of neurodegenerative diseases like AD.

2.2 Self-Supervised and Foundation Model in EEG

Two main strategies are widely used for self-supervised representation learning and foundation model training in EEG: contrastive learning and mask-reconstruction.

Contrastive Learning: BENDR (Kostas et al., 2021) follows a similar contrastive learning pipeline as Wav2Vec (Baevski et al., 2020), but it is trained on EEG data. EEG2Vec (Zhu et al., 2023) explores both contrastive learning and mask-reconstruction for self-supervised pre-training on EEG data. BIOT (Yang et al., 2024) designs a transformer architecture for biomedical signal embedding and applies a self-supervised contrastive framework similar to BYOL (Grill et al., 2020). COMET (Wang et al., 2023) utilizes various data levels in biomedical time series to define positive and negative pairs.

Mask-Reconstruction: Neuro-BERT (Wu et al., 2024) employs masked autoencoding to predict missing amplitude and phase of EEG signals during pre-training. EEG2Rep (Mohammadi Foumani et al., 2024) combines a context encoder with a momentum target encoder to reconstruct context-level representations rather than raw data in self-supervised pre-training. LaBraM (Jiang et al., 2024b), the first large foundation model in the EEG domain, uses a neural tokenizer to reconstruct the Fourier spectrum during self-supervised pre-training. EEGPT (Wang et al., 2024a) is a foundation model for EEG representation learning that integrates reconstruction loss with an alignment loss between the encoder and momentum encoder. CBraMod (Wang et al., 2025) designs a criss-cross transformer to leverage both spatial and temporal features of EEG. LUNA (Döner et al.) introduces a cross-attention mechanism for a topology-agnostic foundation model capable of handling heterogeneous channels. CSBrain (Zhou et al., 2025) utilizes regions of the brain to design cross-window and cross-region dependencies for diverse decoding EEG tasks. NeurIPT (Fang et al., 2025) leverages a progressive mixture-of-experts architecture to build backbones and introduces 3D electrode embedding for EEG channel embedding. REVE (Ouahidi et al., 2025) pretrains on over 60,000 hours of EEG data from 92 datasets spanning 25,000 subjects, demonstrating the largest pre-training resources so far.

Other Strategies. Recent work has begun exploring autoregressive pre-training for EEG, such as NeuroGPT (Cui et al., 2024) and a study also named EEGPT (Yue et al., 2024). Research bridging language and EEG has also emerged, for instance, NeuroLM (Jiang et al., 2024a), which treats EEG as a foreign language for representation learning. Laya applies the JEPA (Assran et al., 2023) approach on EEG representation learning (Panchavati et al., 2026).

Application-Oriented Approaches. Several studies focus on application-specific foundation models. Brant series (Zhang et al., 2023; 2024) train models tailored to intracranial neural signals; PPi (Yuan et al., 2023) develops a self-supervised framework for subject-independent seizure detection; MIRepNet (Liu et al., 2025) designs a foundation model for motor imagery classification. FAPEX (Zheng et al., 2025) utilizes the fractional amplitude-phase expressor for cross-subject seizure prediction.

3 Method

3.1 Preliminaries

Neurodegenerative Diseases Detection. Neurodegenerative disease detection using EEG typically involves assigning a single label to each subject, reflecting their physiological brain state at the time of data acquisition, such as the presence or absence of AD. This formulation assumes that a subject's state remains stable over the relatively short recording period (e.g., several days to several months), so that all EEG data collected from the same subject share the same label. Exceptions arise with longitudinal datasets that track disease progression (e.g., Healthy $\rightarrow$ Mild Cognitive Impairment (MCI) $\rightarrow$ AD (Vecchio et al., 2018; Ge et al., 2025)); in such cases, trials/sessions reflecting a change in diagnosis for the same subject are treated as different subjects during training. Besides, comorbid diseases are often reported in neurodegenerative diseases. For instance, a patient may present with both AD and Vascular Dementia (VD) (Kim et al., 2023). This

introduces an inherently multi-label scenario, in which each disease should be predicted independently via a dedicated classifier formulated as a binary classification task. In this work, consistent with the datasets used in our experiments, **we simplify the problem formulation by assuming each subject is assigned a single label, with no label transitions during data collection and no comorbid conditions.**

Problem Formulation. EEG-based AD detection aims to identify AD and related cognitive conditions from EEG recordings. While most existing deep learning methods focus on classifying individual EEG samples (sample-level classification), our approach emphasizes end-to-end subject-level detection, which is more clinically relevant. Consider an input EEG sample $\boldsymbol{X} \in \mathbb{R}^{C\times T}$, where C denotes the number of channels (electrodes) and T is the number of timestamps. Let the corresponding class label be $y \in \{1, \ldots, K\}$, where K is the number of categories, such as AD, Healthy Controls (HC), Mild Cognitive Impairment (MCI), Frontotemporal Dementia (FTD), or other dementia-related conditions. We further assign each sample a subject ID and a sampling rate marker, denoted by s and r, respectively.

(1) Sample-Level Classification: The objective is to learn a model $f(\cdot)$ that predicts the class label y for each EEG sample $\boldsymbol{X}$. **(2) Subject-Level Detection:** Given multiple EEG samples belonging to the same subject s, we aggregate their sample-level predictions using majority voting (Sec. 3.3). The final subject-level detection is assigned as the class with the highest vote count among all samples from subject s.

3.2 Gated Temporal-Spatial Transformer

Multi-Scale Segmentation. Instead of resampling EEG signals to a single fixed rate, we downsample each recording to multiple sampling rates. Specifically, signals with original rates above 200 Hz are downsampled to 200, 100, and 50 Hz. We then segment the signals into half-overlapping samples of 100, 200, or 400 timestamps (corresponding to 1s, 2s, and 4s at 100Hz). Boundary segments shorter than 100, 200, or 400 timestamps are discarded. Each segmented sample is assigned a sampling rate label r for subsequent use. This segmentation strategy naturally enables multi-scale feature extraction, allowing the model to capture both fine-grained and coarse temporal patterns. Moreover, since our ultimate goal is subject-level AD detection, this approach increases the total number of samples per subject, which benefits both model training and post-hoc majority voting.

Patch Embedding. To handle EEG samples with arbitrary channel configurations and variable recording lengths, we implement a patch embedding strategy as follows. Consider an input EEG sample $\boldsymbol{X} \in \mathbb{R}^{C\times T}$, for a patch length L, we slice each channel into N univariate patches. This yields a patch tensor $\boldsymbol{X}_p \in \mathbb{R}^{C\times N\times L}$. Zero padding is applied when necessary to ensure that T is divisible by L, resulting in $N = \lceil T/L \rceil$. Each patch is then projected into a D-dimensional embedding space through a linear mapping $\boldsymbol{W} \in \mathbb{R}^{L\times D}$, producing patch embeddings $\boldsymbol{E} \in \mathbb{R}^{C\times N\times D}$. A fixed temporal positional embedding $\boldsymbol{PE} \in \mathbb{R}^{N\times D}$ is added and broadcast across all C channels. The final patch embeddings are obtained as:

$$\boldsymbol{E} \leftarrow \boldsymbol{X}_p\boldsymbol{W} + \text{Broadcast}\,(\boldsymbol{PE}), \quad \boldsymbol{E} \in \mathbb{R}^{C\times N\times D} \tag{1}$$

3D Channel Embedding. To enable the model to flexibly adapt to arbitrary channel configurations and to learn spatial relationships among channels, we employ the 3D EEG channel embedding introduced in NeurIPT (Fang et al., 2025), with minor modifications. Let $\boldsymbol{P} \in \mathbb{R}^{C\times 3}$ denote the 3D channel coordinate matrix, where the c-th channel has spatial coordinates $\boldsymbol{P}_c = (x_c, y_c, z_c)$ for $c = 1, \ldots, C$. The $\boldsymbol{P}_c$ denotes the relative positions of EEG channels following the international 10-20 system (Homan et al., 1987). To encode spatial information in a parameter-free manner, we apply sinusoidal positional encoding independently along each spatial axis and concatenate the resulting embeddings to form a D-dimensional channel embedding. Given the target embedding dimension D, we automatically allocate dimensions across the 3 spatial axes as:

$$D_x = \left\lfloor \frac{D}{3} \right\rfloor, \quad D_y = \left\lfloor \frac{D}{3} \right\rfloor, \quad D_z = D - D_x - D_y \tag{2}$$

For an axis coordinate $p \in \{x_c, y_c, z_c\}$, we define an axis-specific sinusoidal embedding $\text{CE}(p) \in \mathbb{R}^{D_p}$, where $D_p \in \{D_x, D_y, D_z\}$, as:

$$\text{CE}(p)_{2i} = \sin\left(p \cdot 10000^{-\frac{2i}{D_p}}\right), \quad \text{CE}(p)_{2i+1} = \cos\left(p \cdot 10000^{-\frac{2i}{D_p}}\right). \tag{3}$$

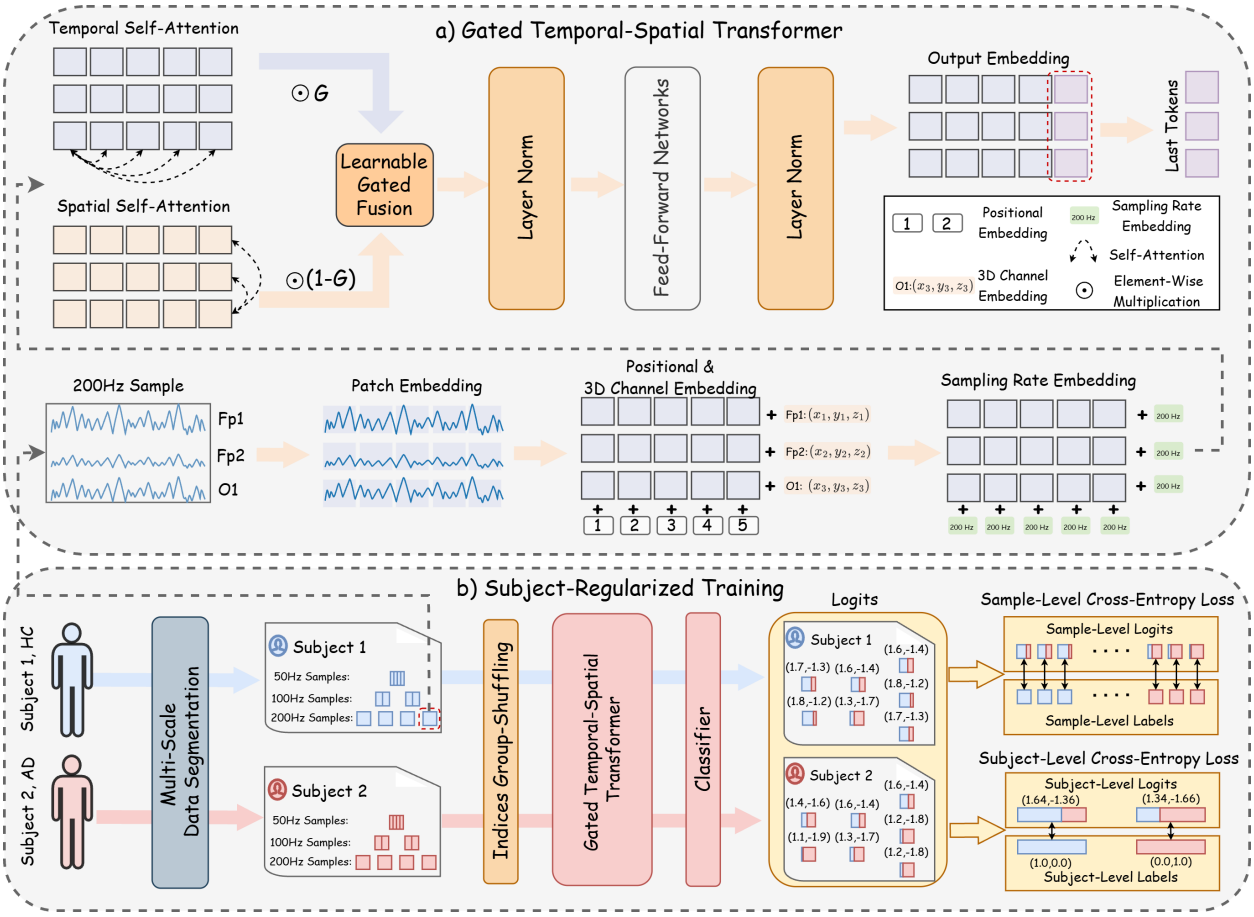

Figure 2: **a) Gated Temporal-Spatial Transformer.** Input EEG samples are first divided into univariate patches. These patches are projected into patch embeddings, to which temporal positional embeddings, 3D channel embeddings, and sampling rate embeddings are added to form the final token representations. A parallel temporal-spatial self-attention mechanism is then applied along both the temporal and channel dimensions. The resulting features from the two branches are adaptively combined through a learnable gated fusion module. Standard Transformer components, including layer norm and feedforward networks, are applied after the gated fusion stage. Finally, the last token of each channel is used as the downstream representation. **b) Subject-Regularized Training.** Each subject's EEG recordings are segmented into samples with multiple sampling rates (e.g., 200 Hz, 100 Hz, and 50 Hz) using the proposed multi-scale segmentation strategy. Index group shuffling ensures that each training batch contains sufficient samples from the same subject, thereby facilitating subject-level learning. The classifier produces sample-level logits, which are further aggregated into subject-level logits. Cross-entropy losses are then computed at both the sample and subject levels.

where $i = 0, 1, \ldots, \left\lfloor \frac{D_p}{2} \right\rfloor - 1$. We compute axis-wise embeddings for each channel and concatenate them to obtain the final 3D channel embedding:

$$\boldsymbol{e}_c = \left[\mathrm{CE}(x_c) \,\|\, \mathrm{CE}(y_c) \,\|\, \mathrm{CE}(z_c)\right] \in \mathbb{R}^D, \quad c = 1, \ldots, C \tag{4}$$

which yields the channel embedding matrix $\boldsymbol{CE} \in \mathbb{R}^{C \times D}$ by stacking $\{\boldsymbol{e}_c\}_{c=1}^{C}$. The channel embeddings are then broadcast and added to all N patch embeddings:

$$\boldsymbol{E} \leftarrow \boldsymbol{E} + \mathrm{Broadcast}\,(\boldsymbol{CE}), \quad \boldsymbol{E} \in \mathbb{R}^{C \times N \times D} \tag{5}$$

In practice, 3D coordinate information can be readily obtained given the electrode montage and channel names using the MNE Python package (Gramfort et al., 2013).

Sampling Rate Embedding. Since the sampling rate critically affects the temporal characteristics of EEG signals, even for recordings of identical duration, we introduce a sampling rate embedding module to explicitly encode the temporal scale of each input. This design enables the model to flexibly generalize across arbitrary sampling rates in both pre-training and downstream tasks, while naturally supporting the multi-scale segmentation strategy described in the previous paragraph. Given that each EEG sample is associated with a sampling rate label r, we define a learnable embedding lookup table that maps each sampling rate to a D-dimensional embedding vector $\boldsymbol{e}_r \in \mathbb{R}^D$. The sampling rate embedding is broadcast along both the channel and temporal dimensions and added to all $C \times N$ patch embeddings:

$$\boldsymbol{E} \leftarrow \boldsymbol{E} + \text{Broadcast}(\boldsymbol{e}_r), \quad \boldsymbol{E} \in \mathbb{R}^{C \times N \times D}. \tag{6}$$

Parallel Temporal-Spatial Self-Attention. Given the input representation $\boldsymbol{E} \in \mathbb{R}^{C \times N \times D}$, where C denotes the number of channels, N the number of temporal patch embeddings per channel, and D the embedding dimension, we apply temporal and spatial self-attention in parallel to jointly model temporal dynamics and spatial dependencies.

Temporal Self-Attention. We use $\boldsymbol{E}_c \in \mathbb{R}^{N \times D}$ to denote the sequence of patch embeddings for the c-th channel. Self-attention is applied independently to each channel along the temporal dimension:

$$\boldsymbol{E}_c \leftarrow \text{Attn}(\boldsymbol{E}_c, \boldsymbol{E}_c, \boldsymbol{E}_c) \in \mathbb{R}^{N \times D}, \quad c = 1, \ldots, C. \tag{7}$$

This allows the model to capture temporal dependencies within each channel c. After updating all channels, we stack the outputs $\boldsymbol{E}_c$ along channel dimension to obtain $\boldsymbol{E}^T \in \mathbb{R}^{C \times N \times D}$.

Spatial Self-Attention. We use $\boldsymbol{E}_n \in \mathbb{R}^{C \times D}$ to denote the set of channel embeddings at the n-th temporal patch. Spatial self-attention is applied independently across channels at each temporal patch index:

$$\boldsymbol{E}_n \leftarrow \text{Attn}(\boldsymbol{E}_n, \boldsymbol{E}_n, \boldsymbol{E}_n) \in \mathbb{R}^{C \times D}, \quad n = 1, \ldots, N. \tag{8}$$

This enables the model to capture spatial correlations among channels at each temporal patch index n. After updating all patch indices, we stack the outputs $\boldsymbol{E}_n$ along the temporal dimension to obtain $\boldsymbol{E}^S \in \mathbb{R}^{C \times N \times D}$.

Gated Fusion. In real-world applications, EEG data collected from different devices, protocols, and tasks may exhibit varying relative importance of temporal and spatial patterns. We therefore design a learnable gated fusion module that allows the backbone model to adaptively determine the contribution of temporal and spatial features based on their respective outputs. Specifically, an element-wise gate is computed as:

$$\boldsymbol{G} = \sigma\big(\big[\boldsymbol{E}^T \,\|\, \boldsymbol{E}^S\big]\boldsymbol{W}_f\big), \quad \boldsymbol{G} \in \mathbb{R}^{C \times N \times D}, \tag{9}$$

where $\sigma(\cdot)$ denotes the sigmoid function, $\|$ denotes concatenation, and $\boldsymbol{W}_f \in \mathbb{R}^{2D \times D}$ is a learnable projection matrix. The final representation is then obtained through element-wise gated fusion:

$$\boldsymbol{E} = \boldsymbol{G} \odot \boldsymbol{E}^T + (1 - \boldsymbol{G}) \odot \boldsymbol{E}^S, \quad \boldsymbol{E} \in \mathbb{R}^{C \times N \times D}, \tag{10}$$

where $\odot$ denotes element-wise multiplication. This adaptive fusion mechanism enables the model to automatically learn the relative importance of temporal and spatial features for different datasets and tasks.

Classifier. After M stacked blocks of self-attention, layernorm, and feedforward network modules, the model outputs an embedding $\boldsymbol{E} \in \mathbb{R}^{C \times N \times D}$. The downstream representation $\boldsymbol{E}_d \in \mathbb{R}^{C \times D}$ is obtained by selecting the **last token** along the temporal dimension for each channel from $\boldsymbol{E}$. This design is similar to using a per-channel `[CLS]` token, but avoids explicit token initialization and naturally generalizes to downstream data with an arbitrary number of channels. The representation $\boldsymbol{E}_d$ is then fed into a linear classifier $c(\cdot)$ to produce predictions.

3.3 Subject-Regularized Training

Majority Voting. Since the model is trained on segmented fixed-length EEG samples, it naturally produces sample-level predictions during inference. Therefore, an additional post-processing step is required to obtain

final subject-level detections. Following prior studies (Ieracitano et al., 2019), we adopt a majority voting strategy. Specifically, given all sample-level predictions belonging to the same subject s, the final subject-level detection is determined as the most frequently predicted class among those samples.

Subject-Level Cross-Entropy Loss. Previous studies have shown that high sample-level prediction accuracy does not necessarily translate into strong subject-level detection performance (Wang et al., 2024d), since subject-level predictions are typically computed post hoc (majority voting) and do not directly participate in the training process. To address this issue, we introduce a subject-level cross-entropy loss that explicitly encourages predictive consistency across samples belonging to the same subject. To distinguish it from the conventional sample-level cross-entropy loss $\mathcal{L}_{\text{ce}}^{\text{sam}}$, we denote the proposed subject-level loss as $\mathcal{L}_{\text{ce}}^{\text{sub}}$. The pseudocode for its computation is provided in Algorithm 1. For each unique subject s appearing in a training batch, we first compute a subject-level logit by averaging the sample-level logits over all samples associated with subject s. The loss $\mathcal{L}_{\text{ce}}^{\text{sub}}$ is then defined as the cross-entropy between these aggregated subject-level logits and the corresponding subject labels. The final training objective combines both sample-level and subject-level supervision:

$$\mathcal{L}_{ce} = \alpha \mathcal{L}_{\text{ce}}^{\text{sam}} + \beta \mathcal{L}_{\text{ce}}^{\text{sub}}, \quad \alpha + \beta = 1, \alpha, \beta \in [0, 1] \tag{11}$$

where α and β are balancing coefficients.

Algorithm 1 Subject-Level Cross-Entropy Loss

Require: Batch logits $\{\mathbf{z}_i\}_{i=1}^{B}$, labels $\{y_i\}_{i=1}^{B}$, subject IDs $\{s_i\}_{i=1}^{B}$ ▷ B is the batch size
Ensure: Subject-level Cross-Entropy Loss $\mathcal{L}_{\text{ce}}^{\text{sub}}$
$\mathcal{U} \leftarrow \text{Unique}(\{s_i\}_{i=1}^{B})$ ▷ Unique subject IDs in the current batch
Initialize subject-level logits list $\mathcal{Z} \leftarrow [\,]$
Initialize subject-level labels list $\mathcal{Y} \leftarrow [\,]$
for each subject $u \in \mathcal{U}$ **do**
  $\mathcal{M}_u \leftarrow \{i \mid s_i = u\}$ ▷ Indices of samples from subject u
  $\bar{\mathbf{z}}_u \leftarrow \frac{1}{|\mathcal{M}_u|} \sum_{i \in \mathcal{M}_u} \mathbf{z}_i$ ▷ Average subject-level logit
  $y_u \leftarrow y_{i^*}$ for any $i^* \in \mathcal{M}_u$ ▷ All samples share the same label
  $\mathcal{Z} \leftarrow \mathcal{Z} \cup \{\bar{\mathbf{z}}_u\}$
  $\mathcal{Y} \leftarrow \mathcal{Y} \cup \{y_u\}$
end for
return CrossEntropy$(\mathcal{Z}, \mathcal{Y})$

Index Group Shuffling. In practical training with a large number of subjects, the probability that multiple samples from the same subject appear within the batch decreases, reducing the effectiveness of subject-level training. To mitigate this issue, we introduce an index group shuffling algorithm as an engineering trick that dynamically reorders indices for each epoch. The pseudocode is provided in the Appendix in Algorithm 2. The computational cost of this algorithm is negligible, as it requires only sorting and shuffling. The algorithm ensures that, for every subject included in a batch, at least a predefined group size of samples from that subject is present, while maintaining enough randomness.

3.4 Domain-Inspired Self-Supervised Pre-training

We employ both sample-level and subject-level medical contrastive learning for self-supervised pre-training (Wang et al., 2023). The sample-level contrastive learning module performs instance discrimination, treating different augmented views of the same sample as positive pairs and views from other samples as negative pairs. The augmentation methods are described in Appendix B. The subject-level contrastive learning module leverages subject IDs as guidance: samples from the same subject are treated as positive pairs, while those from different subjects are treated as negative pairs. Compared with masked autoencoding (He et al., 2022), subject-level contrastive learning is more suitable for our goal of **capturing subject-level representations** for AD detection, under the assumption that all samples from a subject collected in a short period share the same label. We again incorporate the index group shuffling algorithm (Algorithm 2) to maintain the probability that samples with the same subject ID co-occur within a mini-batch. The overall self-supervised pre-training objective is a weighted sum of the two contrastive module losses:

$$\mathcal{L}_{co} = \lambda_1 \mathcal{L}_{co}^{sam} + \lambda_2 \mathcal{L}_{co}^{sub}, \quad \lambda_1 + \lambda_2 = 1, \lambda_1, \lambda_2 \in [0, 1] \tag{12}$$

where λ_1 and λ_2 are balancing coefficients.

4 Experiments

Table 2: **Dataset Statistics.** The **h** and **m** in the total time denote hours and minutes. Abbreviations: **ASSR**: Auditory Steady-State Response; **RS**: Resting State; **HV**: Hyperventilation; **PS**: Photic Stimulation. **HC** - Healthy Controls; **AD**: Alzheimer's Disease; **FTD**: Frontotemporal Dementia; **PD**: Parkinson's disease; **MS**: Multiple Sclerosis; **MCI**: Mild Cognitive Impairment; **SCZ**: Schizophrenia; **DEP**: Depression; **DEM**: Dementia; **FEP**: First Episode Psychosis; **MDD**: Major Depressive Disorder; **ADHD**: Attention Deficit Hyperactivity Disorder; **SMC**: Subjective Memory Complaints; **OCD**: Obsessive-Compulsive Disorder.

Category	Datasets	Paradigm	#Subjects	Sampling Rate	#Channels	Total Time	#Timestamps	#Samples	Classes	Tasks
Non-AD	**BACA-RS**	RS	608	1000–>200,100,50 Hz	65–>19	169.80h	400	1,057,775	HC	Pre-train
	Depression	RS	122	500–>200,100,50 Hz	66–>19	23.36h	400	146,062	HC, DEP	Pre-train
	FEPCR	RS	143	1000–>200,100,50 Hz	64–>19	12.27h	400	76,694	HC, FEP	Pre-train
	MCEF-RS	RS	165	512–>200,100,50 Hz	64–>19	14.14h	400	88,411	HC	Pre-train
	PD-RS	RS	149	500–>200,100,50 Hz	60–>19	6.64h	400	41,174	HC, PD	Pre-train
	PEARL-Neuro	RS	79	1000–>200,100,50 Hz	127–>19	14.36h	400	90,147	HC	Pre-train
	SRM-RS	RS	109	1024–>200,100,50 Hz	64–>19	9.10h	400	56,880	HC	Pre-train
	TDBrain	RS	1273	500–>200,100,50 Hz	33–>19	89.71h	400	555,455	MDD, ADHD, SMC, OCD, etc.	Pre-train
	TUEP	RS	200	256–>200,100,50 Hz	21, etc.–>19	466.24h	400	2,930,241	HC, Epilepsy	Pre-train
AD	**AD-Auditory**	ASSR	35	250–>200,100,50 Hz	19	5.20h	400	32,655	HC, AD, MCI,	Pre-train
	BrainLat	RS	135	512–>200,100,50 Hz	19	17.26h	400	108,206	HC, AD, FTD, PD, MS	Pre-train
	P-ADIC	RS	249	500–>200,100,50 Hz	19	75.39h	400	473,970	HC, AD, MCI, SCZ, DEP	Pre-train
	CAUEEG	RS, PS, HV, etc.	1379	200–>200,100,50 Hz	19	282.37h	400	1,773,814	HC, MCI, DEM, etc.	Pre-train
	ADFSU	RS	92	128–>100,50 Hz	19	24.53m	100	4048	HC, AD	Fine-tune
	ADFTD	RS, PS	88	500–>200,100,50 Hz	19	26.64h	400	167,083	HC, AD, FTD	Fine-tune
	ADSZ	RS	48	128–>100,50 Hz	19	6.80m	100	1128	HC, AD	Fine-tune
	APAVA	RS	23	256–>200,100,50 Hz	16	0.92h	200	9282	HC, AD	Fine-tune
	CNBPM	RS	189	256–>200,100,50 Hz	19	19.51h	400	122,029	HC, MCI, AD	Fine-tune

4.1 Setup

Datasets Curation. We categorize EEG datasets utilized in this paper into two groups: those containing data from subjects with Alzheimer's Disease (**AD** datasets) and those without (**Non-AD** datasets). We conducted a comprehensive review of relevant publications and public EEG repositories (e.g., OpenNeuro, Dryad, figshare) published before 2026. This effort identified 8 public AD datasets. In total, we utilize 8 public and 1 private AD datasets, most of which consist of resting-state recordings: **AD-Auditory** (Lahijanian et al., 2024), **BrainLat** (Prado et al., 2023), **P-ADIC** (Shor et al., 2021), **CAUEEG** (Kim et al., 2023), **ADFSU** (Vicchietti et al., 2023), **ADFTD** (Miltiadous et al., 2023b), **ADSZ** (Alves et al., 2022), **APAVA** (Escudero et al., 2006), and **CNBPM** (Amezquita-Sanchez et al., 2019). Together, these datasets comprise **2,238 subjects and 427.81 hours** of recordings, forming the **world's largest** EEG-AD corpus reported so far. We also include 9 non-AD domain-relevant resting-state datasets that include recordings from both healthy subjects and patients with other neurological conditions (e.g., Parkinson's disease, depression, ADHD). These datasets are: **BACA-RS** (Getzmann et al., 2024), **Depression** (Cavanagh et al., 2019), **FEPCR** (Phalen et al., 2020), **MCEF-RS** (Chenot et al., 2024), **PD-RS** (Singh et al., 2023), **PEARL-Neuro** (Dzianok & Kublik, 2024), **SRM-RS** (Hatlestad-Hall et al., 2022), **TDBrain** (Van Dijk et al., 2022), **TUEP** (Veloso et al., 2017), account for total **2848 subjects, 805.62 hours**. More details on dataset curation are presented in the Appendix D.1.

Data Preprocessing. The preprocessing pipeline is presented in Appendix D.2, and the statistics of the processed datasets are summarized in Table 2. For pre-training, we leverage all 9 Non-AD datasets together with 4 AD datasets, resulting in a total of **4,646 subjects, 1,185.84 hours, and 7,431,484** samples (representing 2s, 4s, and 8s segments at 200Hz, 100Hz, and 50Hz sampling rates). The remaining 5 AD datasets are reserved for downstream evaluation, resulting in a total of **440 subjects, 47.59 hours, and 303,570** samples. Table 2 shows the processed dataset statistics.

Baselines. We compare our method against **16** baselines, including 1 manual-feature approach, 11 supervised deep learning methods, and 4 EEG foundation models. These selected baselines are SOTA methods or have shown strong performance in EEG classification tasks. The feature-based method extracts **Statistical**, **Spectral**, **Power**, and **Complexity** features (see Appendix C) commonly used for EEG-based AD detection, followed by classification with a linear classifier. The 11 supervised learning methods include **EEGConformer** (Song et al., 2022), **EEGInception** (Zhang et al., 2021), **EEGNet** (Lawhern et al., 2018), **iTransformer** (Liu et al., 2023b), **MedGNN** (Fan et al., 2025), **Medformer** (Wang et al., 2024b), **MNet** (Watanabe et al., 2024), **ModernTCN** (Luo & Wang, 2024), **PatchTST** (Nie, 2023), **TCN** (Bai et al., 2018), **TimesNet** (Wu et al., 2023). The 4 EEG foundation models are **BIOT** (Yang et al., 2024), **LaBraM** (Jiang et al., 2024b), **CBraMod** (Wang et al., 2025), **CSBrain** (Zhou et al., 2025).

Training & Parameter Settings. Self-supervised pre-training is conducted for 30 epochs without early stopping, followed by fine-tuning for up to 200 epochs with early stopping (patience=15) based on the best sample-level F1. Batch sizes are 2048 for pre-training and 512 for supervised learning. We use AdamW with learning rates of 2×10^{-4} (pre-training) and 1×10^{-4} (fine-tuning), scheduled by CosineAnnealingLR. We employ 14 evaluation metrics, including sample-level accuracy, precision (macro-averaged), sensitivity (macro-averaged), specificity (macro-averaged), F1 score (macro-averaged), AUROC (macro-averaged), and AUPRC (macro-averaged), and their corresponding subject-level metrics computed via majority voting (Para. 3.3). The contrastive loss coefficients in self-supervised pre-training are set to $\lambda_1 = 0.25$ and $\lambda_2 = 0.75$. Both sample-level and subject-level cross-entropy losses are set to $\alpha = \beta = 0.5$ in the fine-tuning stage. The *group_size* for index group shuffling is set to 32 and 8 for the pre-training and fine-tuning stages, respectively. For self-supervised pre-training, all subjects from the pre-training datasets are used. For supervised learning and fine-tuning, we adopt Monte Carlo cross-validation (Xu & Liang, 2001) with a subject-independent (Wang et al., 2024b) 8:1:1 train/validation/test split, ensuring no subject overlap while preserving randomness across seeds. To enable cross-domain transfer, no pre-training subjects or datasets are used in fine-tuning. For 4 foundation model baselines, we fine-tune their released checkpoints. Each evaluation is repeated with five random seeds (41-45), reporting the mean and standard deviation. Experiments are run on 8 NVIDIA RTX A6000 GPUs with Python 3.10 and PyTorch 2.5.1+cu121. More details for each method are in Appendix C.

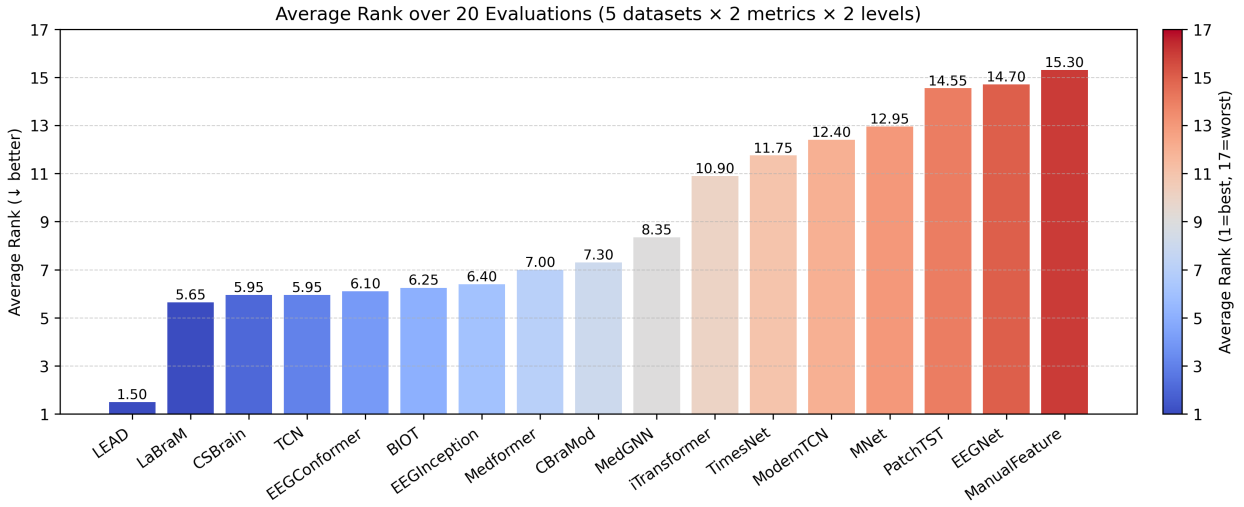

Figure 3: **Average Performance Rank.** Average performance rank of 17 methods across all 20 evaluations. For example, the value 1.5 of LEAD indicates an average rank of 1.5. **Lower ranks and a deeper blue indicate better performance**.

4.2 Results

Performance Comparison with Baselines. We compare LEAD with 16 baseline methods on 5 downstream datasets. The F1 score and AUROC results are reported in Table 3, while the full results for all 7 evaluation metrics are provided in Appendix E.1. Our method achieves top-1 performance in 17 out of 20 evaluations.

Table 3: **Method Comparison Results.** F1 score and AUROC comparison results of 16 baselines and our method. The full results for all 7 metrics are available in Appendix E.1. The **Top-1**, Top-2, and Top-3 results are highlighted in red, blue, and green.

Datasets	ADFSU *(4,048 Samples) (92 Subjects) (HC vs AD)*		ADFTD *(167,083 Samples) (88 Subjects) (HC vs AD vs FTD)*		ADSZ *(1,128 Samples) (48 Subjects) (HC vs AD)*		APAVA *(9,282 Samples) (23 Subjects) (HC vs AD)*		CNBPM *(122,029 Samples) (189 Subjects) (HC vs MCI vs AD)*	
Methods \ Metrics	F1 Score	AUROC	F1 Score	AUROC	F1 Score	AUROC	F1 Score	AUROC	F1 Score	AUROC
Sample-Level Classification										
ManualFeature	75.67±11.41	80.90±8.87	44.73±3.88	62.89±4.31	58.42±21.26	67.20±15.34	63.22±6.51	66.49±5.78	43.00±6.36	61.29±6.44
EEGConformer	94.22±3.28	98.87±0.95	58.59±3.84	80.61±3.90	94.73±3.47	98.50±1.48	73.05±4.93	83.07±3.70	59.17±8.85	81.76±8.20
EEGInception	90.88±2.13	98.73±0.75	62.69±4.06	84.01±2.76	92.60±4.44	98.06±1.66	67.84±4.60	78.04±4.19	58.56±8.02	81.73±8.18
EEGNet	74.48±9.91	85.88±7.95	40.31±6.74	60.93±4.42	64.90±12.88	77.18±19.10	67.84±4.60	78.04±4.19	38.57±6.52	63.51±7.95
iTransformer	76.80±2.31	88.16±1.86	54.33±3.67	73.98±4.02	73.13±4.75	81.86±4.90	72.13±2.30	85.53±1.00	51.37±8.01	74.99±10.16
MedGNN	90.09±2.37	98.06±1.42	67.67±4.60	85.52±2.57	88.59±8.01	95.90±4.10	67.54±7.95	81.39±3.83	58.68±7.50	82.13±7.35
Medformer	87.15±1.91	97.48±1.32	64.83±6.72	84.13±3.27	88.73±5.76	94.18±3.72	67.42±2.44	78.38±2.75	**60.39±7.68**	82.17±7.87
MNet	76.19±8.72	92.75±5.08	59.91±9.77	82.30±5.04	79.21±7.35	88.87±5.92	52.98±8.72	79.85±3.17	47.04±9.76	77.03±11.79
ModernTCN	75.09±3.18	91.94±1.84	58.52±3.71	77.25±3.33	75.73±4.05	84.42±4.30	58.42±0.76	75.88±2.41	55.94±8.10	77.85±9.30
PatchTST	76.03±4.27	89.12±1.95	50.95±4.04	71.30±4.75	71.05±10.02	79.10±12.54	52.39±2.65	73.76±3.11	49.78±6.10	71.52±9.58
TCN	90.10±3.13	98.82±0.73	64.26±2.40	82.94±1.89	88.93±9.23	94.02±6.85	74.81±4.47	85.84±3.92	58.71±5.02	82.01±6.02
TimesNet	82.08±6.29	95.21±2.45	58.66±5.68	79.18±3.39	83.80±3.60	88.12±9.12	55.95±6.89	56.01±4.93	56.19±6.63	79.18±7.74
BIOT	88.94±2.69	96.78±3.04	69.79±5.90	86.85±4.77	90.83±4.72	96.73±2.98	79.55±5.36	**92.92±2.77**	54.79±9.52	77.95±11.46
LaBraM	92.24±3.38	98.90±0.53	75.64±4.68	91.22±2.72	91.12±4.50	97.54±1.88	71.65±3.35	85.61±1.20	52.66±7.11	78.44±9.59
CBraMod	92.32±4.27	98.47±1.49	68.33±4.53	86.95±2.89	83.70±9.28	96.67±3.69	74.10±2.40	83.10±1.85	53.93±6.87	78.68±9.10
CSBrain	91.11±2.34	98.23±0.51	69.39±2.63	86.82±1.28	93.16±4.03	98.87±1.14	61.69±8.63	68.20±1.53	60.30±5.41	**83.39±5.73**
LEAD	**97.06±1.02**	**99.92±0.07**	**81.01±5.02**	**94.03±1.33**	**97.42±2.74**	**100.00±0.00**	**83.70±5.30**	92.51±4.33	60.20±6.83	79.10±10.21
Subject-Level Detection										
ManualFeature	81.05±20.29	80.00±18.71	54.26±7.80	69.80±2.94	63.05±30.68	70.00±24.49	57.33±19.60	65.00±12.25	43.95±7.64	61.43±4.16
EEGConformer	100.00±0.00	100.00±0.00	68.45±15.21	79.13±8.16	100.00±0.00	100.00±0.00	65.33±16.00	70.00±10.00	58.42±8.60	70.00±5.80
EEGInception	100.00±0.00	100.00±0.00	79.47±17.70	86.23±10.23	100.00±0.00	100.00±0.00	70.67±21.33	75.00±15.81	59.77±8.06	70.71±5.71
EEGNet	78.37±21.07	81.25±18.54	54.82±14.34	70.60±8.38	61.38±22.11	66.67±18.26	70.67±21.33	75.00±15.81	34.02±11.43	55.00±7.35
iTransformer	66.01±17.61	65.00±12.25	67.17±9.15	78.21±5.36	100.00±0.00	100.00±0.00	84.00±13.06	85.00±12.25	52.85±8.58	67.14±5.25
MedGNN	100.00±0.00	100.00±0.00	75.23±4.81	83.41±3.30	96.57±6.86	96.67±6.67	65.33±16.00	70.00±10.00	52.18±11.32	65.71±7.69
Medformer	100.00±0.00	100.00±0.00	78.98±12.08	84.84±8.41	100.00±0.00	100.00±0.00	73.33±0.00	75.00±0.00	60.93±6.49	72.14±4.16
MNet	58.82±17.61	60.00±12.25	68.63±17.53	80.08±9.49	89.90±13.38	90.00±13.33	41.33±16.00	55.00±10.00	39.97±4.38	62.14±4.29
ModernTCN	58.82±17.61	60.00±12.25	74.39±11.35	82.82±6.50	89.71±8.40	90.00±8.16	33.33±0.00	50.00±0.00	61.96±9.90	72.86±5.35
PatchTST	77.12±18.02	75.00±15.81	65.04±14.32	77.26±7.62	86.29±19.99	86.67±19.44	33.33±0.00	50.00±0.00	51.35±9.07	67.14±4.16
TCN	96.08±7.84	95.00±10.00	82.80±6.10	87.38±4.32	96.57±6.86	96.67±6.67	73.33±0.00	75.00±0.00	62.66±3.41	72.86±2.86
TimesNet	84.97±21.64	85.00±20.00	64.23±17.73	78.57±9.45	96.57±6.86	96.67±6.67	33.33±0.00	50.00±0.00	55.52±4.93	67.86±3.91
BIOT	100.00±0.00	100.00±0.00	88.10±9.72	90.71±7.61	90.83±4.72	96.73±2.98	95.00±10.00	94.67±10.67	55.74±8.58	67.86±5.98
LaBraM	100.00±0.00	100.00±0.00	91.14±8.64	93.77±6.16	100.00±0.00	100.00±0.00	70.67±21.33	75.00±15.81	50.11±6.53	65.00±3.50
CBraMod	100.00±0.00	100.00±0.00	82.21±6.30	87.10±3.77	89.07±14.85	90.00±13.33	73.33±0.00	75.00±0.00	50.60±10.89	65.71±7.00
CSBrain	100.00±0.00	100.00±0.00	78.33±7.52	84.56±4.61	100.00±0.00	100.00±0.00	57.33±19.60	65.00±12.25	62.41±5.80	72.86±3.64
LEAD	**100.00±0.00**	**100.00±0.00**	**93.95±4.95**	**95.36±3.85**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**65.94±6.66**	**75.00±5.05**

Notably, LEAD consistently outperforms all competing approaches across subject-level F1 score and AUROC, demonstrating strong potential for real-world clinical detection. Furthermore, our method surpasses existing EEG foundation models despite being pre-trained on only 1,185.84 hours of EEG data, which is substantially less than the approximately 2,500 hours used by LaBraM and more than 9,000 hours used by CBraMod and CSBrain. These results highlight the effectiveness of the proposed backbone, subject-level training strategy, and domain-inspired pre-training paradigm, as well as the careful selection of pre-training data.

Table 4: **EEG Paradigms Comparison.** Each ADFTD subject was recorded under two paradigms: resting-state eyes-closed and photic-stimulation eyes-open, with varying recording lengths. We evaluate model performance on each paradigm separately and both.

Datasets	Paradigms	Accuracy	Precision	Sensitivity	Specificity	F1 Score	AUROC	AUPRC
ADFTD	**Resting-State (19.40h)**	55.90±4.43	53.22±6.63	52.29±5.21	76.91±2.39	50.67±5.83	74.02±6.19	59.78±8.63
(Sample-Level)	**Photic-Stimulation (7.25h)**	58.23±9.57	58.13±12.16	56.24±10.83	78.83±4.93	55.60±11.70	72.08±11.15	60.17±13.52
(HC vs AD vs FTD)	**Both (26.64h)**	**81.70±4.74**	**82.93±4.40**	**80.78±4.95**	**90.52±2.45**	**81.01±5.02**	**94.03±1.33**	**90.37±2.14**
ADFTD	**Resting-State (19.40h)**	64.00±10.20	56.14±20.90	61.11±11.52	80.95±5.48	55.49±15.08	71.03±8.48	54.29±10.49
(Subject-Level)	**Photic-Stimulation (7.25h)**	60.00±10.95	61.67±18.05	60.56±11.84	79.68±5.80	58.58±14.37	70.12±8.80	54.32±9.98
(HC vs AD vs FTD)	**Both (26.64h)**	**94.00±4.90**	**95.67±3.59**	**93.89±5.09**	**96.83±2.61**	**93.95±4.95**	**95.36±3.85**	**91.56±6.91**

EEG Paradigms Comparison. We evaluate the ADFTD dataset across three paradigms: resting-state, photic-stimulation, and both. Results are shown in Table 4. Using both paradigms substantially outperforms the individual paradigms, achieving a subject-level F1 of 93.95% versus 55.58% (resting-state) and 58.58% (photic-stimulation). While it remains unclear whether this gain arises from longer recordings or complementary

paradigm features, the evidence indicates that *combining these two paradigms is non-detrimental and likely to be beneficial for AD detection.*

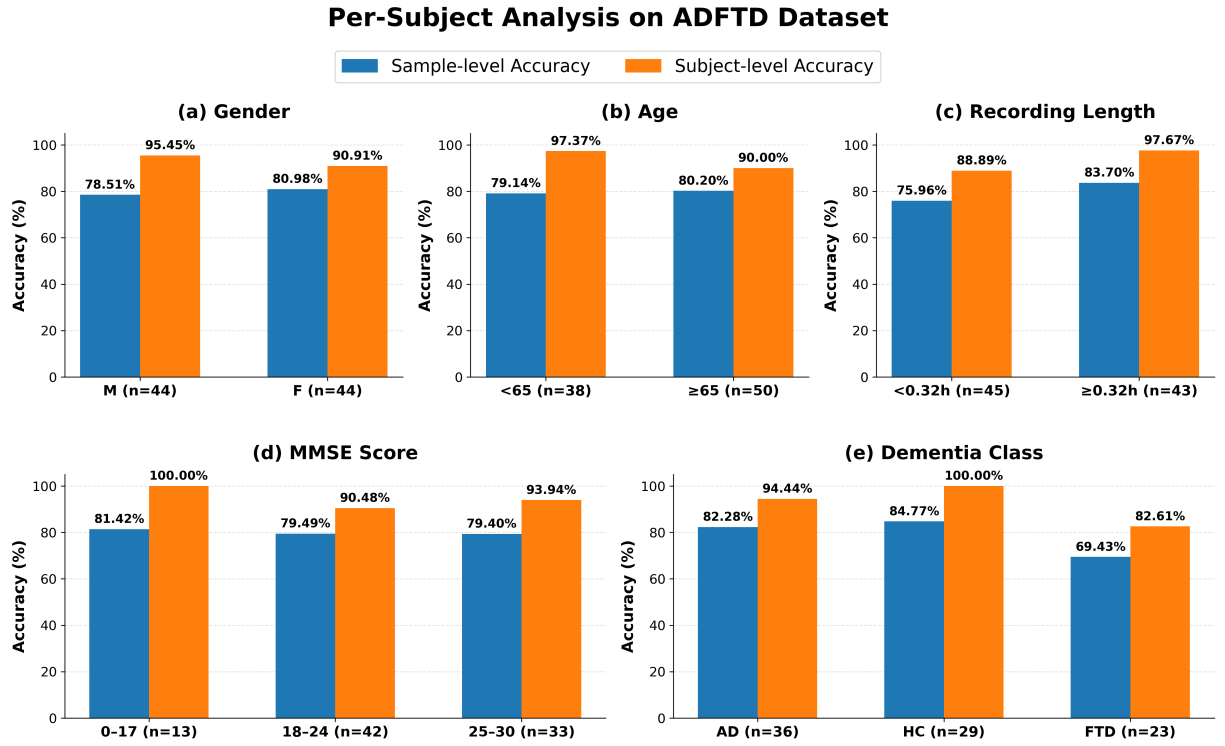

Figure 4: **Grouped Bar Plot of Per-Subject Analysis.** Sample-level and subject-level accuracy are reported for subgroups defined by gender, age, recording length, MMSE score, and dementia class.

Table 5: **Per-Subject Analysis.** We report per-subject results on the ADFTD dataset under the LOSO setting and analyze the factors influencing model performance, including demographic characteristics, recording length, MMSE score, and dementia category.

Label	Gender		Age		Length		MMSE			Dementia Class		
Group	M	F	<65	≥65	<0.32h	>=0.32h	0-17	18-24	25-30	AD	HC	FTD
Subject Number	44	44	38	50	45	43	13	42	33	36	29	23
Sample-level Accuracy	78.51%	**80.98%**	79.14%	**80.20%**	75.96%	**83.70%**	**81.42%**	79.49%	79.40%	82.28%	**84.77%**	69.43%
Subject-level Accuracy	**95.45%**	90.91%	**97.37%**	90.00%	88.89%	**97.67%**	**100.00%**	90.48%	93.94%	94.44%	**100.00%**	82.61%

Per-Subject Analysis. As ADFTD is a public dataset with detailed demographic information, we further explore per-subject performance across demographics, recording length, MMSE score, and clinical subgroups under the leave-one-subject-out (LOSO) setting. The full results of each subject are in Appendix E.2. A comparison among factors is in Table 5 and plotted in Figure 4. According to the results, longer recordings (>=0.32 h) consistently yield higher accuracy at both the sample and subject levels, highlighting the importance of sufficient EEG recording length. Subjects younger than 65 years show higher subject-level accuracy than older subjects, suggesting increased variability in elderly EEG recordings. Performance also varies with cognitive status: subjects with lower MMSE scores (0-17) are classified most reliably, whereas intermediate MMSE ranges (18-24) exhibit reduced accuracy. In terms of dementia category, the model achieves perfect subject-level accuracy in HC, strong performance in AD, and noticeably lower accuracy in FTD, indicating that FTD remains the most challenging to distinguish.

Ablation Study We conduct an ablation study on the ADFTD dataset to evaluate the contribution of each module in our proposed method. The results are reported in Table 6. Overall, removing any individual component results in a degradation of more than 1% in the F1 score at both the sample- and subject-levels, indicating that all modules contribute positively to the final performance. The most significant

Table 6: **Ablation Study.** We investigate the effectiveness of individual modules and design choices in our method through systematic ablation, in which each component is removed separately.

Datasets	Paradigms	Accuracy	Precision	Sensitivity	Specificity	F1 Score	AUROC	AUPRC
ADFTD *(Sample-Level)* *(167,083 Samples)* *(HC vs AD vs FTD)*	**No Pre-training**	76.02±4.61	76.96±3.96	74.58±4.96	87.61±2.46	74.63±4.96	90.16±2.68	83.80±3.60
	No Index Group Shuffling	80.49±4.87	81.33±4.65	79.65±5.06	89.95±2.51	79.73±5.11	93.51±1.65	89.48±2.46
	No Subject Cross-Entropy Loss	80.59±4.87	81.61±4.38	79.76±5.21	89.97±2.55	79.92±5.18	93.59±1.57	89.61±2.51
	No Sampling Rate Embedding	76.73±4.51	78.37±4.53	75.46±4.35	87.93±2.24	75.84±4.49	90.96±1.76	85.66±2.69
	No Multi-Sampling Segmentation	80.54±5.08	81.21±4.69	79.78±5.43	89.98±2.67	79.87±5.34	93.32±1.84	89.24±2.84
	LEAD (Full)	**81.70**±4.74	**82.93**±4.40	**80.78**±4.95	**90.52**±2.45	**81.01**±5.02	**94.03**±1.33	**90.37**±2.14
ADFTD *(Subject-Level)* *(88 Subjects)* *(HC vs AD vs FTD)*	**No Pre-training**	86.00±8.00	89.22±6.97	85.56±8.13	92.54±4.22	85.42±8.45	89.05±6.17	81.02±10.64
	No Index Group Shuffling	92.00±4.00	94.00±3.09	92.22±4.08	95.87±2.09	92.04±4.00	94.05±3.08	88.89±5.58
	No Subject Cross-Entropy Loss	92.00±4.00	94.00±3.09	92.22±4.08	95.87±2.09	92.04±4.00	94.05±3.08	88.89±5.58
	No Sampling Rate Embedding	86.00±8.00	88.44±7.18	85.56±8.13	92.70±4.09	85.06±8.85	89.13±6.10	80.24±11.29
	No Multi-Sampling Segmentation	92.00±4.00	94.00±3.09	92.22±4.08	95.87±2.09	92.04±4.00	94.05±3.08	88.89±5.58
	LEAD (Full)	**94.00**±4.90	**95.67**±3.59	**93.89**±5.09	**96.83**±2.61	**93.95**±4.95	**95.36**±3.85	**91.56**±6.91

performance drops are observed when pre-training is removed and when the sampling rate embedding is excluded. Specifically, disabling pre-training or sampling-rate embedding results in approximately a 5% decrease in sample-level F1 score and an 8% decrease in subject-level F1 score. The degradation caused by removing contrastive pre-training highlights the importance of the proposed domain-inspired pre-training strategy. In addition, the performance drop observed when removing the sampling rate embedding and the multi-sampling segmentation strategy indicates that using sampling rate embedding with a single sampling rate leads to only a minor performance loss, whereas training on multi-sampling data without sampling rate embedding causes a substantial degradation. This suggests that the model struggles to distinguish samples with different sampling rates when no explicit sampling rate information is provided. Other components, including index group shuffling and the subject-level cross-entropy loss, also consistently improve performance at both the sample and subject levels. Overall, the ablation results demonstrate the effectiveness of the proposed design choices and confirm that our method does not contain redundant modules.

5 Conclusion and Limitations

Conclusion. This paper introduces LEAD, the first foundation model for EEG-based Alzheimer's disease detection, trained on the world's largest EEG AD corpus. We design a gated temporal spatial Transformer that adapts to EEG recordings with arbitrary lengths, channel configurations, and sampling rates. We further propose a subject-regularized training strategy, including subject-level cross-entropy, index-group shuffling, and multi-scale segmentation, to enhance subject-level feature learning. Medical contrastive pre-training is conducted on 13 datasets, including 4 AD datasets and 9 non-AD neurological disorder datasets. Extensive experiments on 5 heterogeneous AD downstream datasets demonstrate that LEAD achieves the best overall ranking compared with 16 baseline methods across 20 evaluations. These results validate the effectiveness of our proposed backbone transformer, training strategies, proper pre-training corpus, and pre-training methods. We hope this work will advance research on EEG-based AD detection and support researchers who face challenges training models from scratch due to the limited number of AD subjects, as collecting EEG data from AD patients is costly and time-consuming.

Limitation. Although the performance of classifying AD vs HC and AD vs other dementias, such as FTD, is relatively strong, the performance for dementia stage classification that separates AD, MCI, and HC remains limited, with an F1 score of approximately 66%. This highlights the continued challenge of AD stage detection. AD is a progressive neurodegenerative disease. Healthy older adults may transition into MCI, and some MCI cases eventually progress to AD. However, not all MCI cases advance; some remain stable (SMCI) or even revert to normal cognition, while others progress (PMCI) (Vecchio et al., 2018; Ge et al., 2025). Critically, many PMCI patients may already exhibit AD-related features at the time of data acquisition, although these are not identifiable without longitudinal follow-up, while some SMCI patients are closer to HC. This ambiguity makes certain MCI subjects difficult to distinguish from either AD or HC. Future work should therefore prioritize the collection of longitudinal EEG data to track disease progression.

References

Biosemi. URL https://www.biosemi.com/index.htm. Accessed: 2025-05-04.

Salahuddin Ahmed, Marzia Momin, Jiashu Ren, Hyunjin Lee, Basma AlMahmood, Li-Pang Huang, Archana Pandiyan, Loganathan Veeramuthu, Chi-Ching Kuo, and Tao Zhou. Stick-and-play bioadhesive hairlike electrodes for chronic eeg recording on human. *npj Biomedical Innovations*, 2(1):9, 2025.

Ali H Husseen Al-Nuaimi, Emmanuel Jammeh, Lingfen Sun, and Emmanuel Ifeachor. Complexity measures for quantifying changes in electroencephalogram in alzheimer's disease. *Complexity*, 2018(1):8915079, 2018.

Caroline L Alves, Aruane M Pineda, Kirstin Roster, Christiane Thielemann, and Francisco A Rodrigues. Eeg functional connectivity and deep learning for automatic diagnosis of brain disorders: Alzheimer's disease and schizophrenia. *Journal of Physics: complexity*, 3(2):025001, 2022.

Juan P Amezquita-Sanchez, Nadia Mammone, Francesco C Morabito, Silvia Marino, and Hojjat Adeli. A novel methodology for automated differential diagnosis of mild cognitive impairment and the alzheimer's disease using eeg signals. *Journal of neuroscience methods*, 322:88–95, 2019.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15619–15629, 2023.

Marcos Aviles, Luz María Sánchez-Reyes, José Manuel Álvarez-Alvarado, and Juvenal Rodríguez-Reséndiz. Machine and deep learning trends in eeg-based detection and diagnosis of alzheimer's disease: A systematic review. *Eng*, 5(3):1464–1484, 2024.

Hamed Azami, Steven E Arnold, Saeid Sanei, Zhuoqing Chang, Guillermo Sapiro, Javier Escudero, and Anoopum S Gupta. Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases. *IEEE Access*, 7:68718–68733, 2019.

Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, 33: 12449–12460, 2020.

Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

Zeinab Breijyeh and Rafik Karaman. Comprehensive review on alzheimer's disease: Causes and treatment. *Molecules*, 25(24):5789, 2020.

Raymundo Cassani, Tiago H Falk, Francisco J Fraga, Paulo AM Kanda, and Renato Anghinah. The effects of automated artifact removal algorithms on electroencephalography-based alzheimer's disease diagnosis. *Frontiers in aging neuroscience*, 6:55, 2014.

James F Cavanagh, Andrew W Bismark, Michael J Frank, and John JB Allen. Multiple dissociations between comorbid depression and anxiety on reward and punishment processing: Evidence from computationally informed eeg. *Computational Psychiatry (Cambridge, Mass.)*, 3:1, 2019.

Quentin Chenot, Caroline Hamery, Moritz Truninger, Nicolas Langer, Sébastien Scannella, et al. Investigating the relationship between resting-state eeg microstates and executive functions: A null finding. *cortex*, 178: 1–17, 2024.

LW Chu. Alzheimer's disease: early diagnosis and treatment. *Hong Kong Medical Journal*, 18(3):228, 2012.

Carmina Coronel, Heinrich Garn, Markus Waser, Manfred Deistler, Thomas Benke, Peter Dal-Bianco, Gerhard Ransmayr, Stephan Seiler, Dieter Grossegger, and Reinhold Schmidt. Quantitative eeg markers of entropy and auto mutual information in relation to mmse scores of probable alzheimer's disease patients. *Entropy*, 19(3):130, 2017.

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.

Ozlem Karabiber Cura, Gulce C Yilmaz, H Sabiha Ture, and Aydin Akan. Deep time-frequency feature extraction for alzheimer's dementia eeg classification. In *2022 Medical Technologies Congress (TIPTEKNO)*, pp. 1–4. IEEE, 2022.

Berkay Döner, Thorir Mar Ingolfsson, Luca Benini, and Yawei Li. Luna: Efficient and topology-agnostic foundation model for eeg signal analysis. In *1st ICML Workshop on Foundation Models for Structured Data*.

Patrycja Dzianok and Ewa Kublik. Pearl-neuro database: Eeg, fmri, health and lifestyle data of middle-aged people at risk of dementia. *Scientific Data*, 11(1):276, 2024.

J Escudero, Daniel Abásolo, Roberto Hornero, Pedro Espino, and Miguel López. Analysis of electroencephalograms in alzheimer's disease patients with multiscale entropy. *Physiological measurement*, 27(11):1091, 2006.

Golshan Fahimi, Seyed Mahmoud Tabatabaei, Elnaz Fahimi, and Hamid Rajebi. Index of theta/alpha ratio of the quantitative electroencephalogram in alzheimer's disease: a case-control study. *Acta Medica Iranica*, pp. 502–506, 2017.

Wei Fan, Jingru Fei, Dingyu Guo, Kun Yi, Xiaozhuang Song, Haolong Xiang, Hangting Ye, and Min Li. Towards multi-resolution spatiotemporal graph learning for medical time series classification. In *Proceedings of the ACM on Web Conference 2025*, pp. 5054–5064, 2025.

Zitao Fang, Chenxuan Li, Hongting Zhou, Shuyang Yu, Guodong Du, Ashwaq Qasem, Yang Lu, Jing Li, Junsong Zhang, and Sim Kuan Goh. Neuript: Foundation model for neural interfaces. *39th Conference on Neural Information Processing Systems*, 2025.

Francisco J Fraga, Tiago H Falk, Paulo AM Kanda, and Renato Anghinah. Characterizing alzheimer's disease severity via resting-awake eeg amplitude modulation analysis. *PloS one*, 8(8):e72240, 2013.

Lorena Gallego-Viñarás, Juan Miguel Mira-Tomás, Anna Michela Gaeta, Gerard Pinol-Ripoll, Ferrán Barbé, Pablo M Olmos, and Arrate Muñoz-Barrutia. Alzheimer's disease detection in eeg sleep signals. *IEEE Journal of Biomedical and Health Informatics*, 2024.

Heinrich Garn, Markus Waser, Manfred Deistler, Thomas Benke, Peter Dal-Bianco, Gerhard Ransmayr, Helena Schmidt, Guenter Sanin, Peter Santer, Georg Caravias, et al. Quantitative eeg markers relate to alzheimer's disease severity in the prospective dementia registry austria (prodem). *Clinical Neurophysiology*, 126(3):505–513, 2015.

Yingfeng Ge, Jianan Yin, Caie Chen, Shuo Yang, Yuduan Han, Chonglong Ding, Jiaming Zheng, Yifan Zheng, and Jinxin Zhang. An eeg-based framework for automated discrimination of conversion to alzheimer's disease in patients with amnestic mild cognitive impairment: an 18-month longitudinal study. *Frontiers in Aging Neuroscience*, 16:1470836, 2025.

Stephan Getzmann, Patrick D Gajewski, Daniel Schneider, and Edmund Wascher. Resting-state eeg data before and after cognitive activity across the adult lifespan and a 5-year follow-up. *Scientific Data*, 11(1): 988, 2024.

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, et al. Meg and eeg data analysis with mne-python. *Frontiers in Neuroinformatics*, 7:267, 2013.

Giulia Grande, Martina Valletta, Debora Rizzuto, Xin Xia, Chengxuan Qiu, Nicola Orsini, Matilda Dale, Sarah Andersson, Claudia Fredolini, Bengt Winblad, et al. Blood-based biomarkers of alzheimer's disease and incident dementia in the community. *Nature Medicine*, pp. 1–9, 2025.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

Christoffer Hatlestad-Hall, Trine Waage Rygvold, and Stein Andersson. Bids-structured resting-state electroencephalography (eeg) data extracted from an experimental paradigm. *Data in Brief*, 45:108647, 2022.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

Richard W Homan, John Herman, and Phillip Purdy. Cerebral location of international 10–20 system electrode placement. *Electroencephalography and clinical neurophysiology*, 66(4):376–382, 1987.

Cosimo Ieracitano, Nadia Mammone, Alessia Bramanti, Amir Hussain, and Francesco C Morabito. A convolutional neural network approach for classification of dementia stages based on 2d-spectral representation of eeg recordings. *Neurocomputing*, 323:96–107, 2019.

Wei-Bang Jiang, Yansen Wang, Bao-Liang Lu, and Dongsheng Li. Neurolm: A universal multi-task foundation model for bridging the gap between language and eeg signals. *arXiv preprint arXiv:2409.00101*, 2024a.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci. *The Twelfth International Conference on Learning Representations*, 2024b.

Pramod H Kachare, Sandeep B Sangle, Digambar V Puri, Mousa Mohammed Khubrani, and Ibrahim Al-Shourbaji. Steadynet: spatiotemporal eeg analysis for dementia detection using convolutional neural network. *Cognitive Neurodynamics*, pp. 1–14, 2024.

Paulo Afonso Medeiros Kanda, Lucas R Trambaiolli, Ana C Lorena, Francisco J Fraga, Luis Fernando I Basile, Ricardo Nitrini, and Renato Anghinah. Clinician's road map to wavelet eeg as an alzheimer's disease biomarker. *Clinical EEG and neuroscience*, 45(2):104–112, 2014.

Min-jae Kim, Young Chul Youn, and Joonki Paik. Deep learning-based eeg analysis to classify normal, mild cognitive impairment, and dementia: Algorithms and dataset. *NeuroImage*, 272:120054, 2023.

Dominik Klepl, Fei He, Min Wu, Daniel J Blackburn, and Ptolemaios Sarrigiannis. Adaptive gated graph convolutional network for explainable diagnosis of alzheimer's disease using eeg data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023.

Demetres Kostas, Stephane Aroca-Ouellette, and Frank Rudzicz. Bendr: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of eeg data. *Frontiers in Human Neuroscience*, 15:653659, 2021.

NN Kulkarni and VK Bairagi. Extracting salient features for eeg-based diagnosis of alzheimer's disease using support vector machine classifier. *IETE Journal of Research*, 63(1):11–22, 2017.

Mojtaba Lahijanian, Hamid Aghajan, and Zahra Vahabi. Auditory gamma-band entrainment enhances default mode network connectivity in dementia patients. *Scientific Reports*, 14(1):13153, 2024.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

Colin Lea, Michael D Flynn, Rene Vidal, Austin Reiter, and Gregory D Hager. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 156–165, 2017.

Fangzhou Li, Shoya Matsumori, Naohiro Egawa, Shusuke Yoshimoto, Kotaro Yamashiro, Haruo Mizutani, Noriko Uchida, Atsuko Kokuryu, Akira Kuzuya, Ryosuke Kojima, et al. Predictive diagnostic approach to dementia and dementia subtypes using wireless and mobile electroencephalography: A pilot study. *Bioelectricity*, 4(1):3–11, 2022.

Dingkun Liu, Zhu Chen, Jingwei Luo, Shijie Lian, and Dongrui Wu. Mirepnet: A pipeline and foundation model for eeg-based motor imagery classification. *arXiv preprint arXiv:2507.20254*, 2025.

Xiaokun Liu, Chunlai Zhang, Zheng Ji, Yi Ma, Xiaoming Shang, Qi Zhang, Wencheng Zheng, Xia Li, Jun Gao, Ruofan Wang, et al. Multiple characteristics analysis of alzheimer's electroencephalogram by power spectral density and lempel–ziv complexity. *Cognitive neurodynamics*, 10:121–133, 2016.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. *International conference on learning representations*, 2023a.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2023b.

Donghao Luo and Xue Wang. Moderntcn: A modern pure convolution structure for general time series analysis. In *The twelfth international conference on learning representations*, pp. 1–43, 2024.

Colin L Masters, Randall Bateman, Kaj Blennow, Christopher C Rowe, Reisa A Sperling, and Jeffrey L Cummings. Alzheimer's disease. *Nature reviews disease primers*, 1(1):1–18, 2015.

Andreas Miltiadous, Emmanouil Gionanidis, Katerina D Tzimourta, Nikolaos Giannakeas, and Alexandros T Tzallas. Dice-net: a novel convolution-transformer architecture for alzheimer detection in eeg signals. *IEEE Access*, 2023a.

Andreas Miltiadous, Katerina D Tzimourta, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Dimitrios G Tsalikakis, Pantelis Angelidis, Markos G Tsipouras, Euripidis Glavas, Nikolaos Giannakeas, et al. A dataset of scalp eeg recordings of alzheimer's disease, frontotemporal dementia and healthy subjects from routine eeg. *Data*, 8(6):95, 2023b.

Navid Mohammadi Foumani, Geoffrey Mackellar, Soheila Ghane, Saad Irtza, Nam Nguyen, and Mahsa Salehi. Eeg2rep: enhancing self-supervised eeg representation through informative masked inputs. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5544–5555, 2024.

Aldo Mora-Sánchez, Gérard Dreyfus, and François-Benoît Vialatte. Scale-free behaviour and metastable brain-state switching driven by human cognition, an empirical approach. *Cognitive neurodynamics*, 13: 437–452, 2019.

National Academies of Sciences, Engineering, and Medicine. Reducing the impact of dementia in america: A decadal survey of the behavioral and social sciences. 2021.

BR Nayana, MN Pavithra, S Chaitra, TN Bhuvana Mohini, Thompson Stephan, Vijay Mohan, and Neha Agarwal. Eeg-based neurodegenerative disease diagnosis: comparative analysis of conventional methods and deep learning models. *Scientific Reports*, 15(1):15950, 2025.

Lucy Nelson and Naji Tabet. Slowing the progression of alzheimer's disease; what works? *Ageing research reviews*, 23:193–209, 2015.

Y Nie. A time series is worth 64words: Long-term forecasting with transformers. *International conference on learning representations*, 2023.

Julie Ottoy, Nicole Owsicki, Murat Bilgel, Alexa Pichet Binette, Gemma Salvadó, Min Su Kang, David M Cash, Michael Ewers, Renaud La Joie, Laura EM Wisse, et al. Recent advances in neuroimaging of alzheimer's disease and related dementias. *Alzheimer's & Dementia*, 21(9):e70648, 2025.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. Reve: A foundation model for eeg–adapting to any setup with large-scale pretraining on 25,000 subjects. *39th Conference on Neural Information Processing Systems*, 2025.

Saarang Panchavati, Uddhav Panchavati, Corey Arnold, and William Speier. Laya: A lejepa approach to eeg via latent prediction over reconstruction. *arXiv preprint arXiv:2603.16281*, 2026.

Henry Phalen, Brian A Coffman, Avniel Ghuman, Ervin Sejdić, and Dean F Salisbury. Non-negative matrix factorization reveals resting-state cortical alpha network abnormalities in the first-episode schizophrenia spectrum. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(10):961–970, 2020.

Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197, 2019.

Pavel Prado, Vicente Medel, Raul Gonzalez-Gomez, Agustín Sainz-Ballesteros, Victor Vidal, Hernando Santamaría-García, Sebastian Moguilner, Jhony Mejia, Andrea Slachevsky, Maria Isabel Behrens, et al. The brainlat project, a multimodal neuroimaging dataset of neurodegeneration from underrepresented backgrounds. *Scientific Data*, 10(1):889, 2023.

Carlos Roncero-Parra, Alfonso Parreño-Torres, Roberto Sánchez-Reolid, Jorge Mateo-Sotos, and Alejandro L Borja. Inter-hospital moderate and advanced alzheimer's disease detection through convolutional neural networks. *Heliyon*, 10(4), 2024.

Gemma Salvadó, Kanta Horie, Nicolas R Barthélemy, Jacob W Vogel, Alexa Pichet Binette, Charles D Chen, Andrew J Aschenbrenner, Brian A Gordon, Tammie LS Benzinger, David M Holtzman, et al. Disease staging of alzheimer's disease using a csf-based biomarker model. *Nature Aging*, 4(5):694–708, 2024.

Magali T Schmidt, Paulo AM Kanda, Luis FH Basile, Helder Frederico da Silva Lopes, Regina Baratho, Jose LC Demario, Mario S Jorge, Antonio E Nardi, Sergio Machado, Jéssica N Ianof, et al. Index of alpha/theta ratio of the electroencephalogram: a new marker for alzheimer's disease. *Frontiers in aging neuroscience*, 5:60, 2013.

Xiaocai Shan, Jun Cao, Shoudong Huo, Liangyu Chen, Ptolemaios Georgios Sarrigiannis, and Yifan Zhao. Spatial–temporal graph convolutional network for alzheimer classification based on brain functional connectivity imaging of electroencephalogram. *Human Brain Mapping*, 43(17):5194–5209, 2022.

Oded Shor, Amir Glik, Amit Yaniv-Rosenfeld, Avi Valevski, Abraham Weizman, Andrei Khrennikov, and Felix Benninger. Eeg p-adic quantum potential accurately identifies depression, schizophrenia and cognitive decline. *Plos one*, 16(8):e0255529, 2021.

Arun Singh, Rachel C Cole, Arturo I Espinoza, Jan R Wessel, James F Cavanagh, and Nandakumar S Narayanan. Evoked mid-frontal activity predicts cognitive dysfunction in parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 94(11):945–953, 2023.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2022.

Luke Tait, George Stothart, Elizabeth Coulthard, Jon T Brown, Nina Kazanina, and Marc Goodfellow. Network substrates of cognitive impairment in alzheimer's disease. *Clinical Neurophysiology*, 130(9): 1581–1595, 2019.

Lucas R Trambaiolli, Ana C Lorena, Francisco J Fraga, Paulo AM Kanda, Renato Anghinah, and Ricardo Nitrini. Improving alzheimer's disease diagnosis with machine learning techniques. *Clinical EEG and neuroscience*, 42(3):160–165, 2011.

Lucie Tylova, Jaromir Kukal, and Oldrich Vysata. Predictive models in diagnosis of alzheimer's disease from eeg. *Acta Polytechnica*, 53(2), 2013.

Lucie Tylová, Jaromír Kukal, Václav Hubata-Vacek, and Oldřich Vyšata. Unbiased estimation of permutation entropy in eeg analysis for alzheimer's disease classification. *Biomedical Signal Processing and Control*, 39: 424–430, 2018.

Katerina D Tzimourta, Theodora Afrantou, Panagiotis Ioannidis, Maria Karatzikou, Alexandros T Tzallas, Nikolaos Giannakeas, Loukas G Astrakas, Pantelis Angelidis, Evripidis Glavas, Nikolaos Grigoriadis, et al. Analysis of electroencephalographic signals complexity regarding alzheimer's disease. *Computers & Electrical Engineering*, 76:198–212, 2019a.

Katerina D Tzimourta, Nikolaos Giannakeas, Alexandros T Tzallas, Loukas G Astrakas, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Pantelis Angelidis, Dimitrios G Tsalikakis, and Markos G Tsipouras. Eeg window length evaluation for the detection of alzheimer's disease over different brain regions. *Brain sciences*, 9(4):81, 2019b.

Hanneke Van Dijk, Guido Van Wingen, Damiaan Denys, Sebastian Olbrich, Rosalinde Van Ruth, and Martijn Arns. The two decades brainclinics research archive for insights in neurophysiology (tdbrain) database. *Scientific data*, 9(1):333, 2022.

Fabrizio Vecchio, Francesca Miraglia, Francesco Iberite, Giordano Lacidogna, Valeria Guglielmi, Camillo Marra, Patrizio Pasqualetti, Francesco Danilo Tiziano, and Paolo Maria Rossini. Sustainable method for alzheimer dementia prediction in mild cognitive impairment: Electroencephalographic connectivity and graph theory combined with apolipoprotein e. *Annals of neurology*, 84(2):302–314, 2018.

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.

Mário L Vicchietti, Fernando M Ramos, Luiz E Betting, and Andriana SLO Campanharo. Computational methods of eeg signals analysis for alzheimer's disease classification. *Scientific Reports*, 13(1):8184, 2023.

Jonathan Vöglein, Nicolai Franzmeier, John C Morris, Marianne Dieterich, Eric McDade, Mikael Simons, Oliver Preische, Anna Hofmann, Jason Hassenstab, Tammie L Benzinger, et al. Pattern and implications of neurological examination findings in autosomal dominant alzheimer disease. *Alzheimer's & Dementia*, 19(2):632–645, 2023.

Guangyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. Eegpt: Pretrained transformer for universal and reliable representation of eeg signals. *Advances in Neural Information Processing Systems*, 2024a.

Jing Wang, Yuxing Fang, Xiao Wang, Huichao Yang, Xin Yu, and Huali Wang. Enhanced gamma activity and cross-frequency interaction of resting-state electroencephalographic oscillations in patients with alzheimer's disease. *Frontiers in aging neuroscience*, 9:243, 2017.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. Cbramod: A criss-cross brain foundation model for eeg decoding. *ICLR*, 2025.

Ruofan Wang, Jiang Wang, Shunan Li, Haitao Yu, Bin Deng, and Xile Wei. Multiple feature extraction and classification of electroencephalograph signal for alzheimers' with spectrum and bispectrum. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(1), 2015.

Yihe Wang, Yu Han, Haishuai Wang, and Xiang Zhang. Contrast everything: A hierarchical contrastive framework for medical time-series. *Advances in Neural Information Processing Systems*, 36:55694–55717, 2023.

Yihe Wang, Nan Huang, Taida Li, Yujun Yan, and Xiang Zhang. Medformer: A multi-granularity patching transformer for medical time-series classification. *Advances in Neural Information Processing Systems*, 2024b.

Yihe Wang, Taida Li, Yujun Yan, Wenzhan Song, and Xiang Zhang. How to evaluate your medical time series classification? *arXiv preprint arXiv:2410.03057*, 2024c.

Yihe Wang, Nadia Mammone, Darina Petrovsky, Alexandros T Tzallas, Francesco C Morabito, and Xiang Zhang. Adformer: A multi-granularity transformer for eeg-based alzheimer's disease assessment. *arXiv preprint arXiv:2409.00032*, 2024d.

Markus Waser, Manfred Deistler, Heinrich Garn, Thomas Benke, Peter Dal-Bianco, Gerhard Ransmayr, Dieter Grossegger, and Reinhold Schmidt. Eeg in the diagnostics of alzheimer's disease. *Statistical Papers*, 54:1095–1107, 2013.

Markus Waser, Heinrich Garn, Reinhold Schmidt, Thomas Benke, Peter Dal-Bianco, Gerhard Ransmayr, Helena Schmidt, Stephan Seiler, Günter Sanin, Florian Mayer, et al. Quantifying synchrony patterns in the eeg of alzheimer's patients with linear and non-linear connectivity markers. *Journal of Neural Transmission*, 123:297–316, 2016.

Yusuke Watanabe, Yuki Miyazaki, Masahiro Hata, Ryohei Fukuma, Yasunori Aoki, Hiroaki Kazui, Toshihiko Araki, Daiki Taomoto, Yuto Satake, Takashi Suehiro, et al. A deep learning model for the detection of various dementia and mci pathologies based on resting-state electroencephalography data: A retrospective multicentre study. *Neural Networks*, 171:242–250, 2024.

Di Wu, Siyuan Li, Jie Yang, and Mohamad Sawan. Neuro-bert: Rethinking masked autoencoding for self-supervised neurological pretraining. *IEEE Journal of Biomedical and Health Informatics*, 2024.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*, 2022.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

Qing-Song Xu and Yi-Zeng Liang. Monte carlo cross validation. *Chemometrics and Intelligent Laboratory Systems*, 56(1):1–11, 2001.

Chaoqi Yang, M Westover, and Jimeng Sun. Biot: Biosignal transformer for cross-data learning in the wild. *Advances in Neural Information Processing Systems*, 36, 2024.

Zhizhang Yuan, Daoze Zhang, Yang Yang, Junru Chen, and Yafeng Li. Ppi: Pretraining brain signal model for patient-independent seizure detection. *Advances in Neural Information Processing Systems*, 36:69586–69604, 2023.

Tongtian Yue, Shuning Xue, Xuange Gao, Yepeng Tang, Longteng Guo, Jie Jiang, and Jing Liu. Eegpt: Unleashing the potential of eeg generalist foundation model by autoregressive pre-training. *arXiv preprint arXiv:2410.19779*, 2024.

Ce Zhang, Young-Keun Kim, and Azim Eskandarian. Eeg-inception: an accurate and robust end-to-end neural network for eeg-based motor imagery classification. *Journal of Neural Engineering*, 18(4):046014, 2021.

Daoze Zhang, Zhizhang Yuan, Yang Yang, Junru Chen, Jingjing Wang, and Yafeng Li. Brant: Foundation model for intracranial neural signal. *Advances in Neural Information Processing Systems*, 36:26304–26321, 2023.

Daoze Zhang, Zhizhang Yuan, Junru Chen, Kerui Chen, and Yang Yang. Brant-x: A unified physiological signal alignment framework. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4155–4166, 2024.

Ruizhe Zheng, Lingyan Mao, Dingding Han, Tian Luo, Yi Wang, Jing Ding, and Yuguo Yu. Fapex: Fractional amplitude-phase expressor for robust cross-subject seizure prediction. *39th Conference on Neural Information Processing Systems*, 2025.

Yuchen Zhou, Jiamin Wu, Zichen Ren, Zhouheng Yao, Weiheng Lu, Kunyu Peng, Qihao Zheng, Chunfeng Song, Wanli Ouyang, and Chao Gou. Csbrain: A cross-scale spatiotemporal brain foundation model for eeg decoding. *arXiv preprint arXiv:2506.23075*, 2025.

Qiushi Zhu, Xiaoying Zhao, Jie Zhang, Yu Gu, Chao Weng, and Yuchen Hu. Eeg2vec: Self-supervised electroencephalographic representation learning. *arXiv preprint arXiv:2305.13957*, 2023.

Appendix A Index Group-Shuffling

Algorithm 2 Index Group-Shuffling

Require: Subject IDs $\{s_i\}_{i=1}^{N}$, group size G, batch size B ▷ N denotes the number of all training samples
Ensure: A shuffled index permutation π of all training samples

$\mathbf{p} \leftarrow \text{argsort}(\{s_i\}_{i=1}^{N})$ ▷ Sort sample indices by subject ID
$\mathcal{G} \leftarrow [\,]$ ▷ Initialize the list of groups
for $i = 1$ to N step G **do**
  $\mathcal{G} \leftarrow \mathcal{G} \cup \left\{\mathbf{p}[i : \min(i + G - 1, N)]\right\}$ ▷ Split sorted indices into groups
end for
Shuffle($\mathcal{G}$) ▷ Shuffle the group order
$\mathbf{q} \leftarrow \text{Concat}(\mathcal{G})$ ▷ Flatten all groups
$\mathcal{B} \leftarrow [\,]$ ▷ Initialize the list of batches
for $j = 1$ to N step B **do**
  $\mathbf{u} \leftarrow \mathbf{q}[j : \min(j + B - 1, N)]$ ▷ Form one batch
  Shuffle($\mathbf{u}$) ▷ Shuffle samples within the batch
  $\mathcal{B} \leftarrow \mathcal{B} \cup \{\mathbf{u}\}$
end for
$\pi \leftarrow \text{Concat}(\mathcal{B})$ ▷ Flatten all batches into the final permutation
return π

Appendix B Data Augmentation Banks

For contrastive pre-training, we employ a bank of data augmentation methods to enhance the model's robustness and generalization capabilities. During the forward pass in the training of each iteration, one augmentation method will be picked from available augmentation options with equal probability. The data augmentation methods include temporal flipping, temporal masking, frequency masking, channel masking, jittering, and dropout, and can be further expanded to more choices.

1) Temporal Flippling. We reverse the EEG data along the temporal dimension. The probability of applying this augmentation is controlled by a parameter *prob*, with a default value of 0.5. **2) Temporal Masking.** We randomly mask timestamps across all channels. The proportion of timestamps masked is controlled by the parameter *ratio*, with a default value of 0.1. **3) Frequency Masking.** This method involves converting the EEG data into the frequency domain, randomly masking some frequency bands, and then converting it back. The proportion of frequency bands masked is controlled by the parameter *ratio*, with a default value of 0.1. **4) Channel Masking.** We randomly mask channels across all timestamps. The proportion of channel masked is controlled by the parameter *ratio*, with a default value of 0.1. **5) Jittering.** Random noise, ranging from 0 to 1, is added to the raw data. The intensity of the noise is adjusted by the parameter *scale*, which is set by default to 0.1. **6) Dropout.** Similar to the dropout layer in neural networks, this method randomly drops some values. The proportion of values dropped is controlled by the parameter *ratio*, with a default value of 0.1.

Appendix C Implementation Details

Manual Feature utilize 32 features, including mean, variance, skewness, kurtosis, std, iqr, max, min, mean, median, delta power, theta power, alpha power, beta power, total power, theta alpha ratio, alpha beta ratio, delta relative power, theta relative power, alpha relative power, beta relative power, phase coherence, spectral centroid, spectral rolloff, spectral peak, average magnitude, median frequency, amplitude modulation, spectral entropy, tsallis entropy, and shannon entropy (Tzimourta et al., 2019b;a; Kulkarni & Bairagi, 2017; Kanda et al., 2014; Waser et al., 2013; Tylova et al., 2013; Mora-Sánchez et al., 2019; Wang et al., 2017; Cassani et al., 2014; Wang et al., 2015; Fraga et al., 2013; Tait et al., 2019; Waser et al., 2016; Trambaiolli et al., 2011; Fahimi et al., 2017; Schmidt et al., 2013; Liu et al., 2016; Kanda et al., 2014; Garn et al., 2015; Tylová

et al., 2018; Coronel et al., 2017; Al-Nuaimi et al., 2018). A linear projection layer is then applied to these manually extracted features for final classification.

EEGConformer (Song et al., 2022) uses convolutional modules to learn low-level local features and embed the raw data into patches for self-attention. We set *e_layers* = 12, *n_heads* = 8, *d_model* = 128, and *d_ff* = 256.

EEGInception (Zhang et al., 2021) uses different scales of convolutional kernels, combined with a spatial block for feature extraction. We set *n_blocks* = 3, *channels* = (96,192,384), *kernel_sizes* = (8,16,32), *depth_multiplier* = 2, *bottleneck_channels* = 32.

EEGNet (Lawhern et al., 2018) is a classic deep learning method for EEG decoding. It uses depthwise and separable convolutions to capture spatial and temporal features. We keep the same structure when applying convolutions, normalization, and activations as described in the paper.

iTransformer (Liu et al., 2023a) proposes a novel data-embedding method based on the transformer architecture for multivariate time-series analysis. It reverses the conventional embedding strategy. Instead of using multi-channel samples at a single time point into a single temporal token, iTransformer treats each channel/variate as an independent variable token. We set *e_layers* = 12, *n_heads* = 8, *d_model* = 128, and *d_ff* = 256.

MedGNN (Fan et al., 2025) is a graph-nerual-network-based method for medical time-series classification. It uses a multi-resolution graph transformer architecture to model the dynamic dependencies and fuse the information from different resolutions. We set *–resolution_list* = 2,4,6,8, *–nodedim* = 10, *e_layers* = 12, *n_heads* = 8, *d_model* = 128, and *d_ff* = 256.

Medformer (Wang et al., 2024b) is designed for biomedical time series classification, including EEG and ECG. Cross-channel multi-granularity patch embedding and intra-inter-granularity self-attention are utilized. We extended our work based on this method. We set *e_layers* = 12, *d_model* = 128, and *d_ff* = 256, *patch_len_list* = [5, 10, 20].

MNet (Watanabe et al., 2024) is a convolution-based deep learning method for EEG-based AD detection. It contains 4 Con2D blocks for feature extraction. We keep the same structure and order when applying convolutions, normalization, and activations as described in the paper.

ModernTCN (Luo & Wang, 2024) is a convolutional architecture designed for general time-series analysis and demonstrates strong performance, particularly in classification. It combines depthwise convolutions with multiple pointwise convolutions and introduces channel-independent embedding mechanisms. We set *patch_len* = 20, *stride* = 10, *num_blocks* = 1 1 1 1, *large_size* = 9 9 9 9, *small_size* = 5 5 5 5, *dims* = 32 64 128 128.

PatchTST (Nie, 2023) is a transformer-based model for time series prediction and representation learning, with its core innovation lying in a dual design of patching and channel independence. The model decomposes multi-channel data into multiple one-channel patches, treating them as semantically informative sub-sequence fragments for input. We set *e_layers* = 12, *n_heads* = 8, *d_model* = 128, and *d_ff* = 256, *patch_len* = 25.

TCN (Lea et al., 2017) is a convolutional architecture designed explicitly for time-series modeling. They extend residual networks by incorporating causal and dilated convolutions, enabling effective learning of long-range temporal dependencies. We set *e_layers* = 12, *n_heads* = 8, *d_model* = 128, and *d_ff* = 256.

TimesNet (Wu et al., 2022) is a convolutional model for time-series analysis, especially on classification tasks. It transforms one-dimensional time-series data into two-dimensional tensors across multiple periodicities, thereby leveraging the strengths of 2D convolutional networks for time-series analysis. We set *e_layers* = 2, *top_k* = 3, *d_model* = 32, and *d_ff* = 64.

BIOT (Yang et al., 2024) is the first foundation model for biomedical time-series data, including EEG. It employs single-channel patch embedding to handle biosignals with varying channel counts. Each patch is mapped to tokens, with segment, channel, and positional embeddings added to make the tokens distinguishable. To align the channel configuration of our datasets with their checkpoints, we use a 1D convolution (Conv1D) for channel mapping. We retain the other default architectural parameters from the checkpoint. The

pre-trained checkpoint is available at https://github.com/ycq091044/BIOT/blob/main/pretrained-models/EEG-PREST-16-channels.ckpt.

LaBraM (Jiang et al., 2024b) is the first EEG foundation model, trained on 2,000 hours of EEG recordings collected from multiple datasets. Its pre-training follows a two-step strategy that combines vector quantization with mask-based reconstruction. In our experiments, we use their released pre-trained checkpoint with default parameters and fine-tune it on four AD downstream datasets. To align the channel configuration of our datasets with their checkpoints, we use a 1D convolution (Conv1D) for channel mapping. We retain the other default architectural parameters from the checkpoint. The pre-trained checkpoint is available at https://github.com/935963004/LaBraM/blob/main/checkpoints/labram-base.pth.

CBraMod (Wang et al., 2025) is an SOTA EEG foundation model trained on the 9000 hours TUEG dataset with 19 standard channels of the 10-20 system. They use a criss-cross transformer to perform spatial and temporal attention to capture features along two dimensions. We keep their default parameters to load their pre-trained checkpoint and fine-tune on 5 AD downstream datasets. The pre-trained checkpoint is available at https://huggingface.co/weighting666/CBraMod/blob/main/pretrained_weights.pth.

CSBrain (Zhou et al., 2025) is an SOTA EEG foundation model that leverages brain regions to model cross-window and cross-region dependencies for diverse EEG decoding tasks. This model is also trained on the 9000 hours TUEG dataset with 19 standard channels of the 10-20 system. We keep their default parameters to load their pre-trained checkpoint and fine-tune on 5 AD downstream datasets. The pre-trained checkpoint is available at https://drive.google.com/drive/folders/1je-1TtdHv6klcd-kTlPNkiA1wLrxybva.

LEAD (Ours). We pre-train on 13 datasets (4 AD and 9 Non-AD) and then fine-tune on 5 downstream AD datasets. We set the patch length L to 50 and the stride to 50. The sample- and subject-level cross-entropy loss weights α and β are both set to 0.5. The contrastive loss coefficients λ_1 and λ_2 are set to 0.25 and 0.75. We set *e_layers* = 12, *n_heads* = 8, *d_model* = 128, *d_ff* = 256, and *group_size* = 8. The indices group shuffling, sampling rate embedding, and multi-sampling segmentation are all enabled, with *multi_sampling_rate* set to 200, 100, and 50 Hz. The *montage_name* is set to *standard_1005*, and *channel_names* are exactly matched with the channel configuration of the downstream dataset.

Appendix D Datasets Preprocessing

D.1 Data Curation

We categorize all EEG datasets utilized in this paper into two groups: those containing Alzheimer's Disease subjects (**AD** datasets) and those without (**Non-AD** datasets).

AD Datasets. Despite the promise of EEG for AD detection, the field remains critically limited by the scarcity of accessible datasets. To address this, we conducted a comprehensive review of relevant publications and public EEG repositories (e.g., OpenNeuro, Dryad, figshare) published before 2026. This effort identified 8 public datasets. In total, we utilize 8 public and 1 private AD datasets, most of which consist of resting-state recordings, with minor evoked steady-state response: **AD-Auditory** (Lahijanian et al., 2024), **Brain-Lat** (Prado et al., 2023), **P-ADIC** (Shor et al., 2021), **CAUEEG** (Kim et al., 2023), **ADFSU** (Vicchietti et al., 2023), **ADFTD** (Miltiadous et al., 2023b), **ADSZ** (Alves et al., 2022), **APAVA** (Escudero et al., 2006), and **CNBPM** (Amezquita-Sanchez et al., 2019). Together, these datasets comprise **2,238 subjects and 427.81 hours** of recordings, forming the **world's largest** EEG-AD corpus reported so far.

Non-AD Datasets. While the inclusion of AD datasets represents a major step forward, their scale is still limited for large-scale pre-training, and a portion must be reserved for downstream evaluation. To further expand our training resources, we curated 9 **domain-relevant** non-AD datasets. Unlike existing EEG foundation models that mix paradigms such as resting-state (RS), motor imagery (MI), and event-related potentials (ERP) (Jiang et al., 2024b; Wang et al., 2024a), we deliberately focus on datasets aligned with AD detection. This decision is motivated by the fact that different paradigms reflect distinct neuroscientific processes and cortical activations, usually requiring paradigm-specific preprocessing pipelines. Combining data from heterogeneous paradigms can introduce conflicting patterns and diminish the utility of pre-training resources. Our curated non-AD datasets contain recordings from both healthy subjects and patients with

other brain disorders (e.g., PD, DEP, ADHD). Specifically, we include: **BACA-RS** (Getzmann et al., 2024), **Depression** (Cavanagh et al., 2019), **FEPCR** (Phalen et al., 2020), **MCEF-RS** (Chenot et al., 2024), **PD-RS** (Singh et al., 2023), **PEARL-Neuro** (Dzianok & Kublik, 2024), **SRM-RS** (Hatlestad-Hall et al., 2022), **TDBrain** (Van Dijk et al., 2022), and **TUEP** (Veloso et al., 2017). Collectively, they comprise **2,848 subjects and 805.62 hours** of EEG. Most of these datasets were acquired under resting-state or resting-state-like paradigms, ensuring consistency with our downstream AD detection.

D.2 Unified Data Preprocessing

The statistics of processed datasets are provided in Table 2. The datasets are highly heterogeneous, with substantial variability in channel numbers, sampling rates, and recording lengths. To utilize them for training, we apply the following preprocessing pipeline, applied sequentially:

1) Removal of non-EEG channels: All non-EEG channels are removed, such as EOG, ECG, or coordinate information. **2) Notch and Band-Pass Filtering:** Each trial undergoes a notch filter at 50 Hz or 60 Hz, followed by a band-pass filter between 0.5 Hz and 45 Hz. This step aims to suppress line noise, slow drifts, and high-frequency noise, which are generally outside the frequency range of scalp-recorded brain activity. **3) Average re-referencing:** Average re-referencing is applied to reduce global noise and potential baseline shifts. **1) Artifact Removal:** For datasets lacking prior artifact rejection, we utilize independent component analysis (ICA) combined with the ICLabel algorithm (Pion-Tonachini et al., 2019) to automatically identify and remove components associated with artifacts like eye blinks, or muscle activity. **2) Channel Alignment:** We align all pre-training datasets (downstream datasets could be an arbitrary number of channels and montage) to the standard 19-channel montage based on the international 10-20 system: Fp1, Fp2, F7, F3, Fz, F4, F8, T3/T7, C3, Cz, C4, T4/T8, T5/P7, P3, Pz, P4, T6/P8, O1, and O2 (Homan et al., 1987). If a dataset has more channels, only the 19 channels with these names are selected, and the rest are discarded. For datasets employing different montages (e.g., Biosemi (bio)), signals are projected onto the target 19 channels using their 3D coordinates. All the fine-tuning datasets keep the raw channel. **3) Frequency Alignment:** All datasets are resampled to a uniform sampling frequency of 200 Hz. This rate is widely used, adequately captures the main physiological EEG frequency bands (δ, θ, α, β, γ), and reduces high-frequency noise. **4) Data Segmentation:** We use a multi-scale segmentation (see para. 3.2) strategy to segment recordings into training samples. **6) Z-Score Normalization:** Z-score normalization is applied to each segmented sample, computed independently for each channel.

Appendix E Detailed Results

This section reports detailed experimental results for different tasks, including baseline method comparisons and per-subject analysis. The evaluation is conducted on five downstream datasets: ADFSU, ADFTD, ADSZ, APAVA, and CNBPM. We report all 14 evaluation metrics, including sample-level accuracy, precision (macro-averaged), sensitivity/recall (macro-averaged), specificity (macro-averaged), F1 score (macro-averaged), AUROC (macro-averaged), and AUPRC (macro-averaged), and their corresponding subject-level metrics computed via majority voting (See para. 3.3).

E.1 Method Comparison Results

The detailed results for baseline method comparisons on five downstream datasets are presented in Table 7 (ADFSU), Table 8 (ADFTD), Table 9 (ADSZ) , Table 10 (APAVA), and Table 11 (CNBPM), respectively.

E.2 Per-Subject Analysis

The detailed results for the per-subject analysis under the leave-one-subject-out (LOSO) setting on the ADFTD dataset are presented in Table 12 and Table 13.

Table 7: **Method Comparison Detailed Results on ADFSU.** The detailed results comparison between our method and baseline methods on the dataset ADFSU includes all 7 evaluation metrics. The **Top-1**, Top-2, and Top-3 results are highlighted in red, blue, and green.

Datasets	Models	Accuracy	Precision	Sensitivity	Specificity	F1 Score	AUROC	AUPRC
ADFSU *(Sample-Level)* *(4,048 Samples)* *(HC vs AD)*	**ManualFeature**	87.40±3.88	82.55±9.01	73.87±10.80	73.87±10.80	75.67±11.41	80.90±8.87	77.35±9.05
	EEGConformer	96.53±1.93	96.75±2.12	92.33±4.46	92.33±4.46	94.22±3.28	98.87±0.95	98.36±1.24
	EEGInception	94.60±0.98	94.49±1.15	88.63±4.21	88.63±4.21	90.88±2.13	98.73±0.75	97.99±1.19
	EEGNet	82.27±10.27	77.52±10.71	74.92±6.79	74.92±6.79	74.48±9.91	85.88±7.95	82.41±10.23
	iTransformer	87.80±1.07	86.93±3.61	72.75±2.50	72.75±2.50	76.80±2.31	88.16±1.86	85.59±1.34
	MedGNN	94.13±1.50	94.34±3.59	87.08±1.86	87.08±1.86	90.09±2.37	98.06±1.42	97.08±2.13
	Medformer	92.73±0.80	93.58±0.78	83.33±2.71	83.33±2.71	87.15±1.91	97.48±1.32	96.30±1.18
	MNet	88.40±3.38	90.34±3.43	72.12±8.24	72.12±8.24	76.19±8.72	92.75±5.08	90.30±5.71
	ModernTCN	87.80±1.19	90.98±2.03	70.38±2.93	70.38±2.93	75.09±3.18	91.94±1.84	89.57±2.09
	PatchTST	87.13±1.42	83.81±2.01	72.83±4.55	72.83±4.55	76.03±4.27	89.12±1.95	85.53±2.14
	TCN	94.27±1.70	95.07±2.54	86.79±3.71	86.79±3.71	90.10±3.13	98.82±0.73	98.00±1.23
	TimesNet	90.60±2.39	92.28±2.38	77.88±6.26	77.88±6.26	82.08±6.29	95.21±2.45	93.39±2.27
	BIOT	93.47±1.51	92.80±2.71	86.17±2.83	86.17±2.83	88.94±2.69	96.78±3.04	95.60±3.05
	LaBraM	95.47±1.77	96.14±0.98	89.67±4.97	89.67±4.97	92.24±3.38	98.90±0.53	97.98±0.88
	CBraMod	95.40±2.28	94.56±2.40	90.88±5.60	90.88±5.60	92.32±4.27	98.47±1.49	97.59±2.27
	CSBrain	94.67±1.14	93.96±1.28	89.17±3.69	89.17±3.69	91.11±2.34	98.23±0.51	97.23±0.63
	LEAD (Ours)	**98.18±0.61**	**98.70±0.44**	**95.62±1.54**	**95.62±1.54**	**97.06±1.02**	**99.92±0.07**	**99.85±0.13**
ADFSU *(Subject-Level)* *(92 Subjects)* *(HC vs AD)*	**ManualFeature**	92.00±7.48	85.78±23.02	80.00±18.71	80.00±18.71	81.05±20.29	80.00±18.71	91.56±7.62
	EEGConformer	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	EEGInception	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	EEGNet	88.00±11.66	80.89±23.26	81.25±18.54	81.25±18.54	78.37±21.07	81.25±18.54	92.28±7.50
	iTransformer	86.00±4.90	72.67±26.67	65.00±12.25	65.00±12.25	66.01±17.61	65.00±12.25	85.33±4.35
	MedGNN	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	Medformer	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	MNet	84.00±4.90	61.78±26.67	60.00±12.25	60.00±12.25	58.82±17.61	60.00±12.25	83.56±4.35
	ModernTCN	84.00±4.90	61.78±26.67	60.00±12.25	60.00±12.25	58.82±17.61	60.00±12.25	83.56±4.35
	PatchTST	90.00±6.32	84.67±22.44	75.00±15.81	75.00±15.81	77.12±18.02	75.00±15.81	89.33±6.35
	TCN	98.00±4.00	98.89±2.22	95.00±10.00	95.00±10.00	96.08±7.84	95.00±10.00	97.78±4.44
	TimesNet	94.00±8.00	86.89±23.54	85.00±20.00	85.00±20.00	84.97±21.64	85.00±20.00	93.78±8.12
	BIOT	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	LaBraM	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	CBraMod	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	CSBrain	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	LEAD (Ours)	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**

Table 8: **Method Comparison Detailed Results on ADFTD.** The detailed results comparison between our method and baseline methods on the dataset ADFTD includes all 7 evaluation metrics. The **Top-1**, Top-2, and Top-3 results are highlighted in red, blue, and green.

Datasets	Models	Accuracy	Precision	Sensitivity	Specificity	F1 Score	AUROC	AUPRC
ADFTD *(Sample-Level)* *(167,083 Samples)* *(HC vs AD vs FTD)*	**ManualFeature**	46.98±4.11	45.82±4.61	45.27±4.08	73.13±2.00	44.73±3.88	62.89±4.31	43.69±3.63
	EEGConformer	61.28±2.83	62.39±4.23	58.82±3.53	79.87±1.51	58.59±3.84	80.61±3.90	69.14±5.37
	EEGInception	64.25±3.85	66.62±2.85	64.03±3.63	82.16±1.76	62.69±4.06	84.01±2.76	73.69±2.34
	EEGNet	45.22±3.93	40.83±8.66	42.91±4.03	71.60±1.93	40.31±6.74	60.93±4.42	43.59±5.12
	iTransformer	56.27±3.61	54.65±3.95	54.71±3.55	77.93±1.73	54.33±3.67	73.98±4.02	59.26±5.30
	MedGNN	69.60±3.85	70.22±4.07	68.15±4.74	84.48±2.16	67.67±4.60	85.52±2.57	76.86±3.18
	Medformer	66.66±4.89	67.19±4.49	65.13±6.15	82.80±2.69	64.83±6.72	84.13±3.27	74.59±5.16
	MNet	63.92±6.05	63.84±8.01	61.89±7.28	81.46±3.11	59.91±9.77	82.30±5.04	71.32±8.50
	ModernTCN	60.38±3.42	59.21±3.91	58.59±3.67	79.82±1.71	58.52±3.71	77.25±3.33	63.44±4.55
	PatchTST	55.48±3.23	54.35±6.92	52.24±2.94	76.72±1.53	50.95±4.04	71.30±4.75	56.53±6.94
	TCN	66.04±2.46	64.89±2.47	64.44±2.31	82.68±1.19	64.26±2.40	82.94±1.89	71.43±3.32
	TimesNet	61.85±3.37	61.12±4.68	59.71±4.92	80.38±1.87	58.66±5.68	79.18±3.39	66.48±6.35
	BIOT	70.94±5.66	70.86±5.53	69.86±5.91	85.14±2.96	69.79±5.90	86.85±4.77	78.40±6.69
	LaBraM	77.00±3.76	77.33±2.84	75.79±4.53	88.24±2.02	75.64±4.68	91.22±2.72	84.75±4.41
	CBraMod	69.92±3.95	69.52±4.41	68.58±3.97	84.60±1.94	68.33±4.53	86.95±2.89	78.06±4.26
	CSBrain	70.67±2.12	70.41±2.03	69.60±2.49	85.02±1.11	69.39±2.63	86.82±1.28	78.66±1.93
	LEAD (Ours)	**81.70±4.74**	**82.93±4.40**	**80.78±4.95**	**90.52±2.45**	**81.01±5.02**	**94.03±1.33**	**90.37±2.14**
ADFTD *(Subject-Level)* *(88 Subjects)* *(HC vs AD vs FTD)*	**ManualFeature**	62.00±4.00	56.03±15.90	59.44±4.16	80.16±1.81	54.26±7.80	69.80±2.94	53.37±4.75
	EEGConformer	74.00±10.20	70.95±20.86	72.22±10.97	86.03±5.35	68.45±15.21	79.13±8.16	66.26±12.37
	EEGInception	82.00±13.27	80.67±21.12	81.67±13.68	90.79±6.78	79.47±17.70	86.23±10.23	76.08±17.02
	EEGNet	64.00±10.20	58.81±22.32	60.56±11.17	80.63±5.60	54.82±14.34	70.60±8.38	54.25±10.70
	iTransformer	72.00±7.48	72.56±14.91	70.56±7.16	85.87±3.56	67.17±9.15	78.21±5.36	61.84±6.73
	MedGNN	78.00±4.00	83.33±2.79	77.78±4.65	89.05±2.09	75.23±4.81	83.41±3.30	69.28±4.90
	Medformer	80.00±10.95	84.11±8.15	80.00±11.17	89.68±5.66	78.98±12.08	84.84±8.41	73.44±14.44
	MNet	74.00±12.00	70.30±21.93	73.33±13.10	86.83±5.89	68.63±17.53	80.08±9.49	66.72±14.24
	ModernTCN	78.00±7.48	78.37±13.91	77.22±8.85	88.41±4.16	74.39±11.35	82.82±6.50	71.16±9.55
	PatchTST	72.00±9.80	68.81±21.28	69.44±10.24	85.08±5.00	65.04±14.32	77.26±7.62	62.33±12.48
	TCN	84.00±4.90	88.67±4.40	83.33±5.83	91.43±2.81	82.80±6.10	87.38±4.32	78.49±6.23
	TimesNet	72.00±11.66	61.86±21.27	71.11±13.10	86.03±5.82	64.23±17.73	78.57±9.45	63.38±14.25
	BIOT	88.00±9.80	91.11±7.54	87.78±10.03	93.65±5.19	88.10±9.72	90.71±7.61	83.72±13.29
	LaBraM	92.00±7.48	94.44±4.65	91.67±8.24	95.87±4.09	91.14±8.64	93.77±6.16	88.78±10.19
	CBraMod	84.00±4.90	88.78±4.13	82.78±5.09	91.43±2.48	82.21±6.30	87.10±3.77	77.31±6.86
	CSBrain	80.00±6.32	84.67±4.88	79.44±6.24	89.68±3.01	78.33±7.52	84.56±4.61	72.03±8.23
	LEAD (Ours)	**94.00±4.90**	**95.67±3.59**	**93.89±5.09**	**96.83±2.61**	**93.95±4.95**	**95.36±3.85**	**91.56±6.91**

Table 9: **Method Comparison Detailed Results on ADSZ.** The detailed results comparison between our method and baseline methods on the dataset ADSZ includes all 7 evaluation metrics. The **Top-1**, Top-2, and Top-3 results are highlighted in red, blue, and green.

Datasets	Models	Accuracy	Precision	Sensitivity	Specificity	F1 Score	AUROC	AUPRC
ADSZ *(Sample-Level)* *(1,128 Samples)* *(HC vs AD)*	**ManualFeature**	65.64±13.56	55.33±25.09	64.85±14.35	64.85±14.35	58.42±21.26	67.20±15.34	64.07±12.94
	EEGConformer	94.73±3.47	94.91±3.36	94.91±3.32	94.91±3.32	94.73±3.47	98.50±1.48	98.52±1.49
	EEGInception	92.62±4.43	93.43±3.85	92.93±4.16	92.93±4.16	92.60±4.44	98.06±1.66	98.04±1.75
	EEGNet	66.27±12.39	71.26±13.28	67.19±12.37	67.19±12.37	64.90±12.88	77.18±19.10	78.29±17.21
	iTransformer	73.29±4.86	73.91±5.53	73.30±4.80	73.30±4.80	73.13±4.75	81.86±4.90	82.02±5.15
	MedGNN	88.62±8.00	89.11±7.79	88.84±7.86	88.84±7.86	88.59±8.01	95.90±4.10	96.05±3.98
	Medformer	88.78±5.71	88.81±5.76	88.73±5.77	88.73±5.77	88.73±5.76	94.18±3.72	94.22±3.91
	MNet	79.26±7.34	79.67±7.36	79.42±7.29	79.42±7.29	79.21±7.35	88.87±5.92	89.64±5.23
	ModernTCN	76.15±3.85	77.15±3.25	75.94±3.83	75.94±3.83	75.73±4.05	84.42±4.30	84.40±4.18
	PatchTST	71.13±10.02	71.93±10.56	71.48±10.34	71.48±10.34	71.05±10.02	79.10±12.54	79.27±11.47
	TCN	89.00±9.14	89.05±9.28	88.93±9.25	88.93±9.25	88.93±9.23	94.02±6.85	93.44±7.78
	TimesNet	83.88±3.58	84.16±3.47	83.87±3.57	83.87±3.57	83.80±3.60	88.12±9.12	87.78±9.75
	BIOT	90.88±4.66	91.44±4.13	91.04±4.75	91.04±4.75	90.83±4.72	96.73±2.98	96.52±3.55
	LaBraM	91.23±4.37	92.10±3.37	91.16±4.38	91.16±4.38	91.12±4.50	97.54±1.88	97.58±1.94
	CBraMod	84.46±8.37	88.51±4.55	84.62±8.54	84.62±8.54	83.70±9.28	96.67±3.69	96.77±3.44
	CSBrain	93.18±4.03	93.66±3.56	93.36±3.78	93.36±3.78	93.16±4.03	98.87±1.14	98.92±1.10
	LEAD (Ours)	**97.42±2.74**	**97.60±2.59**	**97.51±2.57**	**97.51±2.57**	**97.42±2.74**	**100.00±0.00**	**100.00±0.00**
ADSZ *(Subject-Level)* *(48 Subjects)* *(HC vs AD)*	**ManualFeature**	70.00±24.49	60.00±33.91	70.00±24.49	70.00±24.49	63.05±30.68	70.00±24.49	70.00±24.49
	EEGConformer	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	EEGInception	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	EEGNet	66.67±18.26	67.00±26.38	66.67±18.26	66.67±18.26	61.38±22.11	66.67±18.26	65.33±18.45
	iTransformer	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	MedGNN	96.67±6.67	97.50±5.00	96.67±6.67	96.67±6.67	96.57±6.86	96.67±6.67	96.67±6.67
	Medformer	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	MNet	90.00±13.33	90.83±13.02	90.00±13.33	90.00±13.33	89.90±13.38	90.00±13.33	87.22±16.25
	ModernTCN	90.00±8.16	92.50±6.12	90.00±8.16	90.00±8.16	89.71±8.40	90.00±8.16	85.00±12.25
	PatchTST	86.67±19.44	87.50±19.36	86.67±19.44	86.67±19.44	86.29±19.99	86.67±19.44	86.67±19.44
	TCN	96.67±6.67	97.50±5.00	96.67±6.67	96.67±6.67	96.57±6.86	96.67±6.67	95.00±10.00
	TimesNet	96.67±6.67	97.50±5.00	96.67±6.67	96.67±6.67	96.57±6.86	96.67±6.67	95.00±10.00
	BIOT	90.88±4.66	91.44±4.13	91.04±4.75	91.04±4.75	90.83±4.72	96.73±2.98	96.52±3.55
	LaBraM	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	CBraMod	90.00±13.33	93.50±8.31	90.00±13.33	90.00±13.33	89.07±14.85	90.00±13.33	87.00±16.61
	CSBrain	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
	LEAD (Ours)	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**

Table 10: **Method Comparison Detailed Results on APAVA.** The detailed results comparison between our method and baseline methods on the dataset APAVA includes all 7 evaluation metrics. The **Top-1**, Top-2, and Top-3 results are highlighted in red, blue, and green.

Datasets	Models	Accuracy	Precision	Sensitivity	Specificity	F1 Score	AUROC	AUPRC
APAVA *(Sample-Level)* *(9,282 Samples)* *(HC vs AD)*	**ManualFeature**	70.40±4.92	77.53±8.45	64.55±5.25	64.55±5.25	63.22±6.51	66.49±5.78	64.17±5.79
	EEGConformer	76.90±3.19	81.85±2.78	72.65±4.10	72.65±4.10	73.05±4.93	83.07±3.70	83.43±3.70
	EEGInception	71.38±2.38	72.95±2.80	68.11±3.64	68.11±3.64	67.84±4.60	78.04±4.19	78.65±3.34
	EEGNet	71.38±2.38	72.95±2.80	68.11±3.64	68.11±3.64	67.84±4.60	78.04±4.19	78.65±3.34
	iTransformer	74.51±1.52	74.88±1.21	71.67±2.24	71.67±2.24	72.13±2.30	85.53±1.00	83.37±1.39
	MedGNN	73.71±4.64	81.38±2.00	68.40±5.99	68.40±5.99	67.54±7.95	81.39±3.83	82.69±3.50
	Medformer	71.99±1.67	74.97±2.64	67.47±2.03	67.47±2.03	67.42±2.44	78.38±2.75	78.33±2.96
	MNet	65.87±4.04	79.55±1.05	58.52±5.31	58.52±5.31	52.98±8.72	79.85±3.17	80.30±3.26
	ModernTCN	67.24±0.54	73.27±2.32	60.87±0.54	60.87±0.54	58.42±0.76	75.88±2.41	76.25±2.12
	PatchTST	64.54±0.37	75.24±6.16	57.34±0.79	57.34±0.79	52.39±2.65	73.76±3.11	72.85±3.27
	TCN	78.16±3.07	82.81±2.51	74.26±4.16	74.26±4.16	74.81±4.47	85.84±3.92	86.33±3.81
	TimesNet	66.71±3.77	74.47±6.25	59.80±4.51	59.80±4.51	55.95±6.89	56.01±4.93	59.13±5.35
	BIOT	81.62±3.83	84.08±2.35	78.92±5.17	78.92±5.17	79.55±5.36	**92.92±2.77**	**92.33±2.89**
	LaBraM	74.24±1.74	75.28±1.25	71.52±3.22	71.52±3.22	71.65±3.35	85.61±1.20	83.91±1.44
	CBraMod	75.51±3.19	76.27±4.83	73.93±1.74	73.93±1.74	74.10±2.40	83.10±1.85	82.97±2.10
	CSBrain	70.16±5.00	78.21±4.83	63.95±6.04	63.95±6.04	61.69±8.63	68.20±1.53	68.80±2.51
	LEAD (Ours)	**84.36±5.33**	**84.41±5.72**	**83.36±4.91**	**83.36±4.91**	**83.70±5.30**	92.51±4.33	92.23±4.80
APAVA *(Subject-Level)* *(23 Subjects)* *(HC vs AD)*	**ManualFeature**	65.00±12.25	60.00±28.58	65.00±12.25	65.00±12.25	57.33±19.60	65.00±12.25	60.00±8.16
	EEGConformer	70.00±10.00	71.67±23.33	70.00±10.00	70.00±10.00	65.33±16.00	70.00±10.00	63.33±6.67
	EEGInception	75.00±15.81	75.00±25.82	75.00±15.81	75.00±15.81	70.67±21.33	75.00±15.81	70.00±16.33
	EEGNet	75.00±15.81	75.00±25.82	75.00±15.81	75.00±15.81	70.67±21.33	75.00±15.81	70.00±16.33
	iTransformer	85.00±12.25	90.00±8.16	85.00±12.25	85.00±12.25	84.00±13.06	85.00±12.25	80.00±16.33
	MedGNN	70.00±10.00	71.67±23.33	70.00±10.00	70.00±10.00	65.33±16.00	70.00±10.00	63.33±6.67
	Medformer	75.00±0.00	83.33±0.00	75.00±0.00	75.00±0.00	73.33±0.00	75.00±0.00	66.67±0.00
	MNet	55.00±10.00	36.67±23.33	55.00±10.00	55.00±10.00	41.33±16.00	55.00±10.00	53.33±6.67
	ModernTCN	50.00±0.00	25.00±0.00	50.00±0.00	50.00±0.00	33.33±0.00	50.00±0.00	50.00±0.00
	PatchTST	50.00±0.00	25.00±0.00	50.00±0.00	50.00±0.00	33.33±0.00	50.00±0.00	50.00±0.00
	TCN	75.00±0.00	83.33±0.00	75.00±0.00	75.00±0.00	73.33±0.00	75.00±0.00	66.67±0.00
	TimesNet	50.00±0.00	25.00±0.00	50.00±0.00	50.00±0.00	33.33±0.00	50.00±0.00	50.00±0.00
	BIOT	95.00±10.00	96.67±6.67	95.00±10.00	95.00±10.00	94.67±10.67	95.00±10.00	93.33±13.33
	LaBraM	75.00±15.81	75.00±25.82	75.00±15.81	75.00±15.81	70.67±21.33	75.00±15.81	70.00±16.33
	CBraMod	75.00±0.00	83.33±0.00	75.00±0.00	75.00±0.00	73.33±0.00	75.00±0.00	68.33±3.33
	CSBrain	65.00±12.25	60.00±28.58	65.00±12.25	65.00±12.25	57.33±19.60	65.00±12.25	60.00±8.16
	LEAD (Ours)	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**

Table 11: **Method Comparison Detailed Results on CNBPM.** The detailed results comparison between ours and baseline methods on the dataset CNBPM includes all 7 evaluation metrics. The **Top-1**, Top-2, and Top-3 results are highlighted in red, blue, and green.

Datasets	Models	Accuracy	Precision	Sensitivity	Specificity	F1 Score	AUROC	AUPRC
CNBPM *(Sample-Level)* *(122,029 Samples)* *(HC vs MCI vs AD)*	**ManualFeature**	48.25±8.11	44.08±6.56	43.80±6.72	73.56±3.95	43.00±6.36	61.29±6.44	41.05±5.17
	EEGConformer	65.30±11.18	59.85±8.02	59.20±9.36	82.13±5.87	59.17±8.85	81.76±8.20	64.85±9.17
	EEGInception	63.46±10.08	59.60±8.14	60.67±7.85	82.05±4.86	58.56±8.02	81.73±8.18	63.95±8.58
	EEGNet	47.94±7.22	44.16±5.91	41.54±6.09	72.20±3.97	38.57±6.52	63.51±7.95	45.73±5.71
	iTransformer	59.25±11.23	52.26±7.65	51.91±7.86	78.98±5.65	51.37±8.01	74.99±10.16	56.17±8.73
	MedGNN	65.34±9.04	59.36±7.76	59.06±7.98	81.88±4.88	58.68±7.50	82.13±7.35	65.80±7.03
	Medformer	66.31±10.00	61.15±8.05	**60.75±8.15**	82.55±5.20	**60.39±7.68**	82.17±7.87	66.25±7.41
	MNet	61.43±13.81	55.07±11.03	50.29±9.33	78.82±6.67	47.04±9.76	77.03±11.79	58.95±10.49
	ModernTCN	61.45±10.53	57.87±7.74	56.64±7.91	80.86±5.12	55.94±8.10	77.85±9.30	59.69±9.22
	PatchTST	57.43±9.39	51.20±6.10	50.79±5.75	77.75±4.22	49.78±6.10	71.52±9.58	53.69±8.24
	TCN	66.04±7.33	58.90±5.09	59.22±4.67	82.65±3.92	58.71±5.02	82.01±6.02	63.40±5.42
	TimesNet	63.56±10.91	57.30±7.40	56.12±6.36	81.31±5.29	56.19±6.63	79.18±7.74	60.17±7.70
	BIOT	61.70±13.03	55.70±9.53	54.96±9.21	80.37±6.36	54.79±9.52	77.95±11.46	59.39±11.05
	LaBraM	61.01±12.20	53.32±7.36	53.59±6.21	79.88±6.05	52.66±7.11	78.44±9.59	59.05±8.20
	CBraMod	61.84±10.71	54.79±7.13	54.14±6.75	79.78±5.67	53.93±6.87	78.68±9.10	60.68±7.13
	CSBrain	**67.11±7.61**	60.96±5.77	60.59±4.80	**82.69±3.85**	60.30±5.41	**83.39±5.73**	**66.55±5.21**
	LEAD (Ours)	66.00±11.35	**62.51±7.23**	59.73±6.56	81.41±5.82	60.20±6.83	79.10±10.21	65.59±9.75
CNBPM *(Subject-Level)* *(189 Subjects)* *(HC vs MCI vs AD)*	**ManualFeature**	48.57±5.55	42.29±10.06	48.57±5.55	74.29±2.78	43.95±7.64	61.43±4.16	42.09±3.50
	EEGConformer	60.00±7.74	68.74±8.17	60.00±7.74	80.00±3.87	58.42±8.60	70.00±5.80	53.18±6.41
	EEGInception	60.95±7.62	61.35±9.41	60.95±7.62	80.48±3.81	59.77±8.06	70.71±5.71	53.16±6.39
	EEGNet	40.00±9.81	35.17±17.28	40.00±9.81	70.00±4.90	34.02±11.43	55.00±7.35	38.50±5.60
	iTransformer	56.19±7.00	57.95±9.63	56.19±7.00	78.10±3.50	52.85±8.58	67.14±5.25	49.03±5.92
	MedGNN	54.29±10.26	56.81±15.83	54.29±10.26	77.14±5.13	52.18±11.32	65.71±7.69	48.00±6.90
	Medformer	62.86±5.55	64.22±7.00	62.86±5.55	81.43±2.78	60.93±6.49	72.14±4.16	54.28±4.42
	MNet	49.52±5.71	34.50±3.91	49.52±5.71	74.76±2.86	39.97±4.38	62.14±4.29	42.79±3.51
	ModernTCN	63.81±7.13	66.42±10.20	63.81±7.13	81.90±3.56	61.96±9.90	72.86±5.35	56.23±5.39
	PatchTST	56.19±5.55	56.83±16.05	56.19±5.55	78.10±2.78	51.35±9.07	67.14±4.16	49.16±5.15
	TCN	63.81±3.81	63.66±3.19	63.81±3.81	81.90±1.90	62.66±3.41	72.86±2.86	54.45±3.21
	TimesNet	57.14±5.22	59.11±9.17	57.14±5.22	78.57±2.61	55.52±4.93	67.86±3.91	50.34±4.43
	BIOT	57.14±7.97	58.87±8.30	57.14±7.97	78.57±3.98	55.74±8.58	67.86±5.98	49.21±6.42
	LaBraM	53.33±4.67	50.22±7.99	53.33±4.67	76.67±2.33	50.11±6.53	65.00±3.50	45.76±3.58
	CBraMod	54.29±9.33	52.97±13.37	54.29±9.33	77.14±4.67	50.60±10.89	65.71±7.00	47.90±6.64
	CSBrain	63.81±4.86	64.93±7.29	63.81±4.86	81.90±2.43	62.41±5.80	72.86±3.64	54.65±4.85
	LEAD (Ours)	**66.67±6.73**	**70.08±8.94**	**66.67±6.73**	**83.33±3.37**	**65.94±6.66**	**75.00±5.05**	**57.90±6.69**

Table 12: **First Part of per-subject results of ADFTD.** Demographic information, MMSE, recording length, and per-subject results using the LOSO setting of all subjects from Sub-001 to Sub-065.

Subject ID	Gender	Age	Class	MMSE	Length	Sample-level Accuracy	Subject-level Accuracy
Sub-001	F	57	AD	16	0.25h	77.13%	✓
Sub-002	F	78	AD	22	0.33h	40.35%	✗
Sub-003	M	70	AD	14	0.16h	82.44%	✓
Sub-004	F	67	AD	20	0.29h	95.04%	✓
Sub-005	M	70	AD	22	0.30h	96.08%	✓
Sub-006	F	61	AD	14	0.28h	58.93%	✓
Sub-007	F	79	AD	20	0.33h	92.98%	✓
Sub-008	M	62	AD	16	0.33h	79.14%	✓
Sub-009	F	77	AD	23	0.28h	73.26%	✓
Sub-010	M	69	AD	20	0.44h	54.76%	✓
Sub-011	M	71	AD	22	0.29h	89.66%	✓
Sub-012	M	63	AD	18	0.34h	74.04%	✓
Sub-013	F	64	AD	20	0.36h	92.89%	✓
Sub-014	M	77	AD	14	0.40h	85.86%	✓
Sub-015	M	61	AD	18	0.33h	87.08%	✓
Sub-016	F	68	AD	14	0.39h	94.99%	✓
Sub-017	F	61	AD	6	0.35h	98.31%	✓
Sub-018	F	73	AD	23	0.34h	98.53%	✓
Sub-019	F	62	AD	14	0.34h	59.57%	✓
Sub-020	M	71	AD	4	0.35h	94.01%	✓
Sub-021	M	79	AD	22	0.30h	79.66%	✓
Sub-022	F	68	AD	20	0.30h	94.78%	✓
Sub-023	M	60	AD	16	0.25h	69.73%	✓
Sub-024	F	69	AD	20	0.23h	98.96%	✓
Sub-025	F	79	AD	20	0.26h	31.06%	✗
Sub-026	F	61	AD	18	0.27h	97.96%	✓
Sub-027	F	67	AD	16	0.26h	99.50%	✓
Sub-028	M	49	AD	20	0.33h	91.12%	✓
Sub-029	F	53	AD	16	0.23h	96.75%	✓
Sub-030	F	56	AD	20	0.18h	99.47%	✓
Sub-031	F	67	AD	22	0.34h	75.28%	✓
Sub-032	F	59	AD	20	0.24h	85.07%	✓
Sub-033	F	72	AD	20	0.22h	70.55%	✓
Sub-034	F	75	AD	18	0.29h	93.91%	✓
Sub-035	F	57	AD	22	0.32h	91.30%	✓
Sub-036	F	58	AD	9	0.26h	62.05%	✓
Sub-037	M	57	HC	30	0.37h	75.66%	✓
Sub-038	M	62	HC	30	0.35h	93.12%	✓
Sub-039	M	70	HC	30	0.34h	88.15%	✓
Sub-040	M	61	HC	30	0.32h	80.61%	✓
Sub-041	F	77	HC	30	0.34h	81.21%	✓
Sub-042	M	74	HC	30	0.33h	91.37%	✓
Sub-043	M	72	HC	30	0.31h	92.28%	✓
Sub-044	F	64	HC	30	0.38h	87.45%	✓
Sub-045	F	70	HC	30	0.34h	77.45%	✓
Sub-046	M	63	HC	30	0.36h	85.24%	✓
Sub-047	F	70	HC	30	0.34h	90.48%	✓
Sub-048	M	65	HC	30	0.38h	88.31%	✓
Sub-049	F	62	HC	30	0.34h	76.68%	✓
Sub-050	M	68	HC	30	0.31h	96.14%	✓
Sub-051	F	75	HC	30	0.33h	94.02%	✓
Sub-052	F	73	HC	30	0.33h	74.64%	✓
Sub-053	M	70	HC	30	0.31h	94.20%	✓
Sub-054	M	78	HC	30	0.35h	92.48%	✓
Sub-055	M	67	HC	30	0.33h	82.26%	✓
Sub-056	F	64	HC	30	0.34h	81.32%	✓
Sub-057	M	64	HC	30	0.32h	92.14%	✓
Sub-058	M	62	HC	30	0.27h	77.62%	✓
Sub-059	M	77	HC	30	0.30h	67.08%	✓
Sub-060	F	71	HC	30	0.33h	95.32%	✓
Sub-061	F	63	HC	30	0.33h	77.05%	✓
Sub-062	M	67	HC	30	0.35h	51.18%	✓
Sub-063	M	66	HC	30	0.34h	84.08%	✓
Sub-064	M	66	HC	30	0.35h	97.25%	✓
Sub-065	F	71	HC	30	0.30h	93.43%	✓

Table 13: **Second Part of per-subject results of ADFTD.** Demographic information, MMSE, recording length, and results using the LOSO setting of all subjects from Sub-066 to Sub-088. The average in the last line applies to all subjects in the two tables, from Sub-001 to Sub-088.

Subject ID	Gender	Age	Class	MMSE	Length	Sample-level Accuracy	Subject-level Accuracy
Sub-066	M	73	FTD	20	0.22h	67.17%	✓
Sub-067	M	66	FTD	24	0.28h	97.35%	✓
Sub-068	M	78	FTD	25	0.22h	26.06%	✗
Sub-069	M	70	FTD	22	0.31h	97.36%	✓
Sub-070	F	67	FTD	22	0.25h	68.77%	✓
Sub-071	M	62	FTD	20	0.31h	46.27%	✓
Sub-072	M	65	FTD	18	0.28h	50.94%	✓
Sub-073	F	57	FTD	22	0.39h	87.97%	✓
Sub-074	F	53	FTD	20	0.37h	94.46%	✓
Sub-075	F	71	FTD	22	0.31h	73.04%	✓
Sub-076	M	44	FTD	24	0.35h	77.96%	✓
Sub-077	M	61	FTD	22	0.27h	94.68%	✓
Sub-078	M	62	FTD	22	0.34h	90.94%	✓
Sub-079	F	60	FTD	18	0.25h	98.09%	✓
Sub-080	F	71	FTD	20	0.28h	94.20%	✓
Sub-081	F	61	FTD	18	0.25h	64.80%	✓
Sub-082	M	63	FTD	27	0.24h	65.38%	✓
Sub-083	F	68	FTD	20	0.27h	58.15%	✗
Sub-084	F	71	FTD	24	0.20h	15.91%	✗
Sub-085	M	64	FTD	26	0.18h	50.53%	✓
Sub-086	M	49	FTD	26	0.18h	20.11%	✗
Sub-087	M	73	FTD	24	0.19h	88.01%	✓
Sub-088	M	55	FTD	24	0.25h	68.81%	✓
Average	–	66.17	–	22.94	0.30h	79.74%	94.32%